The neural representation of facial expression

Richard Jason Harris

Submitted for the degree of PhD

University of York
Psychology

November 2012

Abstract

Faces provide information critical for effective social interactions. A face can be used to determine who someone is, where they are looking and how they are feeling. How these different aspects of a face are processed has proved a popular topic of research over the last 25 years. However, much of this research has focused on the perception of facial identity and as a result less is known about how facial expression is represented in the brain. For this reason, the primary aim of this thesis was to explore the neural representation of facial expression.

First, this thesis investigated which regions of the brain are sensitive to expression and how these regions represent facial expression. Two regions of the brain, the posterior superior temporal sulcus (pSTS) and the amygdala, were more sensitive to changes in facial expression than identity. There was, however, a dissociation between how these regions represented information about facial expression. The pSTS was sensitive to any change in facial expression, consistent with a continuous representation of expression. In comparison, the amygdala was only sensitive to changes in expression that resulted in a change in the emotion category. This reflects a more categorical response in which expressions are assigned into discrete categories of emotion.

Next, the representation of expression was further explored by asking what information from a face is used in the perception of expression. Photographic negation was used to disrupt the surface-based facial cues (i.e. pattern of light and dark across the face) while preserving the shape-based information carried by the features of the face. This manipulation had a minimal effect on judgements of expression, highlighting the important role of the shape-based information in judgements of expression. Furthermore, combining the photo negation technique with fMRI demonstrated that the representation of faces in the pSTS was predominately based on feature shape information.

Finally, the influence of facial identity on the neural representation of facial expression was measured. The pSTS, but not the amygdala, was most responsive to changes in facial expression when the identity of the face remained the same. It was found that this sensitivity to facial identity in the pSTS was a result of interactions with regions thought to be involved in the processing of facial identity. In this way identity information can be used to process expression in a socially meaningful way.

Table of Contents

Abstract..ii
Table of Contents iv
List of Figures.. viii
List of Tables xii
Acknowledgements..xiv
Declaration...xv

Chapter 1...1
  1.1 What information is available from a face? 1
  1.2 Are facial expression and identity processed independently? 2
    1.2.1 The independent processing of expression and identity – behavioural and neuropsychological evidence 2
    1.2.2 An interaction between facial expression and identity 5
  1.3 Neural regions involved in processing information from faces 7
    1.3.1 Face selectivity in the brain..8
  1.4 Facial expressions of emotion...11
    1.4.1 The universality of facial expressions..12
      1.4.1.1 Universal facial expressions...*12*
      1.4.1.2 Culturally dependent facial expressions...*15*
    1.4.2 Categorical and continuous representations of facial expression 17
      1.4.2.1 Categorical representations of facial expressions..........................*17*
      1.4.2.2 Continuous models of facial expression perception *18*
    1.4.3 Neural regions involved in the processing of facial expressions 19
      1.4.3.1 The STS and social communication..*20*
      1.4.3.2 Biological relevance processing in the amygdala *22*
      1.4.3.3 The representation of transient signals in the FFA......................*25*
  1.5 Thesis aims..25
Chapter 2...27
  2.1 Stimuli selection and manipulation 27
    2.1.1 The Facial Action Coding System 27
    2.1.2 The Ekman and Friesen POFA and the FEEST set 28
    2.1.3 Selection of face stimuli...29
    2.1.4 Stimuli for Experiment 1..31

2.1.5 Stimuli for Experiment 2 31

2.1.6 Stimuli for Experiment 3 35

2.1.7 Stimuli for Experiments 4 and 5 36

2.1.8 Stimuli for Experiments 6 and 7 37

2.2 fMRI Methods 37

2.2.1 fMRI 38

2.2.2 ROI approach 38

2.2.3 Functional localisation 40

2.2.3.1 Functional localiser 1 – stimuli and procedure 40

2.2.3.2 Functional localiser 2 – stimuli and procedure 42

2.2.4 fMRI analysis 43

2.2.5 fMRI protocol 45

Chapter 3 46

3.1 Introduction 46

3.2 Experiment 1: Sensitivity to facial expression and identity in face-selective neural regions 48

3.2.1 Introduction 48

3.2.2 Methods 48

3.2.2.1 Subjects 48

3.2.2.2 Face localiser scan 49

3.2.2.3 Experimental scan 49

3.2.2.4 Imaging parameters and fMRI analysis 50

3.2.3 Results 50

3.2.4 Discussion 53

3.3 Experiment 2: Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain 56

3.3.1 Introduction 56

3.3.2 Validation of expression continua 56

3.3.2.1 Expression-classification experiment 56

3.3.2.2 Same/different discrimination task 58

3.3.3 Methods – fMRI experiment 59

3.3.3.1 Subjects 59

3.3.3.2 Face localiser scan 60

3.3.3.3. Experimental scan 60

3.3.3.4 Imagining parameters and fMRI analysis 61

3.3.4 Results –fMRI experiment 61

3.3.5 Discussion 66

3.4 Experiment 3: Dynamic stimuli reveal a selectivity for facial expressions of emotion in the amygdala, but not in other face-selective regions of the human brain 71

3.4.1 Introduction 71

3.4.2 Methods 72

3.4.2.1 Subjects 72

3.4.2.2 Localiser scan 72

3.4.2.3 Experimental scan 72

3.4.2.4 Imaging parameters and fMRI analysis 73

3.4.3 Results 73

3.4.4 Discussion 77

3.5 Conclusions 80

Chapter 4 81

4.1 Introduction 81

4.2 Experiment 4: Photographic negation reveals the importance of shape-based facial cues the perception of expression 85

4.2.1 Introduction 85

4.2.2 Method – Study 1 85

4.2.2.1 Subjects 85

4.2.2.2 Procedure 86

4.2.3 Results – Study 1 87

4.2.4 Discussion – Study 1 89

4.2.5 Method – Study 2 89

4.2.5.1 Subjects 89

4.2.5.2 Procedure 90

4.2.6 Results – Study 2 91

4.2.7 Discussion – Study 2 92

4.2.8 General Discussion 92

4.3 Experiment 5: Shape-based representations of faces in the pSTS 95

4.3.1 Introduction 95

4.3.2 Method 95

4.3.2.1 Subjects 95

4.3.2.2 Face localiser scan 95

4.3.2.3 Experimental scan 95

4.3.3 Results 96

4.3.4 Discussion 100

4.4 Conclusions 103

Chapter 5 .... **105**

5.1 Introduction **105**

5.2 Experiment 6: Neural responses to expression and gaze in the posterior superior temporal sulcus interact with facial identity. **107**

5.2.1 Introduction 107

5.2.2 Methods 107

5.2.2.1 Subjects *107*

5.2.2.2 Stimuli and procedure *107*

5.2.2.3 Imaging parameters *110*

5.2.2.4 Whole brain analysis *110*

5.2.3 Results 111

5.2.4 Discussion 115

5.3 Experiment 7: Interaction of the pSTS and other face-selective neural regions when processing the changeable aspects of a face **118**

5.3.1 Introduction 118

5.3.2 Methods 119

5.3.2.1 Functional connectivity analysis *119*

5.3.3 Results 122

5.3.4 Discussion 125

5.4 Conclusions **128**

General Discussion **129**

6.1 What is the influence of facial identity on the processing of facial expression? **130**

6.2 How are facial expression represented in the brain? **132**

6.3 What information present in the face is critical for facial expression perception? **134**

6.4 Conclusions **135**

Appendices **136**

7.1 Supplementary Figures **136**

7.1.1 Chapter 2 136

7.1.2 Chapter 4 137

7.2 Supplementary Tables **141**

7.2.1 Chapter 2 141

7.2.2 Chapter 3 145

References **146**

List of Figures

Figure 1.1 The Bruce and Young (1986) model of face processing.

Figure 1.2 Haxby et al., (2000) model of face processing.

Figure 2.1 Initial face stimuli selected for Experiments 1-5: five actors posing five expressions.

Figure 2.2 Outline of the method used to generate average facial expressions.

Figure 2.3 Example of a morphed expression continuum: Happiness to disgust for actor F8.

Figure 2.4 Linearity of expression continua: (a) mean image difference and (b) mean image correlations.

Figure 2.5 Example of a movie stimulus: Frames that constituted F8 happiness movie.

Figure 2.6 Contrast reversed stimuli for Experiments 4-5.

Figure 2.7 Examples of the conditions in Localiser Version 1.

Figure 2.8 Examples of the conditions in Localiser Version 2.

Figure 3.1 Examples of the conditions in Experiment 1.

Figure 3.2 Locations of face-selective regions in Experiment 1. Average location of each region of interest (ROI) across participants transformed into standard space.

Figure 3.3 Experiment 1 results: Peak response to the different conditions in face-selective regions of interest.

Figure 3.4 Results from the expression-classification experiment. The figure shows the results for the four continua that were most accurately recognised (a) fear-happy, (b) happy-disgust, (c) disgust-fear (d) disgust-sad.

Figure 3.5 Results from the expression-discrimination experiment. (a) Proportion of ‘different’ responses. (b) Reaction time for the correct responses.

Figure 3.6 Examples of the three expression conditions used in Experiment 2 for (a) the same identity and (b) different identity conditions.

Figure 3.7 Locations of face-selective regions in Experiment 2. Average location of each ROI across participants transformed into standard space.

Figure 3.8 Experiment 2 results. Peak responses to the different conditions in the pSTS and amygdala.

Figure 3.9 Experiment 2 results. Peak responses to the different conditions in the OFA and FFA.

Figure 3.10 Examples of the conditions in Experiment 3.

Figure 3.11 Locations of face-selective regions in Experiment 3. Average location of each ROI across participants transformed into standard space.

Figure 3.12 Experiment 3 results. Peak responses to the four conditions in the pSTS and amygdala.

Figure 3.13 Experiment 3 results. Peak responses to the four conditions in the OFA and FFA.

Figure 4.1 Example trials of the conditions in Study 1 (Experiment 4).

Figure 4.2 Study 1 (Experiment 4) results. (a) Percent error. (b) Reaction time for the correct trials.

Figure 4.3 Example trials of the conditions in Study 2 (Experiment 4).

Figure 4.4 Study 2 (Experiment 4) results. (a) Percent error (b) Reaction time for the correct trials.

Figure 4.5 Locations of face-selective ROIs in Experiment 5. Average location of each ROI across participants transformed into standard space.

Figure 4.6 Experiment 5 results. Peak responses to the different conditions in the pSTS, FFA and OFA.

Figure 5.1 Examples of the conditions used in Experiments 6 and 7. (a) Face conditions (b) Non-face stimulus conditions.

Figure 5.2 Comparisons of the image variability between successive images in the two face conditions in Experiments 6 and 7. (a) Mean pixel difference. (b) Mean image correlation.

Figure 5.3 Experiment 6 whole brain analysis. (a) Responses to the same identity and different identity face conditions. (b) Responses to the same identity faces compared to non-face stimuli. (c) Responses to the different identity faces compared to the non-face stimuli.

Figure 5.4 Experiment 6 whole brain analysis.

Figure 5.5 Method for deriving residual activation within a region of interest for a single individual.

Figure 5.6 Average timecourse of activation across 103 participants for the four regions of interest in Experiment 7 (Occipital pole, pSTS, OFA and FFA).

Figure 5.7 Method for calculating the functional connectivity between face-selective regions. (a) Timecourse of activation for two regions of interest for one participant. (b) General linear model for the two regions. (c) Residual activation. (d) Correlations between the two regions.

Figure 5.8 Experiment 7 results. Mean correlations of residual activity between the three STS, OFA and FFA during the same face identity and different face identity conditions.

Figure 5.9 Results for Experiment 7. Mean correlations of residual activity between three face-selective regions and the occipital pole during the same face identity and different face identity conditions.

Figure A.1 Peak responses in four face-selective regions to faces and non-face stimuli conditions in the localiser scan. (a) Experiment 1 (b) Experiment 2 (c) Experiment 3.

Figure A.2 Study 1 (Experiment 4) results (percent error). (a) Expression judgements, positive contrast. (b) Expression, judgements, negative contrast. (c) Identity judgements, positive contrast. (d) Identity judgements, negative contrast.

Figure A.3 Study 1 (Experiment 4) results (reaction time). Details as reported in Figure A.2.

Figure A.4 Study 2 (Experiment 4) results (percent error). (a) Expression judgments. (b) Identity judgements.

Figure A.5 Study 2 (Experiment 4) results (reaction time). Details as reported in Figure A.4.

Figure A.6 Peak responses from the amygdala to all six conditions in Experiment 5.

List of Tables

Table 3.1 MNI Coordinates of face-selective regions of interest in Experiment 1.

Table 3.2 MNI Coordinates of face-selective regions of interest in Experiment 2.

Table 3.3 MNI Coordinates of face-selective regions of interest in Experiment 3.

Table 4.1 MNI Coordinates of face-selective regions of interest in Experiment 5.

Table 5.1 Locations of neural regions showing a greater response to the same identity faces compared to different identity faces in Experiment 6.

Table 5.2 Locations of neural regions showing a greater response to the different identity faces compared to the same identity faces in Experiment 6.

Table 5.3 Locations of the core face-selective regions, in Experiment 6, defined by the contrast of same identity faces > bodies, objects, places and scrambled images.

Table 5.4 Locations of the core face-selective regions, in Experiment 6, defined by the contrast of different identity faces > bodies, objects, places and scrambled images.

Table B.1 Action units used by each of the selected actors when posing each of the five expressions (as reported in the FEEST set, 2002).

Table B.2 Results from a recognition expression experiment reported in the FEEST set (2002). The results show the recognition rate of the expression for the five actors and five expressions selected as stimuli.

Table B.3 Percent recognition rate for the selected face stimuli when stimuli were presented for 500 ms.

Table B.4 Percent recognition rate for the selected face stimuli when stimuli were presented for 1000 ms.

Table B.5 Percent recognition rate for the selected face stimuli when stimuli were presented for 1500 ms.

Table B.6 Results from a behavioural experiment comparing the recognition rate of dynamic and static facial expression. Percent recognition rate for dynamic expressions.

Table B.7 Results from a behavioural experiment comparing the recognition rate of dynamic and static facial expression. Percent recognition rate for static expressions.

Table B.8 Expression-categorisation experiment results.

Acknowledgements

There are many people who have made the past three years enjoyable and successful, to whom I am grateful. I would like to thank my supervisors Tim and Andy, their doors were always open. Their continued advice and support proved invaluable throughout the PhD. With their guidance I now have a true appreciation for just how aligned two figures can be. I also appreciate the support from my colleagues at YNiC and the Department, in particular Heidi who always provided useful suggestions.

My parents, brother and grandma have been integral to my success over the years. I will always appreciate their kind love and encouragement. My friends, both old and new, have always been there for me both academically and socially. Particularly Ed Silson, his constructive discussions and feedback have proved useful at many points throughout the PhD.

And Laura, your endless love and kindness has made not only my PhD possible but all that I have achieved since I met you. Love you always.

Declaration

This thesis contains original work completed by myself under the supervision of Prof. Timothy J. Andrews and Prof. Andy Young.

Data reported in **Chapter Three (Experiments 1 & 2)** are in press as:

Harris, R.J., Young, A.W., & Andrews, T.J. (In press). Morphing between expressions dissociates continuous from categorical representations of facial expressions in the human brain. *PNAS.*

and were presented at the Vision Sciences Society, 2012, Naples, Florida, USA, as:

Harris, R.J., Young, A.W., & Andrews, T.J. (2012). The role of the pSTS in the pre-categorical coding of emotional expression [Abstract]. *Journal of Vision, 12(9),* 783.

The data reported in **Chapter Three (Experiments 1 & 3)** were presented at the European Conference on Visual Perception, 2011, Toulouse, France as:

Harris, R.J., Young, A.W., & Andrews, T.J. (2011). Representation of expression and identity in face selective regions. *Perception, 40,* 117.

The data reported in **Chapter Five** have been accepted for publication as:

Baseler, H.A. Harris, R.J., Young, A.W., & Andrews, T.J. (In press). Neural responses to expression and gaze in the posterior superior temporal sulcus interact with facial identity. *Cerebral Cortex.*

and were presented at the Vision Sciences Society, 1012, Naples, Florida, USA, as:

Baseler, H.A. Harris, R.J., Young, A.W., & Andrews, T.J. (2012). Identity modulates pSTS response to changeable aspects of faces. *Journal of Vision, 12(9),* 782.

R. J. Harris contributed equally to the design, stimulus generation, methods, analysis and write-up of the experiments reported in Chapter 5.

Chapter 1

The representation of facial expression in the brain

1.1 What information is available from a face?

Faces convey a wealth of information critical for effective social interactions. Information from a face can be used to determine who someone is, what sex they are, how they are feeling, where they are attending, and so on. All faces are essentially identical in design, with two eyes above a centrally placed nose situated above a mouth. This layout is not restricted to human faces but extends to primates and all other mammals. The similarity in configuration of human faces requires human observers to be highly sensitive to subtle variations between faces in order to extract the valuable cues a face provides. These cues can broadly be organised into two types which are thought to be extracted relatively independently from each other; relatively static invariant cues and transient changes in facial musculature.

Static facial cues allow the observer to discriminate between different identities and categorise the face. These invariant cues, such as the 3D structure of the face as well as the surface colour and pigmentation of the skin, allow the observer to accurately judge the age, gender and the identity of the face (Bruce & Langton, 1994; Bruce et al., 1991, 1993; Burt & Perrett, 1995; Ellis, Shepherd, & Davies, 1979; Russell, 2003, 2009). These static cues also allow for the categorisation of more abstract characteristics. Cook (1939), for example, asked participant’s to rate different faces on intelligence. He found a consensus between participant’s judgements even though these judgements did not reflect a person’s actual intelligence.

Dynamic changes in facial musculature, on the other hand, serve as means of communication. Transient changes of facial muscles can signal a person’s emotions, changes in eye gaze direction and head orientation can indicate an individual’s attention, and mouth movements are useful for decoding speech. Sensitivity to these cues and the accurate interpretation of their meaning are important for social communication and our welfare; these signals allow for the effective communication of potential physical threats within the environment. For example a fearful facial expression warns of impending danger, and a disgust expression can indicate harmful substances that are best avoided.

Despite the importance attached to the accurate interpretation of the changeable aspects of a face, much of the research investigating the neural correlates of face perception has focused on the representation of identity. This thesis aims to explore the neural processing and representation of the changeable aspects of a face. Specifically, one of the primary goals of this thesis is to directly compare the representation of expression in the superior temporal sulcus and the amygdala. These two neural regions are thought to be of considerable importance in the processing of facial expression but have seldom been investigated within the same experimental paradigm.

1.2 Are facial expression and identity processed independently?

In order to investigate the representation of facial expression, the extent to which expression interacts with the other information available from a face must be considered. The extent to which expression and identity processing can be considered independent remains an unresolved and controversial issue. Initial cognitive and neuropsychological studies suggested relatively distinct processing of these facial signals and this approach was incorporated into functional models of face processing (Bruce & Young, 1986). More recent findings, however, raise doubt as to whether the processing of identity and expression can be considered completely independent. This section briefly outlines the research concerning the independence and interactions between facial expression and identity.

1.2.1 The independent processing of expression and identity – behavioural and neuropsychological evidence

Due to the considerable importance attached to different types of facial information, the most efficient way to extract this information is thought to involve different neural subcomponents that are optimally tuned for particular types of face signal (Bruce & Young, 1986; 2012; Haxby, Hoffman, & Gobbini, 2000). As such, dynamic changes in a face are thought to be extracted relatively independently from the more static and invariant cues that gives rise to a person's identity.

Initial behavioural experiments highlighted the possibility of the independent processing of identity and expression (Campbell et al., 1996). In one of the first experiments to consider this issue Young, McWeeny, Hay and Ellis (1986) asked participants to judge whether the expression or identity was the same across pairs of famous or unfamiliar faces. Judgements of identity were quicker for famous compared to unfamiliar faces, however, judgements of expression were unaffected by the familiarity of the face. This suggests that cues important for expression judgements are extracted independently from identity cues. Similar findings were also demonstrated by Bruce (1986) who found no effect of face familiarity on judgements of facial expressions.

Rather than manipulating the level of familiarity, Calder, Young, Keane and Dean (2000) investigated expression and identity processing with the composite face effect. The composite face effect was first demonstrated by Young, Hellawell and Hay (1987) in which participants found it harder to recognise the top or bottom part of a face when they were aligned compared to when they were misaligned. Using this same technique Calder et al., found that it was harder to recognise the expression in either the top or bottom half of the face when faces were aligned. They also found the same composite effect for identity. Importantly, the composite face effect for expression and identity were independent from each other; the composite face effect for expression did not differ if the identity of the two halves of the face were the same or different. The same was true for the composite face effect for identity which did not differ whether the two halves of the face had the same or different expression.

Converging evidence from neuropsychological case studies also indicates dissociable processing of facial expression and identity. Patients with prosopagnosia, an inability to recognise the identity of a face, sometimes show a relatively intact recognition of facial expression (Baudouin & Humphreys, 2006; Bruyer et al., 1983; Shuttleworth, Syring, & Allen, 1982; Tranel, Damasio, & Damasio, 1988). Interpretations of neuropsychological case studies are more compelling when a double dissociation is demonstrated. To this effect patients have also been identified that show impaired recognition of facial expression with intact recognition of identity (Humphreys, Donnelly, & Riddoch, 1993; Parry, Young, Saul, & Moss, 1991). However, the inferences that can be made from a double dissociation identified between different experiments are limited; different case studies use different experimental tasks and therefore results from different experiments are not directly comparable. Young, Newcombe, de Haan, Small and Hay (1993) took a more comprehensive approach to this problem by studying 34 ex-servicemen with discrete brain injuries following shrapnel wounds. They investigated patients' abilities on two tests of familiar face recognition, two tests of unfamiliar face matching and two tests of expression recognition. This allowed them to identify several cases of selective impairments in which patients were impaired on both tasks for a particular face ability but not the others.

Early evidence from cognitive experiments and neuropsychological case studies provided the basis for the Bruce and Young (1986) model of face processing. This functional model proposes that after the initial structural encoding of the face information bifurcates with expression, facial speech and identity analysed along independent processing streams (see Figure 1.1). The initial encoding of the face produces viewer centred pictorial codes that are directly used for the analysis of facial expression in which the emotion is categorised and for the analysis of facial speech in which lip movements are categorised. In order for the identity of face to be recognised more abstract codes are required that are free from irrelevant variations in facial expression. These abstract codes are compared with face recognitions units (FRUs) which contain descriptions of familiar faces. Signals are sent from the FRUs to the cognitive system as well as the person identity nodes (PINs) which act as an interface with identity specific semantic codes. This model highlights that not only are facial expression and identity processed separately after the initial structural encoding of the face but the processing of this information is underpinned by different codes. In order to recognise identity, expression-free codes are required; otherwise a change in expression may lead to the perception of a different identity.

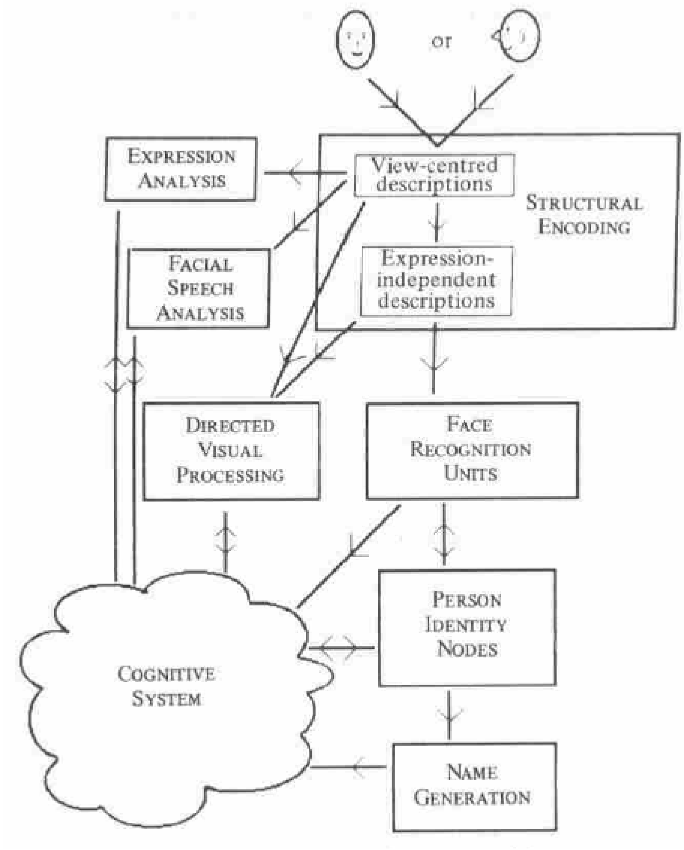

Figure 1.1 *The Bruce and Young (1986) model of face processing*

1.2.2 An interaction between facial expression and identity

A complete separation between the processing of facial expression and identity may be an over-simplification. In order to be socially meaningful facial expression must be tracked across the same identity. Furthermore, the interpretation of and the physiological response to expression is dependent on static cues such as the person's identity, age or gender (Ekman, 1975); for example, the relevance and response to an expression may be different if posed by a middle-aged male stranger rather than a familiar young female. Schweinberger and Soukup (1998) demonstrated the importance of identity information on the analysis of facial expression. They asked participants to judge either the identity or the expression of two faces. The unattended dimension could either stay the same or vary. They found that irrelevant variations in identity had an effect on judgements of expression. This relationship was asymmetrical and irrelevant variations in expression did not affect judgements of identity. In a follow up experiment, Schweinberger, Burton, and Kelly (1999) demonstrated that this influence of identity on expression was not due to the relative processing speed of identity and expression.

Another line of evidence for the interaction of identity and expression comes from the consideration of expression adaptation after-effects. Webster, Kaping, Mizokami, and Duhamel (2004) and Hsu and Young (2004) initially reported adaptation after-effects for facial expressions; prolonged presentation of a particular expression led to decreased sensitivity to that expression on a subsequent presentation. Using this perceptual after-effect, follow-up experiments have investigated the effect of identity on after-effects of expression by manipulating the identity of the adapting and test stimulus. These studies found that although adaptation after-effects were present when the adapting and test stimuli had different identities, the effect was more pronounced when the identity remained constant (Campbell & Burke, 2009; Ellamil, Susskind, & Anderson, 2008; Fox & Barton, 2007).

Interestingly the interaction between expression and identity appears asymmetrical. In the above studies there was an effect of irrelevant variations of identity on judgements of expression but there was little influence of expression on judgements of identity (Fox & Barton, 2007; Schweinberger et al., 1999; Schweinberger & Soukup, 1998). This asymmetry is consistent with the view that identity information may be used to interpret the expression in a socially meaningful way.

In the preceding section, evidence from neuropsychology was presented showing a possible double dissection between expression and identity processing. Inferences, however, from neuropsychological case studies must be interpreted with caution due to the inherent limitations of this approach. This is particularly evident when considering prosopagnosic patients that demonstrate relatively intact facial expression recognition, as it is rare to find prosopagnosic patients that do not show some impairment in facial expression (Calder & Young, 2005).

A further limitation of some previously reported prospagnosia studies is the failure to identify the cause of the identity problem. Much of the evidence to support the double dissociation comes from studies in which the cause of the identity deficit was not identified precisely (for example Etcoff, 1984; Parry et al., 1991; Shuttleworth et al., 1982; Young et al., 1993). As such, identity deficits may not necessarily reflect a visual-perceptual problem but stem from impairments with learning or access to person knowledge (Calder & Young, 2005).

In order to show a clear double dissociation the identity and expression tasks must be matched on their level of difficulty. It is seldom the case that both expression and identity tasks are matched in this way; identity tasks usually involve the recognition of famous people, whereas expression tasks often involve matching expressions (see Calder & Young, 2005; Young et al., 1993). This is of particular importance as if the two tasks are not matched for relative difficulty then patients' better performance in one task may reflect the relative ease of that task rather than a spared ability.

Neuropsychological evidence, then, is at present unable to conclusively demonstrate a dissociation between facial expression and identity processing. Often, impairments in identity recognition are accompanied by impairments on expression recognition; however these impairments are often overlooked due to problems mentioned above. The greater impairment on identity compared to expression recognition may reflect a difference in task difficulty. The lack of a conclusive double dissociation in conjunction with the above cognitive literature suggests there is at least in some part an interaction between the processing of identity and expression. The influence of identity on the neural processing of the changeable aspects of a face is addressed in Chapter 5.

1.3 Neural regions involved in processing information from faces

The neural regions important in the processing of information from faces are now considered. First, neural regions that demonstrate face-selective responses (i.e. neural regions in which the response is modulated more by faces than non-face stimuli) across a range of experimental methods, such as single cell recordings, ERP and fMRI are detailed. Then a neuro-anatomical model of face processing which integrates these face-selective neural regions and defines their roles within the face processing network is outlined.

1.3.1 Face selectivity in the brain

Single cell recordings provide a unique contribution to the understanding of the cortical representation that sub-serves face perception. Recording from individual neurons allows the identification of the response properties of a single cell. This technique has identified a number of neural regions in which there are cells that respond selectively to faces compared to non-face stimuli such as complex 3D objects and scrambled faces (Baylis, Rolls, & Leonard, 1985; Leonard, Rolls, Wilson, & Baylis, 1985; Perrett & Rolls, 1983; Sanghera, Rolls, & Roper-Hall, 1979). Furthermore these cells are also insensitive to variations in colour and the size of the face, whilst some of these face-selective neurons responded more to isolated facial features or the combinations of these features or specific orientations (Bruce, Desimone, & Gross, 1981; Hasselmo, Rolls, & Baylis, 1989; Perrett, Rolls, & Caan, 1979, 1982). An interesting approach to investigating the response properties of neurons at a cellular level was conducted by Tsao, Freiwald, Tootell, and Livingstone (2006). They first identified face-selective regions in monkey cortex using fMRI and then investigated the response of the individual neurons within face-selective voxels. They found that 97 % of the neurons tested in this region were selective for faces.

Additional support for face selective neural regions is provided by electrophysiological recordings from ERP and MEG studies. Using ERP, studies have found a larger negative ERP component at 170 ms at occipito-temporal electrodes to faces compared to non-face stimuli (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Bötzel, Schulze, & Stodieck, 1995; Jeffreys, 1996). Similarly, studies using MEG have also found a similar M170 component to faces (Halgren, Raij, Marinkovic, Jousmäki, & Hari, 2000; Harris & Nakayama, 2008; Liu, Higuchi, Marantz, & Kanwisher, 2000). Recently, however, Thierry, Martin, Downing, and Pegna (2007) have suggested that the N170/M170 component for faces can be attributed to variations in lower-level image properties; the high degree of similarity that naturally exists between faces results in less image variability than that of an array of objects. This image-based interpretation of the N170/M170 seems unlikely as studies which have controlled for image variability and category membership still produce the N170/M170 face component (Goffaux, Gauthier, & Rossion, 2003; Rossion & Jacques, 2008).

Improving upon the limited spatial resolution associated with MEG and ERP, fMRI studies are able to identify clusters of voxels that show face-selective responses. A common method for identifying face-selective regions in the brain is to conduct functional localiser scans. These scans look for clusters of voxels that respond to a specific stimulus category. Face localisers, for example, look for clusters that respond more to faces than non-face stimuli such as objects or scenes. Functionally defining a region of interest increases the statistical power of the experimental paradigm by decreasing the number of voxels investigated. A further advantage of this approach is highlighted when comparing responses across participants. A whole-brain analysis, by registering all the participants' data to the same common space, also allows for the comparison across participants. However, due to the variability of the cortex between participants, registration to a common space is imperfect. Defining regions functionally within participants allows the identification of landmarks that can then be compared across participants (Saxe, Brett, & Kanwisher, 2006).

Localiser scans have been used extensively in studies of face processing. In a ground-breaking study Kanwisher, McDermott, and Chun (1997) found that a face localiser scan consistently identified a region, across participants, in the fusiform gyrus that responded more to faces than non-face stimuli; they termed this region the fusiform face area (FFA). They continued to investigate the face-selective properties of the localised FFA by comparing this region's response to faces, houses, hands and scrambled stimuli. The FFA consistently responded more to faces than all the non-face stimuli. This face-selective response in the FFA has been demonstrated across a range of studies (Andrews & Schluppeck, 2004; Andrews, Schluppeck, Homfray, Matthews, & Blakemore, 2002; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; O'Craven & Kanwisher, 2000) with the response in this region often two to three times greater for faces than non-face stimuli (Yovel & Kanwisher, 2005). Interestingly, the response in this region to faces correlates with behavioural measures of face detection and identification (Grill-Spector, Knouf, & Kanwisher, 2004).

Face-selective responses are not confined to the fusiform gyrus and the use of functional localiser scans has identified other face-selective neural regions. These regions include the inferior occipital gyri (occipital face area, or OFA), the superior temporal sulcus (STS) and the amygdala (Andrews & Ewbank, 2004; Fox, Moon, Iaria, & Barton, 2009). Localisation of these face-selective regions appears robust across the different tasks used in different face-localiser scans; for example Berman et al., (2010) found no difference in localisation of the FFA when the task was varied (passive viewing, 1 back memory task, 2 back memory task).

The aforementioned neural regions typically identified in face localiser scans have been integrated into a neuro-anatomical model of face processing. Haxby and colleagues (2000) model parallels the cognitive model proposed by Bruce and Young (1986) by suggesting different neural subcomponents are optimally tuned to different aspects of the face. They divide their model into a core system comprised of neural regions in which the response is predominantly driven by faces, and an extended system in which regions, not necessarily face selective, are recruited to further analyse the information from a face. Within the core system the inferior occipital gyri (OFA) is responsible for the initial structural encoding of the face. From this region information bifurcates. A route to the lateral fusiform gyrus (FFA) is responsible for the processing of the invariant features of a face such as identity. In a parallel route to the STS information regarding the changeable aspects of a face such as expression and eye gaze is represented. The STS then has reciprocal connections with the amygdala, situated in the extended system, which is recruited for further appraisal of the emotion information (see Figure 1.2).

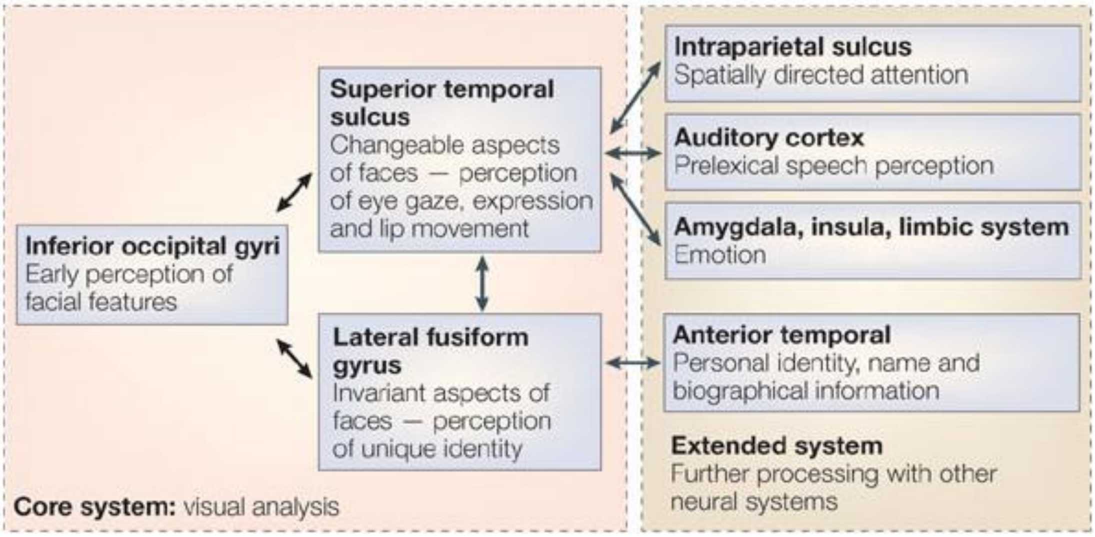

Figure 1.2 *Haxby et al., (2000) model of face processing. This model integrates neural regions typically identified by face localiser scans (from Calder & Young, 2005).*

1.4 Facial expressions of emotion

A specific aim of this thesis is to explore the processing and representation of facial expression in the brain. The remainder of this chapter outlines some of the important issues concerning the perception of facial expressions. Facial expressions of emotions are displayed as transient changes in the facial musculature with highly important communicative value. These changes in facial musculature primarily alter the shapes of facial features (eye widening in fear, upturned corners of the mouth in happiness, and so on), but they can also sometimes introduce local changes in brightness and texture patterns (for example, opening the mouth in surprise or showing the teeth in a grin). Facial expressions are categorised based on the holistic analysis of critical combinations of these facial features (Calder et al., 2000). As such, it is the more commonly encountered overall shape changes in the face that are thought to be of particular importance in the representation of facial expression (White & Li, 2006). This is reflected in the accurate recognition of expression from line drawings of faces that only provide feature shape-based information (Kirita & Endo, 1995; Magnussen, Sunde, & Dyrnes, 1994; Mckelvie, 1973).

Whether these dynamic signals, however, are recognised universally around the world or whether displays and the interpretation of expressions are relative to each individual culture remains a controversial issue. Furthermore, this controversy extends to how facial expressions are recognised. For example, are expressions recognised as belonging to discrete emotion categories or as points along continuous dimensions? This section will first explore universal and cultural accounts of facial expressions recognition followed by how these expressions are represented. Finally the neural regions involved in processing facial expressions are outlined.

1.4.1 The universality of facial expressions

There remains a long-standing debate as to whether facial expressions of emotions are displayed and recognised consistently across the world or whether facial expressions are learnt products of culture and therefore vary between different societies. These two approaches to facial expressions are now considered.

1.4.1.1 Universal facial expressions

Charles Darwin was one of the earliest proponents of the universality theory of facial expressions. In his book *the expression of the emotions in man and animals* Darwin (1872) explored the notion that expressions were not only important in our evolutionary past due to their communicative value but they also served some biologically adaptive function. The specific patterns of facial movements associated with particular expressions of emotions have a physiological benefit to the organism. For example, when disgusted, the tightening of the lips and wrinkling of the nose serves to prevent the ingestion of harmful substances (Darwin, 1872; Ekman, 1972). In support of this, Susskind et al., (2008) found posing fearful facial expressions resulted in participants having an increased nasal volume, air flow, and eye movements. This is a highly appropriate physiological response to fear, placing the organism in a state of readiness in order to respond to a potentially harmful situation. The biological benefits as well as the powerful communicative tool facial expressions provided in our evolutionary past ensured they became inherited and universally posed, not just by humans but by primates and other animals.

In developing this universality approach Darwin (1872) drew upon a variety of sources. To demonstrate that specific facial expression were inherited and not learnt, he cited evidence from observations of blind patients and studied his children from birth. To demonstrate the universality of expressions he sent questionnaires to British settlers across the world detailing questions set to probe the displays of expressions in the variety of cultures they encountered. Furthermore he described how facial expressions were not only present in man but can also be described in animals. Finally he presented photographs of facial expressions to British people and asked them to judge what expressions were presented.

There are, of course, obvious limitations with Darwin's methods for collecting evidence which were open to bias and could easily be misinterpreted. Facial expressions displayed by his own children and blind patients can be easily misjudged by a biased viewer. The questions that Darwin sent to British settlers often stated which pattern of facial movements are associated with particular expressions and thus biased the interpretations of the observers. This ethnocentric approach allowed critics to suggest that universal expression were only demonstrated due to the misinterpretation of a biased viewer (Ekman, 1980).

Darwin's universal theory of facial expressions was readdressed nearly a hundred years later by Ekman and his colleagues. Ekman, like Darwin, agreed that for a number of primary emotions there were distinctive patterns of facial movements that were panhuman (Ekman, 1972, 1973, 1980). These specific facial movements are thought to occur automatically without cognitive processing, they are specific to an emotion, distinguishable from other emotions and are universally recognised (Ekman, 1972). Ekman refers to this as the 'facial affect programme' which links each primary emotion to a distinct pattern of neural firing which moves the facial muscles. Seven primary universal expressions fall within the facial affect programme; happiness, sadness, anger, fear, disgust, surprise and interest. Later, the emotion state of interest was removed from the list of basic emotions (see Ekman, 1980).

Ekman did not disregard the role of culture in his universal approach to facial expression display and recognition. Ekman's theory was in fact a neural-cultural theory of facial expression; although there are universal patterns of facial movements for each emotional expression, culture can influence how these are displayed

(Ekman, 1972). Culture can modify facial expressions of emotions in two ways; (1) what elicits an emotion and the associated pattern of facial movements varies between cultures – what elicits anger in one culture could elicit sadness in another. (2) Display rules, which refers to cultural norms that control how facial expressions are displayed; for example certain cultures may exaggerate facial movements for a particular expression others may not. Although the basic emotions have specific patterns of facial movements, cultural differences inevitably led to variations in the display of facial expressions.

To demonstrate that the recognition of facial expression was universal Ekman sought to test judgements of facial expression across a range of cultures. Ekman and colleagues presented participants from five countries (USA, Brazil, Chile, Argentina and Japan) with facial expressions and asked them to select what expression was shown. If facial expressions are learnt and culturally dependent then judgements of expressions should vary with culture. The judgements of expressions were, however, highly correlated within and between cultures. Furthermore, high correlations were found between the cultures when judging the intensity of expressions on a 7-point scale (the minimum correlation was an r^2 of 0.93 between Chile and the USA) (Ekman, 1972, 1980). Although, this demonstrates universality of expression recognition, the recognition of facial expressions could have been influenced by western media in all the countries studied. To circumvent this limitation, Ekman and Friesen tested judgements of facial expression in a preliterate population in New Guinea isolated from western culture. They presented over 300 participants with different facial expressions and asked them to select which one best matched a short-story designed to elicit a single emotion. Mean accuracy for this task was 75 % and this was decreased due to trouble dissociating fear and surprise. Furthermore, to demonstrate that not only was the recognition of facial expressions universal but the patterns of facial movements used to pose expressions were also panhuman they filmed the facial expression of the New Guinea participants in response to the short-stories. These films were played to American students who had little trouble in identifying the expression being displayed (Ekman, 1972, 1980).

In further support of this theory Ekman (1994) conducted a review of studies investigating the cross-cultural recognition of facial expressions. Ekman reviewed a total of twenty studies on western cultures and eleven on non-western cultures in which participants judged facial expression on a 6 AFC. Ekman compared recognition rates to what would be expected if participants were performing at chance. Rather than setting chance at 16 % (1 in 6) Ekman used a more stringent measure of chance; for example chance for happiness was set at 50 % (see Ekman, 1994). Across the experiments the mean recognition rate for the western cultures was 48% above chance and for non-western cultures 35 %. This review clearly demonstrates the recognition of primary expressions spans cultures.

1.4.1.2 Culturally dependent facial expressions

The universal approach to facial expressions is not without its critics. Russell (1994), in a lengthy review, critiques the evidence supporting the universality of facial expressions. Russell raises questions over the level of agreement required for judgements of facial expression to be considered panhuman; if facial expressions are truly universal then recognition should be approaching near perfect. He reviews nine studies of facial expression recognition and reports a significant interaction between expression and culture. Russell further raises concerns over methodological issues; what is the best way to test different cultures, should single emotional labels be used or are rating scales more appropriate? What if there is no exact word to describe a particular emotion in every culture?

In a reply to this review, Ekman (1994) address the criticisms highlighted by Russell (1994). In order to demonstrate the universality of expression, an absolute agreement on judgements of facial expressions need not be obtained. In fact, cultural display rules will inevitably prevent an absolute agreement. What needs to be demonstrated is that expressions are recognised significantly above chance across cultures. As mentioned in the above section, this has been found. Many of the concerns raised over methodological issues have been addressed by the range of techniques used to collect participant’s judgements of expressions; for example, forced-choice paradigms were abandoned when testing participants in New Guinea due to the difficulty in translating emotional words.

Cultural variations in both the posing and recognition of expression have recently been identified. Elfenbein, Beaupré, Lévesque, and Hess (2007) found that although similar muscles were used to pose ten emotions, there were also significant differences in the muscles used across participants. Furthermore, a review of 87 studies found that although expression are universally recognised above chance, recognition increased when emotions were expressed by and recognised by members of the same national, ethnic, or regional group (Elfenbein & Ambady, 2002), suggesting a benefit for in-group recognition of facial expressions.

Eye movements when interpreting the meaning of a facial expression are also culturally bound. Jack, Blais, Scheepers, Schyns, and Caldara (2009) asked eastern and western participants to judge facial expressions whilst their saccades were recorded. They found consistent judgements of all 7 facial expressions apart from fear and disgust. Recordings of eye movements showed the eastern participants focused more on the eye region.

Although the above studies demonstrate cultural variations in the way expression are posed, recognised and what facial features are focused upon, this does not necessarily limit the universality of the neural-cultural model. The above studies are all susceptible to display rules that influence how expressions are posed and recognised; for example, in Elfenbein et al., (2007) participants were asked to pose expression in a way that their friends would be able to recognise them. Cultural variations in the expression of emotions are to be expected, but the degree of similarity these studies report, especially in regards to the basic emotions, demonstrates a universal aspect of facial expressions.

Taken together, the neural-cultural model of facial expression (Ekman, 1972) can account for the similarity displayed when posing expressions and the accuracy displayed when judging facial expressions across cultures. The neural component of the model highlights panhuman patterns of facial movements used to pose particular expressions. The cultural aspect of this model explains smaller variations between different cultures in how expressions are posed and understood. It appears the evidence to date can be reconciled within this model.

1.4.2 Categorical and continuous representations of facial expression

From the debate on the universality of facial expressions a further controversial issue regarding the processing of facial expressions has arisen; is the perception of facial expressions based on the assignment of expression into discrete categories of emotion or are expressions represented as gradations along continuous dimensions?

1.4.2.1 Categorical representations of facial expressions

As mentioned above, a key component of Ekman's neural-cultural model is the idea that discrete categories of basic emotions are recognised universally. Section 1.4.1.1 extensively reports Ekman's and colleagues work on the recognition of facial expressions which shows that distinct patterns of movements associated with six basic expressions are readily and consistently assigned into discrete emotion categories (Bruce & Young, 2012; Ekman, 1999). Categorical perception is demonstrated when exemplars of the same category are judged as more perceptually similar than they actually are, whereas stimuli from different categories are judged as more perceptually different than they actually are. This stringent test of categorical perception has been applied to the study of facial expressions across a range of experiments (Bimler & Kirkland, 2001; Calder, Young, Perrett, Etcoff, & Rowland, 1996; Etcoff & Magee, 1992; Young, Rowland, Calder, Etcoff, Seth & Perrett, 1997). In these experiments, expression continua are generated by morphing between two different expressions. Pairs of stimuli are then selected from these morphed continua. These pairs can display the same physical expression, can have a different expression but belong to the same emotion category, or have a different expression from different emotion categories. Participants are then asked to determine whether the expression is the same or different across the pairs of stimuli. These studies demonstrate categorical perception of expression as they find an increase in accuracy for discriminating pairs of expressions from different emotion categories compared to pairs of stimuli from within the same emotion.

A purely categorical model of perception of facial expressions would, however, be unable to account for the ability to discriminate differences in the intensity of the same expression category. For example we can detect changes in expression that go from slightly happy to very happy (Calder et al., 1996). This ability is evident in studies of categorical perception where although participant’s performance decreases for within compared to between category discriminations, the ability to discriminate within-category changes is still above chance (Young et al, 1997).

1.4.2.2 Continuous models of facial expression perception

The ability to detect changes in the intensity of expression can be readily accounted for by continuous models of facial expression. Furthermore, continuous models such as that suggested by Woodworth and Schlosberg (1954) and Russell (1980) are able to explain and predict the systematic confusions that occur when labelling facial expressions (Bruce & Young, 2012). In devising their model, Woodworth and Schlosberg looked at studies in which participants had to spontaneously label emotional expressions. Although participants differed in their responses they did so in a non-random way. Woodworth and Schlosberg grouped the expression labels that were used interchangeably and arranged these groups into a circle in which the expressions that were most often confused were placed next to each other; for example surprise was positioned next to fear, and anger neighboured disgust. They identified two orthogonal axes (pleasant-unpleasant and attention-rejection) that could best describe the position of an expression around the circle. Recognition of an expression is then dependent on coding the expression as falling at some point along these two dimensions. This can account for our ability to decode the intensity of the expression; less intense expressions will fall towards the middle of the circle and more intense will be positioned along the extremes.

Continuous models can also account for the systematic variations in how different expressions are posed. Rozin, Lowery, and Ebert (1994) demonstrated that different types of events that elicit disgust are associated with different disgust facial expressions; offensive smells were associated with a nose wrinkle whereas a protruding tongue was a characteristic of offensive foods. Continuous accounts would suggest that these variants of disgust fall at similar points along the aforementioned dimensions and are thus perceived as disgust.

Finally the flexibility of continuous models can account for the influence of contextual information on the interpretation of facial expression. Context is often encoded with the face and aids the interpretation of the expression (Barrett & Kensinger, 2010; Bruce & Young, 2012). Aviezer et al., (2008) demonstrated the importance of context in judging facial expressions; participants would incorrectly interpret facial expressions of disgust when shown in conjunction with body information that cues another emotion. Russell and Fehr (1987) demonstrated participants would reinterpret the perception of expression based on the contextual information provided. Rigid categorical models in which information is automatically decoded from facial expressions are unable to explain this contextual effect.

In summary, there is evidence consistent with both categorical and continuous models of facial expression perception. Categorical perception of expression has been demonstrated and participants are highly consistent at labelling basic emotions. Despite this, the contextual information, the systematic confusions in expression recognition and the ability to judge the intensity of expression lends support to continuous accounts. Rather than having a single categorical or continuous perception of expression, the brain may draw on both types of representation and depending on what information is required a categorical or a continuous representation may be most appropriate. Whether categorical or continuous representations underlie the neural representation of expression is explicitly addressed in Chapter 3, Experiment 2.

1.4.3 Neural regions involved in the processing of facial expressions

The STS and amygdala are two neural regions implicated in the processing of facial expressions (Haxby et al., 2000). Transient changes in facial musculature provide signals important in social communication as well as biologically relevant information such as potential physical threats in the environment. Little is known regarding the relative roles of the STS and amygdala in extracting these cues: are both regions equally involved in representing the social and biological aspects of facial expressions or do these regions contribute differently to the representations of expression? The involvement of the STS and amygdala in processing facial expression is briefly outlined. A further region, the FFA, is also considered here, although this region is suggested to be predominantly involved in processing the invariant features of a face (Haxby et al., 2000), more recent research suggests the FFA might also be sensitive to the changeable aspects of a face.

1.4.3.1 The STS and social communication

A neural region sensitive to socially meaningful patterns of facial movement is the STS (Bruce & Young, 2012; Haxby et al., 2000). The role of the STS in face perception has been relatively under-explored compared to the involvement of the FFA in facial identity. A possible explanation for this may be due to the difficulty in identifying the STS from a functional localiser scan; the STS is less reliably identified than other face-selective regions (Kanwisher & Barton, 2011).

Neural sensitivity to facial expression is demonstrated in neuroimaging experiments by comparing the response to emotional facial expressions with that to neutral expressions. This comparison allows the dissociation of responses specific to the emotional content of the face rather than a general response to the presence of a face. This type of comparison yields a greater response in the STS to emotional facial expressions (Engell & Haxby, 2007; Kesler-West et al., 2001; Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001). The importance of the STS in decoding facial expressions is further demonstrated when considering the pattern of response across this region to different facial expressions. Using high-resolution fMRI with multi-voxel pattern analysis (MVPA) Said, Moore, Engell, Todorov, and Haxby (2010) demonstrated that the pattern of response across the STS was able to discriminate between six different facial expressions (although, this could not discriminate a representation of sadness). Furthermore they found that the similarity structure of the pattern of responses correlated with participants' similarity ratings of the expressions.

Facial expressions, of course, are not the only socially relevant information that can be extracted from a face. Eye gaze is a dynamic property of a face that can indicate a person's attention and intentions and is hugely important in social interactions (Allison, Puce, & McCarthy, 2000). The sensitivity of the STS region to eye gaze has been demonstrated in a range of studies investigating the response in this region to both direct and averted gaze (Hoffman & Haxby, 2000). Furthermore, the STS has neurons sensitive to different gaze directions (Calder et al., 2007; Pelphrey, Viola, & McCarthy, 2004) and these neural representations are distinct but overlap those coding for facial expression (Engel & Haxby, 2007). Rather than simple visual analysis of eye gaze, the STS, appears tuned to the social relevance of this information (Mosconi, Mack, McCarthy, & Pelphrey, 2005; Nummenmaa & Calder, 2009; Pelphrey, Morris, & McCarthy, 2004; Wyk, Hudac, Carter, Sobel, & Pelphrey, 2009).

Cues to social communication are not restricted to a face and can be gained from a variety of modalities. Bodily motion, gestures and vocal intonation provide useful social information and the STS is indeed sensitive to these cues (Ethofer, Pourtois, & Wildgruber, 2006; Grossman & Blake, 2002; Hagan et al., 2009; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Wyk et al., 2009). Importantly, it is not just any motion that drives the response in the STS (Pelphrey, Singerman, Allison, & McCarthy, 2003) but rather it is the emotional information that these modalities provide that particularly engages this region (Gallagher & Frith, 2004). The cross-modality response to emotion in the STS was demonstrated comprehensively by Peelen, Atkinson, and Vuilleumier (2010). They presented participants with a variety of emotions displayed by face movements, body movements and vocal intonations. Using MVPA they found emotion-specific patterns of responses across the STS; importantly these patterns were independent of modality. This integration of cross-modality information in the STS occurs at an early stage of processing (Hagan et al., 2009).

Taken together, the STS is a region involved in the multimodal integration of socially relevant information. This region is sensitive to transient changes in a face such as eye gaze and expression. Sensitivity to other modalities has also been demonstrated such as vocal intonations and bodily movements. In particular, the STS seems to extract the social component of these communicative signals.

1.4.3.2 Biological relevance processing in the amygdala

As well as providing socially relevant signals, facial expressions can communicate a wealth of biologically relevant information pertinent to survival; for example fearful facial expressions can indicate potential physical threats in the environment, whereas a disgust expression can inform of potentially harmful substances that should be avoided. The amygdala, situated in the medial temporal lobe, is implicated in the representation of these biologically relevant signals.

Neuroimaging studies investigating facial expression point towards the amygdala as important in interpreting the emotional meaning of expression. In particular, the amygdala has been implicated in detecting fearful expressions as reflected in neuroimaging studies which have shown an increased BOLD response to fear compared to other facial expression such as happy, anger and disgust (Breiter et al., 1996; Morris et al., 1996; Phillips et al., 1998; Whalen et al., 2004). Although the amygdala involvement in expression analysis has been mostly reported for fear, more recent neuroimaging studies have provided support for the role of the amygdala in the appraisal of other facial expressions (Blair, Morris, Frith, Perrett, & Dolan, 1999; Phan, Wager, Taylor, & Liberzon, 2002; Schienle et al., 2002; Sergerie, et al., 2008; Winston, O’Doherty, & Dolan, 2003).

The importance of the amygdala in processing emotion is reflected in studies of patients with amygdala lesions (Adolphs et al., 1999; Adolphs, Tranel, Damasio, & Damasio, 1994, 1995; Broks et al., 1998; Calder et al., 1996). SM, a woman with bilateral amygdala damage, was asked to rate the intensity of different facial expressions which revealed impairments in interpreting fearful expressions (Adolphs et al., 1994). However, impairments in interpreting facial expression after amygdala damage are not restricted to fear (Graham, Devinsky, & Labar, 2007; Sato et al., 2002). Calder et al., (1996) report two patients with amygdala lesions with impaired recognition of facial expression of fear and anger but their discrimination of happiness, sadness were comparable to controls.

The appraisal of facial expression in the amygdala is thought to reflect its role in detecting threat and ambiguity in the environment. In their review Sander, Grafman, and Zalla (2003) make clear that the amygdala should not be considered a sub-system dedicated only to the evaluation of negative emotions. Instead they see the response in the amygdala as an emergent property of its role in detecting salient stimuli pertinent to survival. From this perspective, fearful expressions are ambiguous and require more information to be gathered in order to make appropriate responses to impending danger, hence reflecting the role of the amygdala in detecting ambiguity. Interestingly, impairments in detecting negative emotions after amygdala lesions are often accompanied by attenuated responses to threat stimuli (Broks et al., 1998; Feinstein, Adolphs, Damasio, & Tranel, 2011; Sprengelmeyer et al., 1999); demonstrating the role of the amygdala in interpreting signals pertinent to survival. This is reflected in a neuroimaging study in which participants were played movies of faces depicting expressions of fear and happiness. In order to manipulate the emotional content of the stimuli the movies were played both forward and backward; movies played in a forward direction increase the intensity of the expression whereas the opposite is true for movies played in reverse. The response in the left amygdala was greater to both happiness and fear movies that were played in a forwards direction than backwards. Furthermore there was a correlation between the intensity of the expression experienced and the response in the amygdala (Sato, Kochiyama, & Yoshikawa, 2010). This provides strong evidence that rather than performing visual analysis of facial expressions, the amygdala is sensitive to the content relevance of the expression.

The processing of biological relevant information in the amygdala is highlighted by its response to other facial characteristics that signal potential threats. Neuropsychological evidence has demonstrated the importance of the amygdala in judging trustworthiness from faces; patients with bilateral amygdala damage judge faces as more trustworthy than controls (Adolphs, Tranel, & Damasio, 1998). The BOLD response in the amygdala also correlates with trustworthiness; with a greater response to faces rated as untrustworthy (Singer, Kiebel, Winston, Dolan, & Frith, 2004; Winston, Strange, O’Doherty, & Dolan, 2002; but see Said, Baron, & Todorov, 2008; Todorov & Engell, 2008). The apparent critical involvement of the amygdala in judgements of trustworthiness reflects a wider role in detecting biologically relevant information.

Biologically relevant information is not only provided by facial expressions but can also be extracted from eye gaze. Eye gaze can signal the intentions of others; for example a mutual gaze may indicate threat whereas an averted gaze may indicate submission. The response in the amygdala has been shown to be sensitive to both direct and averted gaze (Kawashima et al., 1999) and lesions to this region impair the ability to use information from the eye region effectively (Adolphs et al., 2005). It is this emotional component of eye gaze that the amygdala may be particularly sensitive to. Hooker et al., (2003) found modulation of the amygdala in a task which involved participants having to detect when eyes gazed directly at them, compared to a task where eye-gaze was used to determine directional-information. They conclude that this reflects the importance of the amygdala in extracting the emotional or biological relevance component of eye gaze.

Dynamic changes in a face are not the only signals for communicating threats, ambiguity and the need for increased vigilance. Other cues, such as vocal expressions and bodily postures also provide information regarding impending danger. Scott et al., (1997) report the case of D.R. who after amygdala damage was impaired at perceiving intonation patters critical for vocal affect recognition. A further patient with bilateral amygdala damage, N.M., was unable to accurately recognise fear from emotional sounds or bodily postures (Sprengelmeyer et al., 1999).

To summarise, the amygdala is a region involved in detecting and processing highly salient biologically relevant stimuli that signal ambiguity and threats within the environment. One way of communicating this information is through facial expressions. It is not, however, the expression that necessarily drives the response in the amygdala, rather it is the biological relevance that is of importance to this region. The response in the amygdala is not restricted to facial expression or even faces but is also sensitive to eye gaze, vocal intonations and bodily postures, which can also signal threats within the environment.

1.4.3.3 The representation of transient signals in the FFA

Predominantly the FFA is a region implicated in processing the invariant features of a face (Grill-Spector et al., 2004; Rotshtein, Henson, Treves, Driver, & Dolan, 2005). Recent neuroimaging studies, however, have demonstrated sensitivity of this region to facial expressions (Cohen Kadosh, Henson, Cohen Kadosh, Johnson, & Dick, 2010; Narumoto et al., 2001) and the response in this region appears to be modulated by both changes in expressions as well as identity (Fox et al., 2009; Vuilleumier, Armony, Driver, & Dolan, 2001). Using direct causal modelling (DCM) to measure the connectivity between regions, Fairhall and Ishai (2007) found that both facial expressions and identity modulated the connectivity between the inferior occipital gyri and the fusiform gyrus, suggesting the involvement of the fusiform gyrus in expression processing. In an fMRI experiment Ganel, Valyear, Goshen-Gottstein, and Goodale (2005) informed participants to attend to either the identity or the expression of faces. The response in the FFA was modulated when participants attended to expression. There was also an increase in response when participants directed attention to identity and expression varied. Finally, under passive viewing of faces, the FFA showed a greater response to changes in expression compared to constant expressions. The authors conclude that the expression may be computed relative to each individual's face and the FFA may extract the information needed to do this.

1.5 Thesis aims

The overall objective of this thesis is to explore the neural representation of facial expression. The main aims of the thesis are: (1) to directly compare the neural representation of facial expression in the STS and amygdala. (2) To investigate whether expression sensitive neural regions hold categorical or continuous representations of facial expression. (3) To investigate the relative contributions of shape and surface-based information in the representation of facial expressions. (4) To investigate the influence of identity on the representation of the changeable aspects of faces. The aims are specifically addressed across the following four Chapters:

- **Chapter 2** – Face stimuli for the experiments reported in this thesis were derived from the Ekman and Friesen series. These faces were carefully selected and manipulated to produce well controlled stimuli for each experiment. Chapter 2 details the criteria for stimuli selection and the manipulations used to generate the faces for each experiment. This chapter also reports methodological parameters common across all fMRI experiments.
- **Chapter 3** – This chapter explores the nature of expression representations in the brain. The experiments in this chapter first identify neural regions that are sensitive to changes in facial expression. Next, using morphs between different expressions, categorical and continuous neural expression codes are directly compared. In a final experiment, the neural representation of expression is investigated using more ecologically valid dynamic stimuli.
- **Chapter 4** – having explored the neural representation of facial expression in the previous chapter, this chapter sought to identify what facial information underpins the neural coding of expression. Specifically, is the representation of facial expression based predominantly on shape or surface based information? Photographic negation was used to disentangle the contribution of these two types of information by disrupting the surface-based cues whilst largely preserving the shape-based information.
- **Chapter 5** – This chapter investigates the influence of identity information on the representation of expression. In order to be socially meaningful expression needs to be extracted across the same individual. The experiments in this chapter investigate the neural responses to expression with same and different identity faces.

Chapter 2

Stimuli and general methods

2.1 Stimuli selection and manipulation

Previous studies investigating emotion processing have often assumed that facial expressions for a specific emotion are posed in the same way. Work by Ekman and Friesen (1978), however, demonstrates a degree of variability in the patterns of muscle movements that can be used to pose the same emotion. With this in mind, great care was to taken to ensure that the stimuli selected for the reported experiments showed consistency in their facial expressions for each emotion category; not just in their visual representation of the expression but also in the underlying muscles used to pose each emotion category.

The stimuli for Experiments 1-5 were derived from faces in the Ekman and Friesen (1976) Pictures of Facial Affect (POFA) series, carefully selected from the Young et al., (2002) FEEST set. The same five actors and expressions were used in Experiments 1-5. This section outlines the criteria used to select these face images and how they were manipulated to generate the final stimuli for each experiment.

One of the criteria implemented to select stimuli was the muscle groups used to pose each expression; this was based on the Facial Action Coding System (FACS) developed by Ekman and Friesen (1978). First this section provides a brief outline of the FACS. Next, the POFA and FEEST sets are described. Finally the image manipulations used to generate the stimuli needed to investigate specific hypothesis of each experiment, such as morphing and averaging, are described.

2.1.1 The Facial Action Coding System

The Facial Action Coding System was developed by Ekman and Friesen (1978) to provide an objective and comprehensive system to distinguish all possible visual facial movements. Rather than simple descriptions of facial feature movements, Ekman and Friesen employed an anatomical approach to coding facial movements by investigating the underlying muscles used to move the face. Defining facial expressions by the muscles involved overcomes the problems of describing facial movements in relation to the landmarks of the face, in which small individual variations in facial landmarks may lead to uninformative variations in the movement of the face. As individual muscles can be combined to produce a single facial appearance, or divided depending on which part of the muscles moved, the FACS refers to measurements as action units rather than the specific muscles that moved.

Ekman and Friesen developed FACS by learning to fire each individual facial muscle in their own faces and then photographing the resulting movement. The resulting photographs were then randomly reordered and examined to determine whether the separate muscles that had been fired could be distinguished from appearance alone. They coded facial movements by considering single action units and the possible combinations of action units that were used for specific facial movements. Ekman and Friesen described 33 action units that could reliably describe all possible facial movements. Most action units involve a single muscle but some involve a combination of two or three muscles. Actions units offer a useful and objective way to ensure that the face images selected for this thesis had the highest possible consistency in their facial expressions for each emotion.

2.1.2 The Ekman and Friesen POFA and the FEEST set

The stimuli for Experiments 1-5 were faces from the Ekman and Friesen POFA (1976) selected from the FEEST set (2002). The POFA were developed to provide photographs that could be used in cross cultural studies of facial expression recognition. Ekman and Friesen generated hundreds of photographs of facial affect by asking actors to move specific muscles thought to underlie each emotion (Ekman and Friesen were developing the FACS and the POFA concurrently and muscles identified in the FACS were used to inform the muscles moved in the POFA). The only spontaneous expression used in the POFA was happiness. From this database of photographs Ekman and Friesen selected the final images to be included in the POFA based on two criteria. Firstly, they selected the photographs judged to be the best representation of the expression based on the muscle movements used. A second criterion was to use photographs showing the best perceptual representation of the expressions. These photographs were identified by two recognition experiments; one a forced choice paradigm, the other using rating scales. A total of 110 photos were included in the final POFA and all photographs achieved above 70 % accuracy in the recognition experiments (all but 11 photographs achieved above 80 % accuracy). The images from this set have become the most widely used and validated set of images in face perception research (Young et al., 2002).

The POFA was the database from which images were selected for the FEEST set (Young et al., 2002). From the POFA 70 images were selected (10 actors, six expressions, and one neutral face for each actor). The main selection criteria for the FEEST set was the inclusion of actors for which all six expressions and a neutral pose were available.

2.1.3 Selection of face stimuli

For the experiments reported in this thesis (Experiments 1-5) five actors posing five basic expressions (anger, disgust, fear, happiness and sadness) were selected from the FEEST set. These five expressions were selected as they are thought to be universally recognised (Ekman 1972; see Chapter 1.4.1.1). Surprise was not included due its confusability with fear and the debate as to whether it should be considered a basic emotion (Bruce & Young, 2012). The display of facial expressions of specific emotion categories can vary between different actors (Ekman, 1972), so that when different facial expressions of the same emotion are seen in succession, it is possible that these variations in expression could give rise to the perception of a different emotion category. Therefore, in order to minimise variations in the expression of each emotion across actors, great care was taken when selecting stimuli. This was achieved based on the following three selection criteria:

1. *Visual similarity of the posed expression*. Although universally recognised, there are variations in how each expression can be posed (Ekman, 1972). This is seen, for example, in the expression of happiness which can be posed with either an open or closed mouth. The selected actors showed the greatest possible similarity of their visual representation of each emotion category.

2. *Action units.* The FEEST set includes Ekman's own data on the underlying action units that each actor used to pose each expression (as coded by FACS). To ensure the visual similarity of expressions was not based purely on the consistency of anatomical landmarks the muscle groups used to pose expressions were used as a second selection criterion. As the face provides biologically relevant signals it is likely the human observer is sensitive to the muscles used to pose expressions (Darwin, 1872). The five individuals were selected as they demonstrated the greatest possible consistency in the action units used to pose each emotion. The action units used are summarised in Table B.1 (Appendix).
3. *Recognition rates.* Actors with highly recognisable expressions were selected. The FEEST includes recognition rates for the Ekman and Friesen faces in a 6 AFC experiment. The mean recognition rate for the selected actors and their expressions was 93 %. A full breakdown of the recognition rates for the 5 selected actors and their five expressions is detailed in the Table B.2 (Appendix).

The five actors that were selected based on these three criteria can be seen in Figure 2.1.

A behavioural experiment was conducted to validate the recognisability of each of the selected facial expressions as well as the presentation time needed to recognise the expressions. Ten participants were presented with a face followed by a blank screen, during which they had to indicate the expression they thought was presented (5 AFC). Faces were presented for three different time periods (500, 1000 or 1500 ms). Each face was presented twice at each presentation time. The results from this experiment demonstrate that faces were accurately recognised at each time period: mean recognition rate at 500 ms was 90 %, at 1000 ms was 94 % and at 1500 ms was 93 %. A full break-down of recognition rates for each actor and each expression can be seen in Table B.3-B.5 (Appendix). A 5 x 5 x 3 ANOVA with the factors Expression (anger, disgust, fear, happy, sadness), Actor (F5, F6, F8, M1, M6) and Presentation Time (500, 1000 and 1500 ms) found no significant interaction between Expression x Actor x Presentation Time ($F_{(32,608)} = 1.15$, $p = 0.26$). Based on these results a presentation time of approximately 1000 ms was used in subsequent experiments.

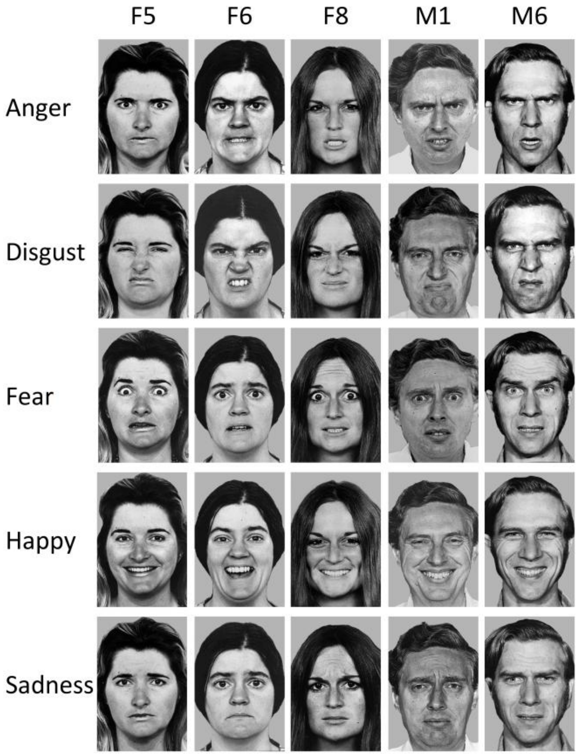

Figure 2.1 *Initial stimuli for Experiments 1-5. Five individuals posing five expressions were selected from the FEEST set. Each row shows the facial expression of one emotion by the five selected actors.*

2.1.4 Stimuli for Experiment 1

The five individuals and five expressions displayed in Figure 2.1 were used in Experiment 1. The prototype images from the FEEST set were used.

2.1.5 Stimuli for Experiment 2

Experiment 2 aimed to investigate categorical and continuous representations of expression in the brain. In order to achieve this aim, expression continua that varied from one expression to another in equally-graded steps were required. The stimuli for this experiment were derived from the prototype faces selected from the FEEST set and used in Experiment 1.

Although the initial stimuli were carefully selected based on stringent criteria, small variations in the facial expressions were still present between actors. In order to remove these variations and produce completely consistent expressions PsychoMorph (Tiddeman, Burt, & Perrett, 2001) was used to generate an average expression for each emotion category. In PsychoMorph, the first step involves delineating the locations of key facial features and contours of each face by specifying 178 fiducial points. Using these fiducial points a mesh of triangle-shaped regions (tessellations) of colour and brightness is overlaid. The average shape of a given emotion expression can then be generated for each emotion category by the program's calculating the average position of each fiducial point across the five actors. The shapes of the tessellations in each constituent image are then deformed and blended together to match the average shape of the fiducial points. This gives the average shape of all the faces that make up the image. The original texture from each individual actor’s face was then applied to the average face shape to produce five distinct identities with the same (i.e. equivalently shaped) facial expression (see Figure 2.2) (Tiddeman et al., 2001). This averaging method was conducted for all five emotion categories. The procedure ensured that differences between images were tightly dependent on changes in identity for each emotion category, rather than any variability between examples of the same expression.

The expression continua were then generated for this experiment by taking two of the averaged expressions from one actor and morphing between them using PsychoMorph. The morphing procedure creates expression continua by manipulating, in evenly-graded steps, the fiducial points present in one image towards the fiducial points specified in another image. In this way expression continua that run between any two expressions can be generated. To ensure that every image in the expression continua was a morph, the continua started from the 99 % image and from this image continua were generated in 11 % steps. An example of a happiness to disgust expression continuum for actor F8 is shown in Figure 2.3. Expression continua were generated for every possible combination of expressions, leading to 10 expression continua for each actor. The four most effective expression continua (there were two actors for each expression continuum) were selected for use in Experiment 2. These continua were selected based on two factors: (1) visual effectiveness – morphing between certain expressions caused artefacts that degraded the quality of the images and these were therefore removed from the stimulus set. (2) Participants' accuracy on a recognition experiment (see Experiment 2, Chapter 3.3.2.1 for details). Continua with the most accurately recognised expressions were selected. The selected expression continua were: fear to happy, happy to disgust, disgust to fear, and disgust to sad.

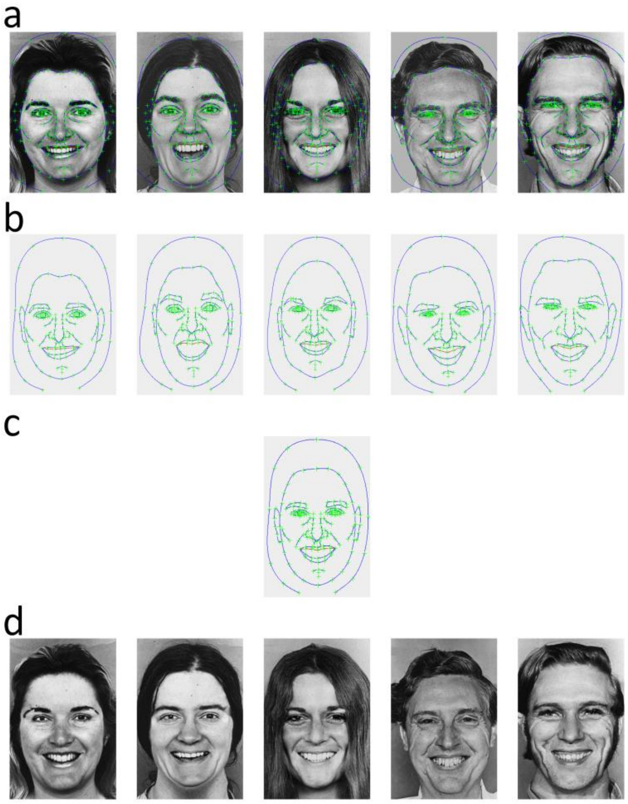

Figure 2.2 *Outline of the method used to generate average expressions. a. The five selected actors' initial happiness expression; facial features have been delineated by 178 fiducial points. b. Texture free representation of the fiducial points used to delineate the facial features. c. Average shape of the five actors' happiness expressions. d. Each individual texture is overlaid onto the average happiness shape to produce five distinct identities with equivalently shaped expressions. This process was repeated for all five emotions.*

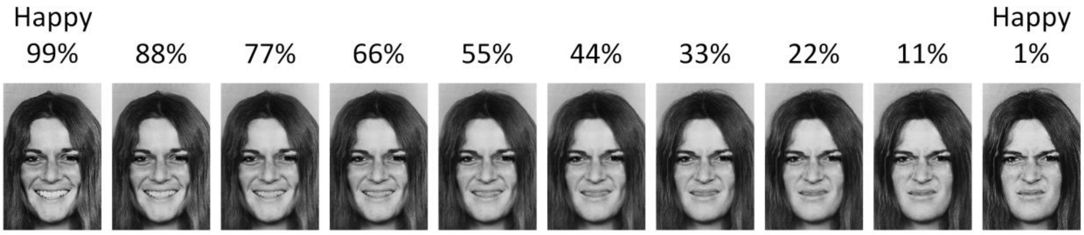

Figure 2.3 *Example expression continuum. Happiness to disgust for F8.*

To verify that each expression continuum involved an evenly-graded linear transition a measure of image change across each continuum was calculated. Each image along the continuum was divided into its constituent pixels and each pixel was assigned a value based on brightness. Using these values the differences between each pixel in one image and the corresponding pixel in another image were determined. These differences were then averaged across the image to provide a mean difference between two images. In this way the image difference between the 99 % image and each image along the expression continuum was compared. These image statistics were calculated for the four expression continua used in the fMRI experiment. The differences between the 99 % and each image were averaged across the different expression continua and are shown in Figure 2.4a. The image differences increased linearly across the expression continua ($r^2 = 0.97$). A second measure of image variability, correlation between pixels, was also calculated. The correlation between images was computed in the same way as the image differences and is shown in Figure 2.4b. A linear correlation was found across the morphs, with a decreasing correlation between images further along the expression continua ($r^2 = 0.94$).

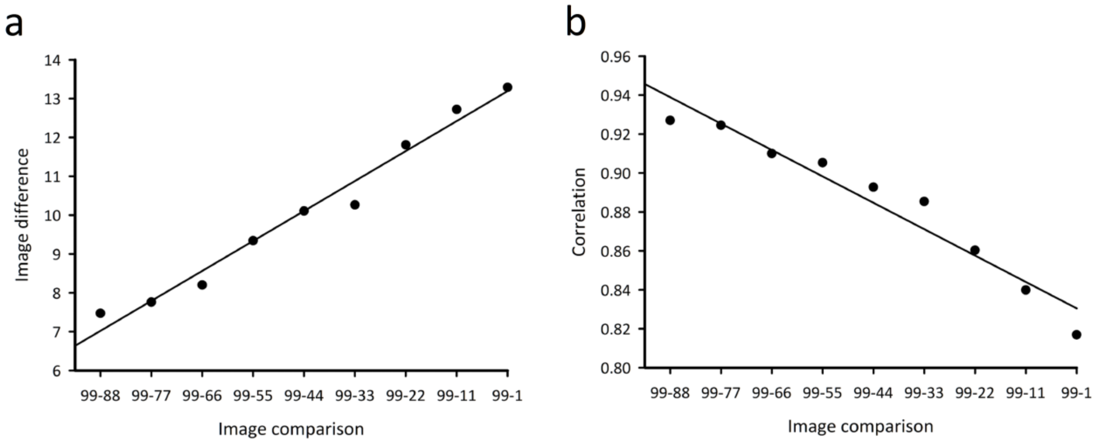

Figure 2.4 *Linearity of expression continua. a. Mean image difference between the 99 % image and each image along the expression continua averaged across the expression continua used in Experiment 2. b. Mean correlations between the 99 % image and each image along the continua averaged across the expression continua used in Experiment 2.*

2.1.6 Stimuli for Experiment 3

Experiments 1 and 2 investigate the representation of expression using static changes in expression. Humans, however, are sensitive to the temporal properties of changes in expression (Ambadar, Schooler, & Cohn, 2005; Kamachi et al., 2001; Wehrle, Kaiser, Schmidt, & Scherer, 2000), so Experiment 3 investigated the representation of expression using more ecologically valid dynamic stimuli. The stimuli for this experiment were movies consisting of a dynamic change in the intensity of the expression from a neutral to a prototype expression. The five expressions and five actors used in the previous experiments were selected for this experiment. The frames for the movies were generated by morphing between each individual's neutral expression and each of their prototype expressions in 5 % steps using PsychoMorph (Tiddeman et al., 2001) and the morphing procedure described above. Movies were generated by playing the morphed images in sequence using Adobe Premiere Pro. The first (neutral) frame was played for 160 ms and the final frame (prototype expression) was played for 280 ms. The 18 intermediate frames were each played for 40 ms (Figure 2.5)

Figure 2.5. *Example of a movie stimulus. Frames that constituted F8 happiness movie.*

Validation of the movie stimuli was demonstrated in an expression-classification experiment, in which recognition rates of the dynamic expressions were compared to the recognition rate for the equivalent original prototype expression. This was thought necessary as in the dynamic stimuli the full prototype expression is available for less time than the static stimuli and may, therefore, prove more difficult to recognise. Participants either classified the static or dynamic expressions in a 5AFC task. 20 participants (11 female; mean age 29) rated the static expressions and 20 participants (12 female; mean age 27) rated the dynamic expressions. Both the dynamic and static expressions were presented for 1160 ms followed by a 2 s gray screen, during which participants could make their response. This experiment found that recognition accuracy for the static expression was 83.6 % and for the dynamic expressions 84.3 %. A full break-down of the accuracy for each expression can be seen in Table B.6 and B.7 (Appendix). A 2 x 5 repeated measures ANOVA with the factors Stimulus Type (static, dynamic) and Expression (anger, disgust, fear, happiness, sadness) revealed a significant interaction between Stimulus Type and Expression ($F_{(4,152)} = 3.44$, $p = 0.01$). There was also a significant main effect of Expression ($F_{(4,152)} = 30.31$, $p < 0.0001$), however there was no main effect of Stimulus Type ($F_{(1,38)} = 0.13$). The significant interaction was driven by a significantly greater accuracy to dynamic anger compared to static anger ($t_{(38)} = 20.1$, $p = 0.005$) and a significantly greater accuracy for static disgust compared to dynamic disgust ($t_{(38)} = 2.97$, $p = 0.05$). The results suggest that the static and dynamic stimuli were equally effective overall at conveying the emotion category.

2.1.7 Stimuli for Experiments 4 and 5

This experiment aimed to investigate the relative contribution of shape-based and surface-based information in the representation of facial expression. Photo negation was used to dissociate surface-based cues from shape-based cues. Photo negation reverses the gray-level relationships within the image, making the white areas black and the black areas white. This manipulation affects the surface-based information, whilst preserving much of the shape-based information carried by the edges of facial features (Benton, 2009).

Experiment 4 was comprised of two behavioural experiments and Experiment 5 involved the use of fMRI. The five actors and expressions selected from the FEEST set were the initial stimuli for these experiments. To generate the photo negative images the contrast-relationships were reversed using Photoshop. Images were cropped using Photoshop to remove the external features of the face. Images were cropped to an ellipse shape which was kept constant within gender (male actors were only paired with males and females with females). The ellipse shape was held constant across actors to ensure participants could not make judgements of identity purely on the external contour of the ellipse. Examples of the stimuli can be seen in Figure 2.6. F8 was removed from the cropped image stimulus set as her hair could still be seen once images were cropped and this could have been used as a potential cue to her identity.

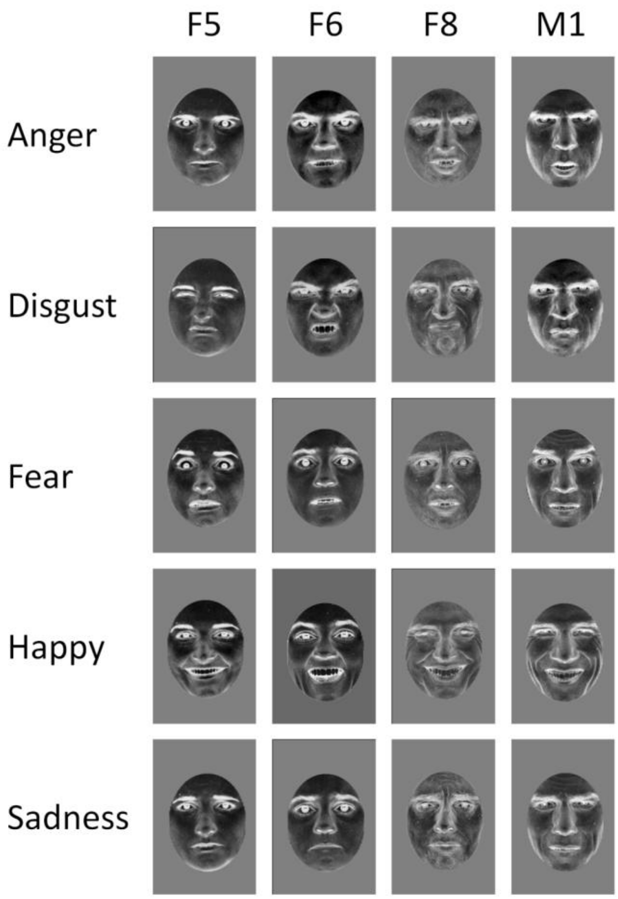

Figure 2.6 *Contrast reversed stimuli used in Experiments 4 and 5. The original stimuli were the five expressions and actors selected for Experiments 1-3.*

2.1.8 Stimuli for Experiments 6 and 7

Experiments 6 and 7 involved the analysis of localiser data collected across numerous experiments. Details on the stimuli for the localiser are reported below in section 2.2.3.1 and 2.2.3.2

2.2 fMRI Methods

To investigate the neural representation of facial expression, fMRI was recruited across six experiments. In Experiments 1-3 and 5 a region of interest approach with functional localisation of regions was implemented. First, this section briefly outlines this approach. Next, the stimuli and procedure for two localiser scans used throughout this thesis are detailed. Finally, the fMRI parameters for data collection and the procedure for data analysis are described here (the fMRI parameters are consistent across all reported fMRI experiments).

2.2.1 fMRI

Functional magnetic resonance imagining was used to investigate the neural representation of facial expression. By measuring changes in the hemodynamic response over time this neuroimaging method creates images based on an indirect measure of neural activity. An increase in neural activity results in a higher metabolic rate, increasing the cells' requirements for glucose and oxygen. The required oxygen is transported to cells bound to haemoglobin. Deoxygenated haemoglobin has magnetic properties detectable in fMRI; changes in the concentration of deoxygenated haemoglobin can therefore provide an indirect measure of neural activity. This is referred to as the blood-oxygen-level dependent (BOLD) response, and changes in BOLD can be measured in fMRI.

Although, in comparison to other techniques such as EEG and MEG, fMRI has good spatial resolution, it is not without limitations. As an indirect measure of neural activity fMRI is associated with poor temporal resolution resulting from the time-lag between neural activity and the replacement of oxygen at those cells. Neural activity (e.g. action potentials) occurs almost immediately after the onset of a stimulus, whereas the replacement of oxygen peaks at approximately six seconds. Although this is an inherent limitation of using an indirect measure, the BOLD response does, however, correlate with action potentials (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001).

2.2.2 ROI approach

In order for the data derived within individuals from fMRI experiments to be useful, the results need to be generalised across individuals. One way to compare individuals would be to conduct a whole brain group analysis on the data. In this approach each individual brain is aligned to a standard space and the responses compared. Although a group analysis of this type allows for the comparison of active regions across individuals, there are two potential problems with this. Firstly, each individual brain is unique; neuroanatomy varies between individual brains and it is often difficult to achieve a perfect registration in which the anatomical landmarks of each individual brain neatly align. A second problem is the number of voxels that are analysed; fMRI allows the measurement of changes in BOLD across the whole brain and this can result in data from tens of thousands of voxels that can be compared. Inherent in a whole-brain analysis, therefore, are multiple comparisons that require strict corrections to prevent false results (Nichols & Hayasaka, 2003). These statistical tests can reduce the power associated with the experimental paradigm and make it increasingly difficult to find an effect.

One way to compare individuals whilst circumventing the problems associated with a group analysis is to implement a region of interest (ROI) approach. Using an ROI approach, activations in specific neural regions identified independently for each individual are compared. This significantly reduces the number of voxels investigated, in turn reducing the number of statistical comparisons that are made and therefore the need for strict corrections. This increases the signal to noise ratio, and thus increases the power of the experimental paradigm. A further advantage of this approach is that by allowing for the identification of ROIs within an individual's brain it avoids the issues associated with attempts to align each individual brain to a standard space.

Two ways of identifying ROIs are most prominent in the literature; the use of anatomical landmarks and functional localisation. The use of anatomical landmarks to identify ROIs is appropriate for clearly defined regions such as the amygdala. Most regions of the brain, however, lack such a clearly defined neuroanatomical architecture; in particular the ventral-temporal cortex has highly variable sulci and gyri that make it difficult to define the same anatomical landmarks across participants. A second problem inherent in anatomical identification of ROIs is that the defined ROI may not be homogenous in its response properties. The superior temporal sulcus, for example, is a large piece of cortex that stretches the length of the temporal lobe. This sulcus includes neurons with potentially different response properties; for example spatially distinct neural representations of expression and eye-gaze have been found along this sulcus (Engel & Haxby, 2007). Anatomical definition of this region, therefore, may include a diverse set of neurons coding different information and the positions of which could vary across individuals.

Functional localisations of ROIs can overcome the problems associated with anatomical identification. By using a separate localiser scan, parts of the cortex with specific response properties can be identified. This technique has been used extensively in face processing. Originally implemented by Kanwisher et al., (1997) they first identified parts of the cortex that responded preferentially to faces and in separate experimental scans tested the response properties of these predefined ROIs. Functionally defined ROIs can be compared across individuals and importantly, across experiments. A region of interest approach was adopted in Experiments 1-3 and 5. ROIs were defined functionally by a separate localiser scan.

2.2.3 Functional localisation

Functional localisers were used in Experiments 1-3 and 5 to identify regions in the brain that respond to faces more than non-face stimuli. The use of a localiser scan that was independent from the main experimental manipulations ensures unbiased localisation of ROIs. Two versions of a face-selective localiser were devised. Version one was implemented to define ROIs in Experiments 1 and 3, version two was implemented in Experiments 2 and 5. Details of the localisers are now outlined.

2.2.3.1 Functional localiser 1 – stimuli and procedure

The localiser used a block design with six different conditions: (1) *same-identity faces*, (2) *different-identity faces*, (3) *bodies*, (4) *inanimate objects*, (5) *places*, and (6) *scrambled images* of the former categories (Figure 2.7). Face images were taken from the Psychological Image Collection at Stirling (PICS; http://pics.psych.stir.ac.uk/). These images varied in viewpoint (frontal, ¾ view, profile) and expression (neutral, happy, speaking) within a block. The changes in viewpoint correspond to changes in gaze direction, which is often signalled in real life by movements of both eyes and head (Bruce & Young, 2012). The face images in each block therefore varied in both expression and gaze direction, but in one face condition the face identity was constant across the images in the block and in the other face condition identity varied across the block. Both male and female faces were used, but gender was held constant within a block.

Body images were taken from a collection at the University of Bangor (http://www.bangor.ac.uk/~pss811/), and contained clothed male and female headless bodies in a variety of postures. Images of places consisted of a variety of unfamiliar indoor scenes, houses and buildings, city scenes and natural landscapes. Stimuli in the object condition consisted of different inanimate objects including tools, ornaments, and furniture. Fourier-scrambled images were created by randomizing the phase of each two-dimensional frequency component in the original image, while keeping the power of the components constant. Scrambled images were generated from the images used in the other stimulus categories.

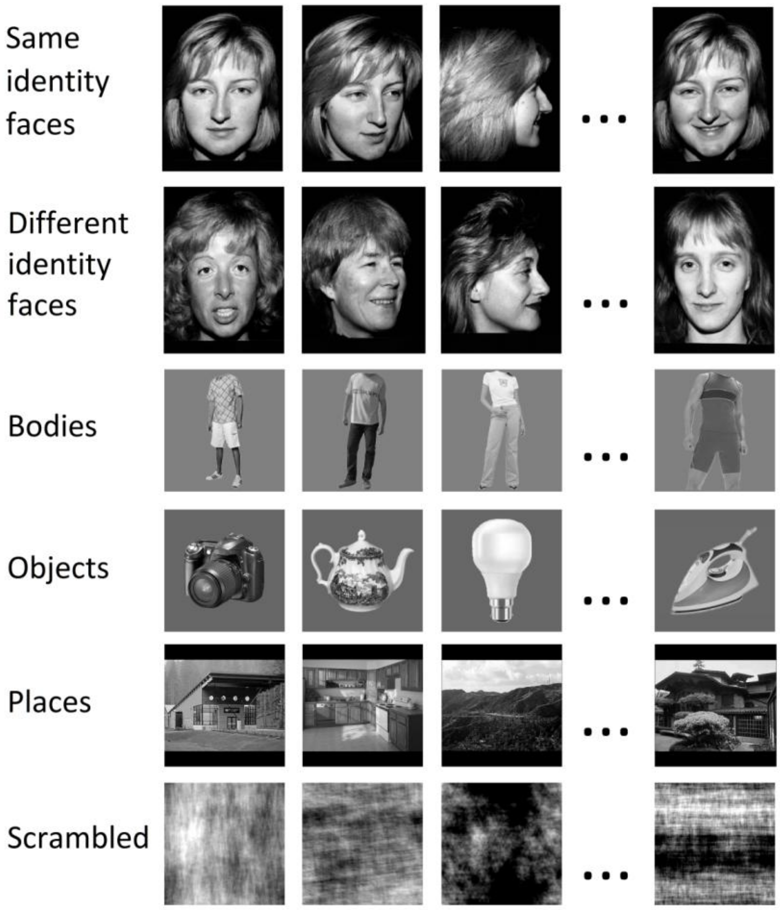

Figure 2.7 *Examples of the conditions in Localiser version 1.*

All images (approx. 8º x 8º) were presented in gray scale and were back-projected onto a screen located inside the bore of the scanner, approximately 57 cm from participants' eyes. Each block consisted of 10 images from a single stimulus condition; each image was presented for 700 ms and followed by a 200 ms blank screen, resulting in a total block length of 9 s. Stimulus blocks were separated by a 9 s gray screen with a central fixation cross. Each condition was repeated four times in a counterbalanced design resulting in a total scan length of 7.2 min. All participants viewed the same sequence of blocks and images.

Participants were required to monitor all images for the presence of a red dot that was superimposed on one or two images in each block. Participants were required to respond, with a button press, as soon as they saw the image containing the target. The target could appear in any location on the image, and was counterbalanced across conditions.

2.2.3.2 Functional localiser 2 – stimuli and procedure

The localiser used a block design with seven different conditions: (1) *same-identity, varying-expression*, (2) *different-identity, varying-expression*, (3) *same-identity, varying-viewpoint*, (4) *different-identity, varying-viewpoint*, (5) *inanimate objects*, (6) *places*, and (7) *scrambled images* of the former categories (Figure 2.8). Face images were taken from the Radboud database (Langner et al., 2010). Blocks varying in viewpoint involved frontal, ¾ and profile views and faces were shown so viewpoint changed in a coherent manner. Blocks varying in expression included the five basic expressions; anger, disgust fear, happiness and sadness. Both male and female faces were used, but gender was held constant within a block. Non-face stimulus conditions were the same as version one of the localiser and detailed above.

All images (approx. 8º x 8º) were presented in gray scale and were back-projected onto a screen located inside the bore of the scanner, approximately 57 cm from participants' eyes. Each block consisted of 5 images from a single stimulus condition; each image was presented for 1 s and followed by a 200 ms blank screen, resulting in a total block length of 6 s. Stimulus blocks were separated by a 9 s gray screen with a central fixation cross. Each condition was repeated five times in a counterbalanced design resulting in a total scan length of 8.75 min. All participants viewed the same sequence of blocks and images.

Participants were required to monitor all images for the presence of a red dot that was superimposed on one image in each block. Participants were required to respond, with a button press, as soon as they saw the image containing the target. The target could appear in any location on the image, and was counterbalanced across conditions.

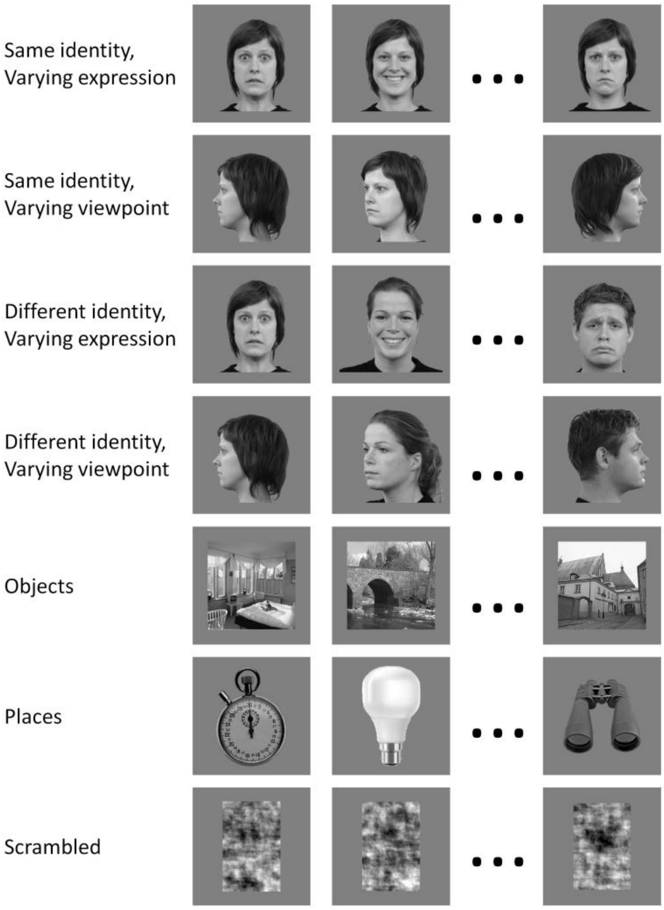

Figure 2.8 *Examples of the conditions used in localiser version 2.*

2.2.4 fMRI analysis

The same procedure for analysing fMRI data was implemented for Experiments 1-3 and 5. Statistical analysis of the fMRI data was performed using FEAT (http://www.fmrib.ox.ac.uk/fsl). The initial 9 s of data from each scan were removed to minimize the effects of magnetic saturation. Motion correction was followed by spatial smoothing (Gaussian, FWHM 6 mm) and temporal high-pass filtering (cutoff, 0.01 Hz). Face-selective regions were individually defined in each individual using the localiser scan by the average of the following two contrasts: (1) *same-identity faces* > non-face stimuli and (2) *different-identity faces* > non-face stimuli (in localiser version two only the face conditions varying in expression, and not viewpoint, were used to define ROIs). Statistical images were thresholded at $p < 0.001$ (uncorrected). In this way, contiguous clusters of voxels located in the inferior fusiform gyrus, in the posterior occipital cortex and in the superior temporal lobe of individual participants could be identified as the FFA, OFA and the STS respectively.

A different approach had to be taken to define the amygdala. Signals in the anterior regions of the temporal lobe are typically noisy, because of the proximity to the ear canals. The lower within-subject signal-to-noise makes it difficult to determine face-selectivity at the level of individual subjects in the amygdala. A face-responsive ROI in the amygdala was therefore defined from the face-selective statistical map at the group level, thresholded at $p < 0.001$ (uncorrected). This ROI in the amygdala was then transformed into the individual MRI space for each participant. The time-course of response in the amygdala ROI was then evaluated for each participant to ensure that it responded more to faces than non-face stimuli. In addition to these functional criteria, we were also able to define the amygdala based on anatomy. Despite the difference in the way that the amygdala was defined, Figure A.1 (Appendix) shows that the face-selective voxels that are located in the amygdala for Experiments 1-3 show a corresponding face-selectivity to the other ROIs. In all other respects, the data were processed in exactly the same way for all ROIs.

The time-course of response for each ROI was then evaluated to ensure that it responded more to faces than non-face stimuli. For each experimental scan, the time series of the filtered MR data from each voxel within a ROI was converted from units of image intensity to percentage signal change. All voxels in a given ROI were then averaged to give a single time series for each ROI in each participant.

Individual stimulus blocks were normalized by subtracting every time point by the zero point for that stimulus block. The normalized data were then averaged to obtain the mean time course for each stimulus condition.

2.2.5 fMRI protocol

The following imaging parameters were used to collect data for all reported experiments. All imaging experiments were performed using a GE 3 tesla HD Excite MRI scanner at York Neuroimaging Centre at the University of York. A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage, radio-frequency coil tuned to 127.4 MHz. A gradient-echo EPI sequence was used to collect data from 38 contiguous axial slices (TR = 3, TE = 25 ms, FOV = 28 x 28 cm, matrix size = 128 x 128, slice thickness = 4 mm). These were co-registered onto a T1-weighted anatomical image (1 x 1 x 1 mm) from each participant. To improve registrations, an additional T1-weighted image was taken in the same plane as the EPI slices.

Chapter 3

The neural representation of facial expression

3.1 Introduction

The ability to visually encode changes in facial musculature that reflect emotional state is essential for effective communication (Ekman & Friesen, 1978). Models of face processing have proposed that expression is either represented as belonging to discrete categories of emotion (Darwin, 1872; Ekman, 1999) or as continuous representations varying along graded dimensions (Russell, 1980; Woodworth & Scholsberg, 1954). Although these models are treated as incompatible, there is evidence to support both approaches (Bruce & Young, 2012).

The categorical perspective is based on the notion that discrete cognitive states underpin a set of basic emotions (Ekman, 1999). Evidence for categorical perception of expression is shown by the consistency with which the basic emotions are recognised (Ekman, 1972) and the greater sensitivity to changes in facial expression which alter the perceived emotion (Calder et al., 1996; Etcoff & Magee, 1992). In contrast, continuous or dimensional models are better able to explain the systematic confusions that occur when labelling facial expressions (Woodworth & Scholsberg, 1954), can account for variation in the way that basic emotions are expressed (Rozin et al., 1994) and the fact that we are readily able to perceive differences in intensity of a given emotional expression (Calder et al., 1996; Young et al., 1997).

The aim of this chapter was to identify regions in the brain that are sensitive to expression and explore how those regions represent facial expression information. Models of face processing propose that the transient signals that give rise to facial expressions are processed largely independently from those important for facial identity (Bruce & Young 2012; Haxby et al., 2000). As such, the neural subcomponents sensitive to expression are thought to be spatially distinct from those sensitive to more invariant features of the face such as identity (Haxby et al., 2000). A route from the occipital face area (OFA) involving the superior temporal sulcus (STS) and amygdala is thought to be sensitive to the changeable aspects of a face. A parallel route from the OFA to the fusiform face area (FFA) and anterior temporal lobe is involved in processing facial identity.

Despite the important roles played by the STS and amygdala in processing facial expression (Morris et al., 1996; Narumoto et al., 2001) little is known about how these regions represent the expression information. This chapter aims to identify face-selective neural regions that are sensitive to expression and investigate how these regions represent the expression information. In Experiment 1, neural regions that were more sensitive to changes in facial expression than changes in facial identity were identified. Experiment 2 aimed to investigate whether these expression sensitive regions hold primarily categorical or continuous representations of expression. Experiment 3 further investigated the neural representation of facial expression using more ecologically valid dynamic stimuli.

3.2 Experiment 1: Sensitivity to facial expression and identity in face-selective neural regions

3.2.1 Introduction

Neuroimaging studies have identified a number of face-selective regions that are involved in the perception of facial expression (Breiter et al., 1996; Fox et al., 2009; Winston, Henson, Fine-Goulden, & Dolan, 2004). The occipital face area (OFA) is thought to be involved in the early perception of facial features and has connections to the superior temporal sulcus (STS). The connection between the OFA and STS is thought to be important in processing dynamic changes in the face, such as changes in expression and gaze, which are important for social interactions (Calder et al., 2007; Engell & Haxby, 2007; Pelphrey, et al., 2004; Puce, Allison, Bentin, Gore, & McCarthy, 1998). Information from the pSTS is then relayed to regions of an extended face processing network including the amygdala for further analysis of facial expression. Although both the STS and amygdala are implicated in the processing of facial expression, very few studies have investigated the response in these regions to facial expression and facial identity within the same experimental paradigm.

This experiment aimed to determine how different regions in the face processing network are involved in the perception of emotion. This experiment compares the response to faces that changed in both facial expression and identity. Regions that are selective to facial expression show a greater response to changes in expression compared to changes in identity.

3.2.2 Methods

3.2.2.1 Subjects

Twenty participants took part in this experiment (15 females; mean age, 23). All participants were right-handed and had normal or corrected-to-normal vision. Visual stimuli (8˚ x 8˚) were back-projected onto a screen located inside the magnetic bore, 57 cm from participants' eyes. All participants provided written consent and the study was given ethical approval by the York Neuroimaging Centre Ethics Committee.

3.2.2.2 Face localiser scan

A separate face localiser scan (Version 1) was used to independently identify regions in each individual’s brain that responded more to faces than non-face stimuli (see Chapter 2.2.3.1 for protocol).

3.2.2.3 Experimental scan

There were four conditions in this experiment: (1) *same-expression, same-identity* (2) *same-expression, different-identity* (3) *different-expression, same-identity* (4) *different-expression, different-identity* (Figure 3.1). Face stimuli were gray-scale Ekman faces selected from the FEEST set (see Chapter 2.1.4). Faces were mounted on gray background and the bridge of the nose was aligned with the fixation cross to prevent images moving around the visual field. Stimuli were presented in blocks, with 5 images per block. Each face was presented for 1100 ms and separated by a gray screen presented for 150 ms. Stimulus blocks were separated by a 9 s fixation gray screen. Each condition was presented 10 times in a counterbalanced order, giving a total of 40 blocks. To ensure participants maintained attention throughout the experiment, participants had to push a button when they detected the presence of a red dot, which was superimposed onto 20 % of the images.

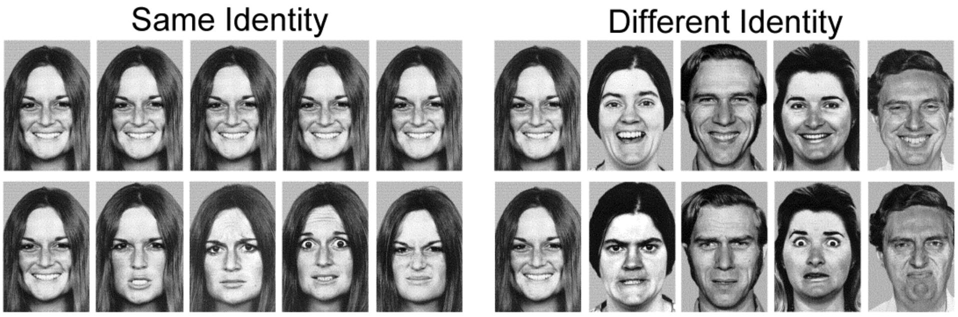

Figure 3.1 *Experiment 1 conditions. Images from the four conditions (Top row: same-expression, same-identity (left), same-expression, different-identity (right). Bottom row: different-expression, same-identity (left), different-expression, different-identity (right).*

3.2.2.4 Imaging parameters and fMRI analysis

Imaging parameters and details regarding the initial processing steps of the fMRI data are reported in Chapter 2.2.4 and 2.2.5.

3.2.3 Results

The Localiser scan identified seven regions of interest that responded more to faces than non-face stimuli. These regions, left and right FFA, left and right OFA, left and right amygdala and the right posterior STS (pSTS), are show in Figure 3.2 and their locations are detailed in Table 3.1. A face-selective part of the left pSTS was only identified in a limited number of subjects and was therefore not included in this analysis.

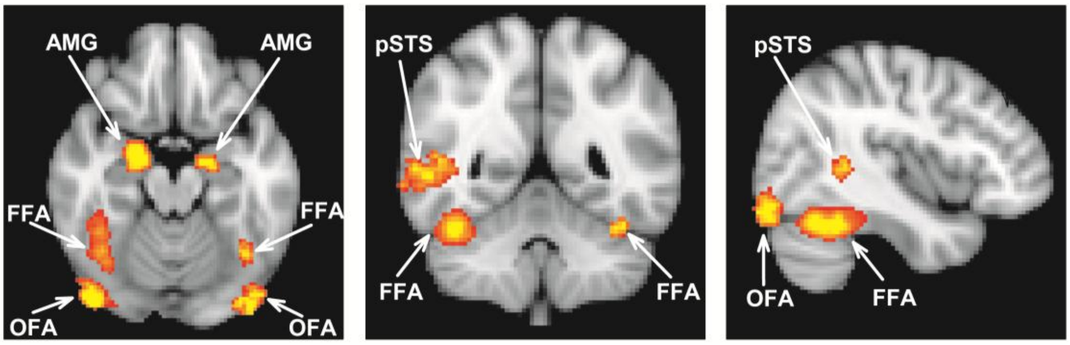

Figure 3.2 *Location of face selective regions in Experiment 1. Average location of each ROI across all participants and transformed into group space. All brain images are depicted in radiologic convention, i.e. coronal and axial slices are left/right reversed.*

A 4 x 3 x 2 ANOVA with the factors Condition (same-expression, same-identity; same-expression, different-identity; different-expression, same-identity; different-expression, different-identity) Region (FFA, OFA, amygdala) and Hemisphere (left, right) was conducted to determine whether the two hemispheres of the same region of interest responded differently. The pSTS was not included in this part of the analysis as it was only identified in the right hemisphere. The results revealed no significant interaction between Condition x Region x Hemisphere ($F_{(6,78)} = 0.39$), nor was there a significant interaction between Condition x Hemisphere ($F_{(3,39)} = 0.71$), nor Region x Hemisphere ($F_{(2,26)} = 1.24$, $p = 0.31$), nor was the a significant main effect of Hemisphere ($F_{(1,13)} = 2.82$, $p = 0.12$). As there was no significant effect of hemisphere the timecourses were averaged across hemispheres resulting in four regions of interest; FFA, OFA, amygdala and right pSTS.

Table 3.1 *MNI coordinates (mm) of face-selective regions in Experiment 1. Coordinates reported are the centre of gravity of each ROI averaged across all participants and transformed into standard space. Standard error is reported in parenthesis.*

Region	*n*	*x*	*y*	z
FFA	20			
L		-41 *(0.8)*	-56 *(2.1)*	-20 *(0.9)*
R		42 *(0.8)*	-54 *(1.2)*	-22 *(1.0)*
OFA	20			
L		-40 *(1.6)*	-85 *(1.4)*	-15 *(1.8)*
R		41 *(1.2)*	-83 *(1.1)*	-12 *(1.1)*
STS	18			
R		51 *(1.6)*	-48 *(2.0)*	5.3 *(1.2)*
Amygdala	16			
L		-19	-9	-15
R		20	-8	-17

The peak responses of the face-selective regions were analysed using a 4 x 2 x 2 ANOVA with Region (pSTS, amygdala, FFA, OFA) Expression (same, different) and Identity (same, different) as the factors. There were significant effects of Expression ($F_{(1,13)} = 4.46$, $p = 0.05$) and Region ($F_{(3,39)} = 48.26$, $p < 0.0001$) but not Identity ($F_{(1,13)} = 2.52$, $p = 0.14$). There was also a significant interaction between Region x Expression ($F_{(3,39)} = 12.73$, $p < 0.0001$). Therefore to investigate which face-selective regions were sensitive to expression the response in each individual ROI is now considered. Note that the differing degrees of freedom in the post-hoc pairwise comparisons for this analysis and other ANOVAs presented in this thesis reflect the way that SPSS handles missing data. Because some ROIs could not be functionally localised in some participants, not all participants provided data for each ROI (Table 3.1).

Figure 3.3 shows the response from the pSTS in this experiment. A 2 x 2 ANOVA with the factors Expression (same, different) and Identity (same, different) revealed a significant effect of Expression ($F_{(1,17)} = 12.84$, $p = 0.002$), but not Identity ($F_{(1,17)} = 1.98$, $p = 0.18$). There was no significant interaction between Expression and Identity ($F_{(1,17)} = 0.04$). The effect of expression was driven by a significantly bigger response to the *different-expression* conditions compared to the *same-expression* conditions in both the *same-identity* conditions ($t_{(17)} = 2.20$, $p = 0.04$) and *different-identity* conditions ($t_{(17)} = 2.75$, $p = 0.01$).

The amygdala revealed a similar pattern of results to that found in the pSTS (Figure 3.3). A 2 x 2 repeated measures ANOVA found a significant main effect of Expression ($F_{(1,15)} = 11.13$, $p = 0.01$) but not Identity ($F_{(1,15)} = 0.09$). There was no significant interaction between Expression and Identity ($F_{(1,15)} = 2.69$, $p = 0.12$). Again, the main effect of expression was driven by the significantly bigger response to *different-expression* compared to *same-expression* in the *same-identity* conditions ($t_{(14)} = 2.18$, $p = 0.05$) and the *different-identity* conditions ($t_{(14)} = 2.23$, $p = 0.04$).

The FFA was sensitive to both changes in expression and identity (Figure 3.3). A 2 x 2 ANOVA revealed significant main effects of Expression ($F_{(1,19)} = 18.06$, $p < 0.0001$) and Identity ($F_{(1,19)} = 4.53$, $p = 0.05$). There was also a significant interaction between Expression and Identity ($F_{(1,19)} = 7.18$, $p = 0.02$). The main of effect of Expression was due to a bigger response to the *different-expression* condition compared to the *same-expression* conditions for *same-identity* faces ($t_{(19)} = 4.39$, $p < 0.0001$). However, there was no significant difference between the *different-expression* and *same-expression* conditions for *different-identity* faces ($t_{(19)} = 0.86$).

The OFA shows a similar pattern of response to that found in FFA (Figure 3.3). There were significant main effects of Expression ($F_{(1,19)} = 12.71$, $p < 0.002$) and Identity ($F_{(1,19)} = 9.91$, $p = 0.01$). There was also a significant interaction between Expression and Identity ($F_{(1,19)} = 8.58$, $p = 0.01$). There was a significantly bigger response to the *different-expression* condition compared to *same-expression* condition for the *same-identity* ($t_{(19)} = 4.24$, $p < 0.0001$) but not for the *different-identity* conditions ($t_{(19)} = 0.16$).

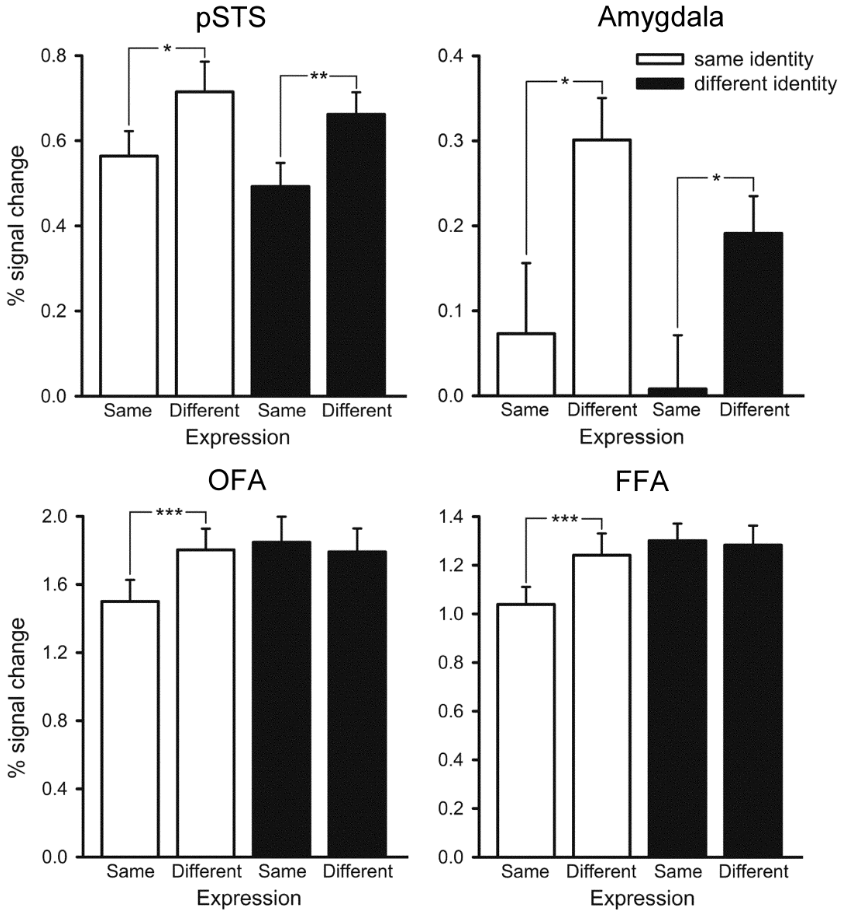

Figure 3.3 *Experiment 1 results. Peak responses to the different conditions in the pSTS, amygdala, OFA and FFA. Error bars represent standard error. * $p < 0.05$, ** $p < 0.01$,*** $p < 0.001$.*

The results from Experiment 1 therefore show selectivity to changes in facial expression (stronger responses to changes in expression than to changes in identity) for pSTS and amygdala.

3.2.4 Discussion

The aim of this study was to identify face-selective neural regions that were more sensitive to changes in facial expression than facial identity. This experiment found that the pSTS and the amygdala were sensitive to faces that changed in expression and that this sensitivity was largely independent of changes in facial identity.

Because of the considerable importance attached to different types of facial information, the most efficient way to analyse this information is thought to involve different neural subcomponents that are optimally tuned for particular types of facial signal (Bruce & Young, 1986, 2012; Haxby et al., 2000). Models of face perception suggest that the analysis of the changeable cues from a face, such as expression, occurs largely independently of the processing of the invariant cues such as identity. This study revealed that the response of the pSTS and amygdala to facial expression was largely independent of changes in facial identity. This is consistent with previous neuroimaging studies that have demonstrated that both the STS and amygdala are sensitive to a range of facial expressions (Engell & Haxby, 2007; Kesler-West et al., 2001; Morris et al., 1996; Narumoto et al., 2001; Sergerie et al., 2008). Although the pSTS has also been shown to be influenced by identity (Andrews & Ewbank, 2004; Fox et al., 2009; Winston et al., 2004), these results suggest that the neural responses in pSTS as well as the amygdala are primarily driven by changeable aspects of the face, such as expression.

In contrast, this experiment found that both the OFA and FFA were sensitive to changes in both expression and identity. The OFA is a region implicated in the structural encoding of the face (Haxby et al., 2000; Rotshtein et al., 2005; see Pitcher, Walsh & Duchaine, 2011); consistent with this, this experiment found that the OFA represented any change in a face. The FFA, however, is thought to be involved in extracting the invariant features of a face that give rise to a person's identity (Grill-Spector et al., 2004; Rotshtein et al., 2005; Yovel & Kanwisher, 2005). The results from this experiment are consistent with other studies which show the FFA is involved in judgements of identity and expression (Cohen Kadosh et al., 2010; Fox et al., 2009; Ganel et al., 2005). It is also possible that these results show that the FFA is sensitive to any structural change in the image. Recent findings have demonstrated that the FFA is sensitive to image variations across the same identity (Davies-Thompson, Gouws, & Andrews, 2009; Xu, Yue, Lescroart, Biederman, & Kim, 2009).

In conclusion, by directly comparing the response to facial expression and identity in the same paradigm, this experiment was able to reveal two regions that were more sensitive expression than facial identity; the pSTS and the amygdala.

3.3 Experiment 2: Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain

3.3.1 Introduction

In Experiment 1 two face selective neural regions, the pSTS and the amygdala, showed sensitivity to changes in facial expression that were largely independent of changes in facial identity. However, how these regions represent the expression information remains relatively unknown. This experiment used morphs between different images of facial expressions to ask whether primarily categorical or continuous representations are used in these different regions of the human brain. To achieve this, the face images used could be physically identical (‘same expression’), could differ in physical properties without crossing the category boundary (‘within-expression change’) or could differ in physical properties and cross the category boundary (‘between-expression change’). Importantly both the within-expression and between-expression conditions involved an equivalent 33 % shift along the morphed continuum. Brain regions that hold a categorical perception of expression should be sensitive to ‘between-expression’ changes in expression, but not ‘within-expression’ changes. However, regions with a continuous representation should be equally sensitive to both ‘between-expression’ and ‘within-expression’ changes.

3.3.2 Validation of expression continua

Continua for the experiment were generated by morphing between different expression images (Chapter 2.1.5). Validation of the morphing procedure was demonstrated in two behavioural experiments. First, an expression-categorisation experiment was conducted to identify which expression continua were most accurately recognised. Although this experiment also provides a measure as to whether the expressions were perceived categorically, a second more stringent test of categorical perception was conducted using a same/different task.

3.3.2.1 Expression-classification experiment

An expression-classification experiment was conducted to determine which of the expression continua were most accurately recognised. The recognition of images from the following expression continua were tested in this experiment: fear-happy, happy-disgust, disgust-sad, disgust-fear, happy-anger, sad-happy, fear-anger (two actors for each continuum). Four images were selected along the appropriate expression continua (99 %, 66 %, 33 %, 1 %) and participants were asked to make a 5 AFC. 26 participants (19 female; mean age 22) took part in this experiment. Face stimuli were presented for 1000 ms followed by a 2 s gray screen during which participants could make their response. Each face image was presented three times resulting in a total experiment length of 8.4 minutes.

This experiment identified four expression continua in which both actors were most accurately recognised; fear-happy, happy-disgust, disgust-fear, disgust-sad. Figure 3.4 shows the results for the four selected expression-continua (averaged across both actors). The results for the other morph continua can be seen in Table B.8 (Appendix). The results clearly demonstrate that for each set of images there was a clear discontinuity in the perception of emotion near the midpoint of the morphed continuum. Participants were more likely to perceive the 99 and 66 % image as belonging to one emotion category and the 33 and 1 % image were perceived as the other emotion along the morph continua. These four continua were selected for the proceeding same/different discrimination task and the fMRI experiment.

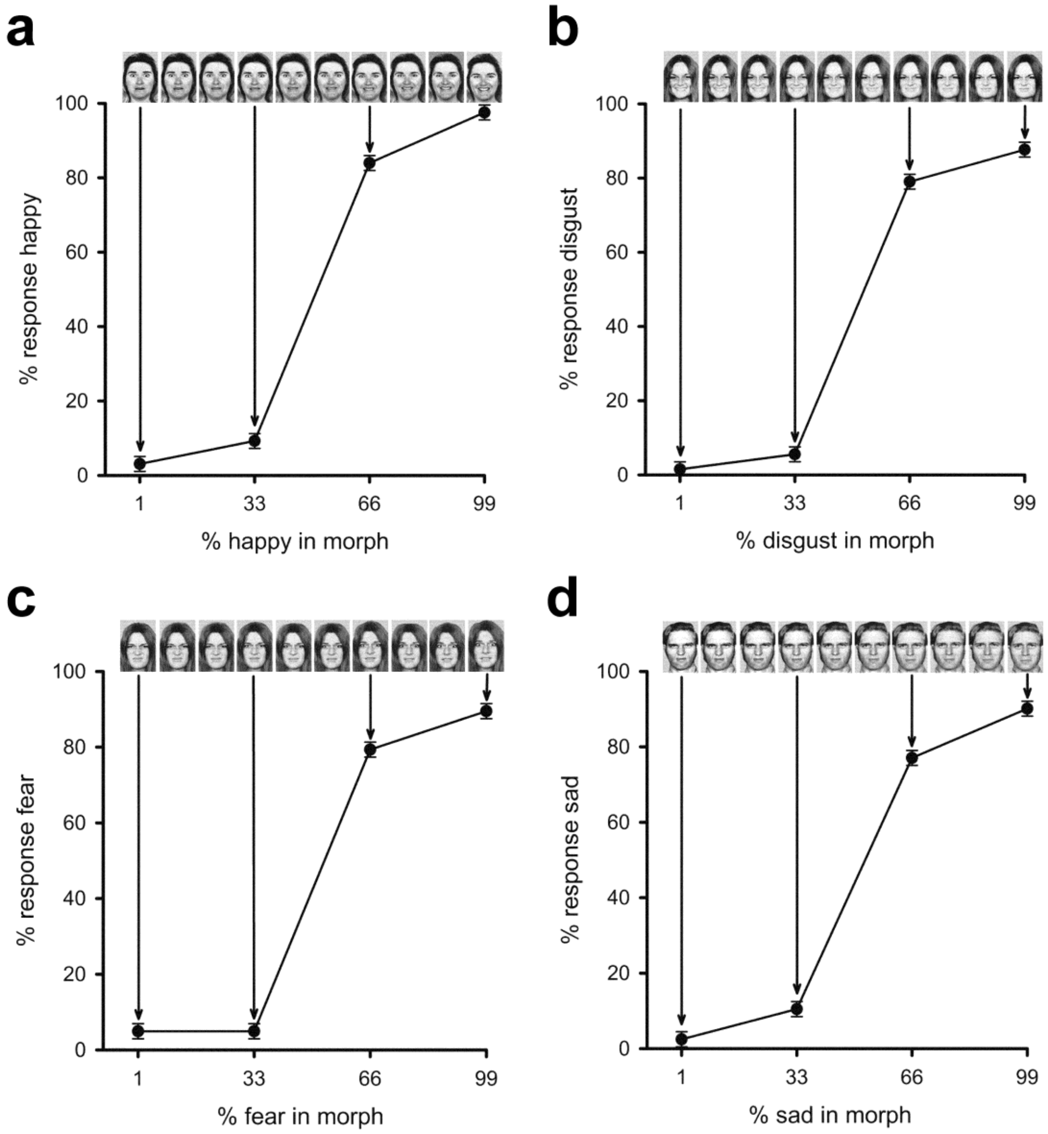

Figure 3.4 *Behavioural results from expression-classification experiment. The figure shows the results from the four continua that were most accurately recognised: a) fear-happy, b) happy-disgust, c) disgust-fear, d) disgust-sad, averaged across participants. The x-axis shows the four morphs levels from the continua that were used in the experiment. The graph represents the proportion of participant's responses that used the name given on the Y-axis, averaged across participants.*

3.3.2.2 Same/different discrimination task

The second behavioural experiment involved a more stringent test of categorical perception using a same-different task (Calder et al., 1996). Fourteen participants (11 females; mean age 24) were shown two sequentially presented faces with the same identity and had to judge whether the images were identical or different (2 AFC). There were 3 conditions (1) *same emotion* (99 % and 99 %, or 66 % and 66 %), (2) *within-emotion change* (99 % and 66 % images), (3) *between-emotion change* (66 % and 33 % images). Face images were each presented for 900 ms with an ISI of 200 ms. Trials were followed by a 2 s gray screen, during which participants could make their response (2 AFC).

This experiment found that images in the *between-emotion* condition were correctly judged as different more often relative to the *within-emotion* condition ($t_{(12)} = 6.47$, $p < 0.001$, Figure 3.5). Moreover, participants responded faster on correct responses to the *between-emotion* compared to the *within-emotion* condition ($t_{(2,24)} = 4.19$, $p < 0.001$). These results show that facial expressions that differ in perceived emotion are discriminated more easily than facial expressions that are perceived to convey the same emotion. This finding is widely considered to form the strongest test for behavioural evidence of categorical perception (Young et al., 1997).

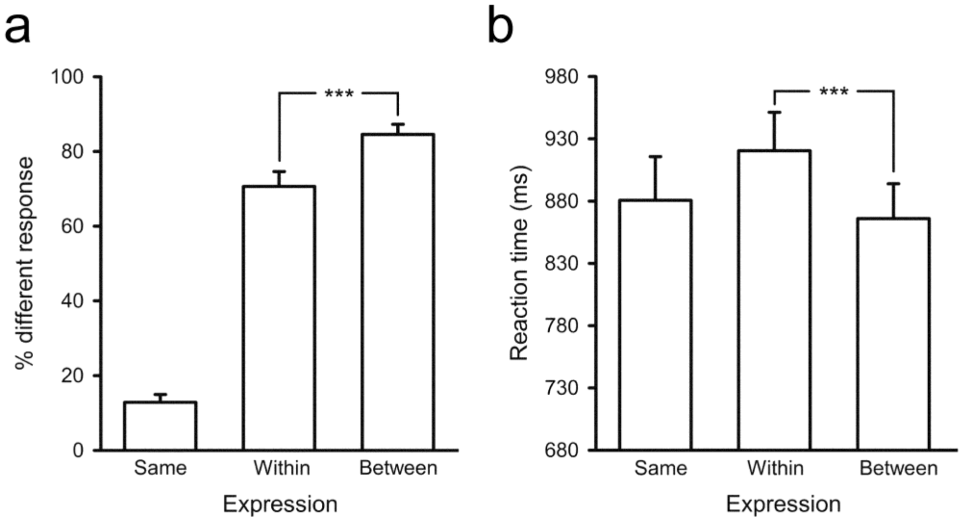

Figure 3.5 *Behavioural results from same/different discrimination experiment. a) Proportion of 'different' responses averaged across all participants for all three conditions. b) Reaction time (ms) for the correct responses, averaged across participants for the three conditions. Error bars represent standard error. *** p < 0.001.*

3.3.3 Methods – fMRI experiment

3.3.3.1 Subjects

All participants were right-handed and had normal or corrected-to-normal vision. All participants provided written consent and the study was given ethical approval by the

York Neuroimaging Centre Ethics Committee. Each participant took part in one of two experimental scans recording neural responses to conditions of interest, and a separate functional localiser scan to provide independent identification of face-selective regions. 25 participants took part in this experiment (19 females; mean age, 25).

3.3.3.2 Face localiser scan

A face localiser scan (Version 2; see Chapter 2.2.3.2 for protocol) was used to identify face-selective neural regions within each individual's brain.

3.3.3.3 Experimental scan

In this experiment, stimuli in a block were selected from 3 faces along the morphed continuum (99 %, 66 %, 33 %). The within-emotion condition used two faces from the morph continua that were on the same side of the category boundary (99 % and 66 %). The between-emotion condition used two faces along the morph continua that crossed the category boundary (66 % and 33 %). Importantly, the physical difference between images was therefore matched across within-emotion and between-emotion conditions (both 33 % difference). The same emotion condition had 2 identical images (99 % and 99 % or 66 % and 66 %). The identity of the faces was also varied to give six conditions: (1) *same-expression, same-identity,* (2) *within-expression, same-identity,* (3) *between-expression, same-identity,* (4) *same-expression, different-identity,* (5) *within-expression, different-identity,* (6) *between-expression, different-identity* (Figure 3.6). Faces were shown for 700 ms and separated by a 200 ms gray screen. Faces were presented in an AB block design with 6 faces per block. Successive blocks were separated by a 9 s fixation cross. Each condition was repeated 8 times in a counterbalanced order, giving a total 48 blocks. Faces were presented mounted on gray background and the bridge of the nose was aligned with the fixation cross to prevent images moving around the visual field. To ensure participants maintained attention they had to press a button on detection of a red dot which was superimposed onto 20 % of the images.

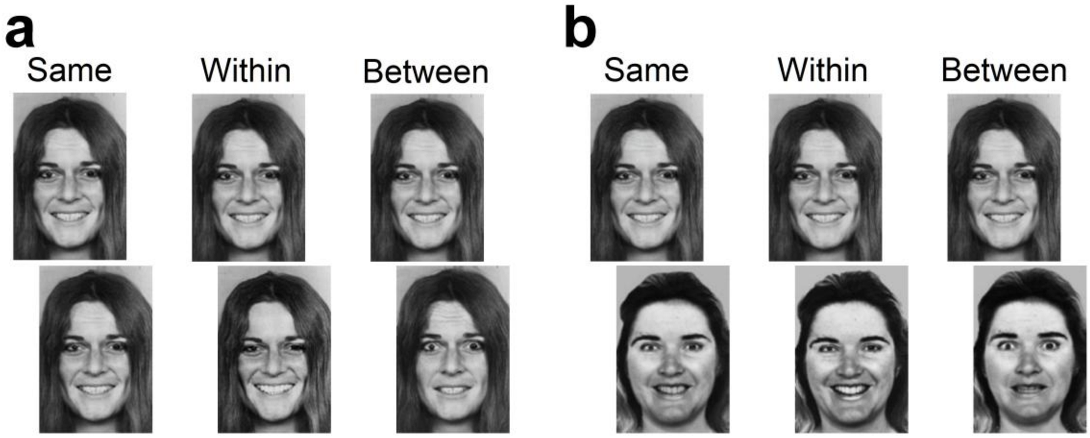

Figure 3.6 *Experiment 2 conditions. a) Images from the three expression conditions for the same identity. b) Images from the three expression conditions for the different identity.*

3.3.3.4 Imagining parameters and fMRI analysis

Imagining parameters and the steps involved in the fMRI analysis are reported in Chapter 2. See section 2.2.4 and 2.2.5.

3.3.4 Results –fMRI experiment

The localiser scan identified seven regions of interest that showed a greater response to faces than non-face stimuli. These regions, left and right OFA, left and right FFA, the left and right amygdala and the right pSTS, are show in Figure 3.7 and their locations are detailed in Table 3.2. As a face-selective part of the left STS could only be reliably identified within a limited number of participants it was not included in this analysis.

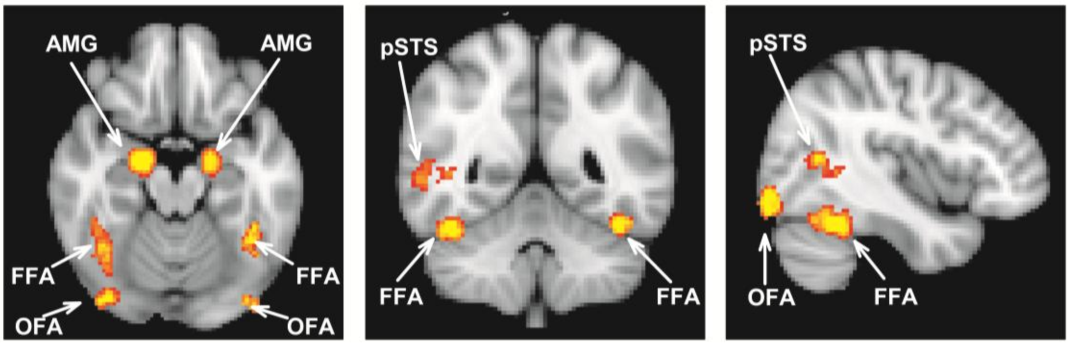

Figure 3.7 *Location of face selective regions in Experiment 2. Average region of interest across participants transformed into standard space. All brain images are depicted in radiologic convention, i.e. coronal and axial slices are left/right reversed.*

An ANOVA was conducted with the factors Condition (same-expression, same-identity; within-expression, same-identity; between-expression, same-identity; same-expression, different-identity; within-expression, different-identity; between-expression, different-identity), Region (OFA, FFA and amygdala) and Hemisphere (left, right). The pSTS was not included in this ANOVA as it was only reliably identified in the right hemisphere. The ANOVA revealed a significant Condition x Region x Hemisphere interaction ($F_{(10,70)} = 3.77$, $p < 0.001$). In order to unpack this interaction with hemisphere further comparisons were conducted. However there was no significant interaction between Hemisphere x Condition ($F_{(5,35)} = 1.92$, $p = 0.12$) or Hemisphere X Region ($F_{(2,14)} = 1.95$, $p = 0.18$) nor was there a main effect of Hemisphere ($F_{(1,7)} = 2.96$, $p = 0.13$). Post-hoc comparisons, using Tukey's HSD, were conducted to determine whether there were any significant differences between the responses to each condition across the hemispheres for each region of interest. This comparison revealed only the response in the amygdala to the same expression, different identity condition was significantly different in the right and left hemisphere. All other 17 cross-hemisphere comparisons did not exceed significance. This is clearly a complex interaction and one that would be interesting for future research. However, due to the conditions used and the lower SNR associated with the amygdala, further investigation of the difference between the right and left amygdala for the same expression, different identity condition is beyond the scope of this thesis. Therefore data from the left and right hemisphere were combined across hemispheres for all ROIs.

The peak responses of the face-selective regions were analysed using a 4 x 3 x 2 ANOVA with Region (pSTS, amygdala, FFA, OFA), Expression (same, within, between) and Identity (same, different) as the factors. There were significant effects of Expression ($F_{(2,24)} = 15.39$, $p < 0.0001$) and Region ($F_{(3,36)} = 49.12$, $p < 0.0001$) but not identity ($F_{(1,12)} = 3.88$, $p = 0.07$). There was also a significant interaction between Region x Expression ($F_{(6,72)} = 2.43$, $p = 0.03$). Therefore, to investigate which face-selective regions were sensitive to expression, and in what way each was sensitive to differences in expression, the response in each individual ROI is now considered.

Table 3.2 *MNI coordinates (mm) of face-selective regions in Experiment 2. Coordinates for the centre of gravity averaged across all participants in standard space is reported. Standard error is reported in parenthesis.*

Region	*n*	*x*	*y*	z
FFA	25			
L		-41 *(0.8)*	-56 *(2.1)*	-20 *(0.9)*
R		42 *(0.8)*	-54 *(1.2)*	-22 *(1.0)*
OFA	20			
L		-40 *(1.6)*	-85 *(1.4)*	-15 *(1.8)*
R		41 *(1.2)*	-83 *(1.1)*	-12 *(1.1)*
STS	17			
R		51 *(1.6)*	-48 *(2.0)*	5.3 *(1.2)*
Amygdala	21			
L		-19	-9	-15
R		20	-8	-17

Figure 3.8 shows the response from the pSTS in Experiment 2. A 3 x 2 ANOVA with the factors Expression (same, within, between) and Identity (same, different) was conducted to determine the sensitivity of the pSTS to these changes. This revealed a significant effect of Expression ($F_{(2,32)} = 13.19$, $p < 0.0001$) but no significant effect of Identity ($F_{(1,16)} = 0.60$) nor a significant interaction ($F_{(2,32)} = 0.25$). The main effect of expression for the *same-identity* conditions was due to significantly bigger responses to the *within-expression* ($t_{(16)} = 3.00$, $p = 0.01$) and *between-expression* ($t_{(16)} = 3.88$, $p = 0.001$) conditions compared to the *same-expression* condition. There was no significant difference between the *within-expression* and *between-expression* conditions ($t_{(16)} = 0.40$). A similar pattern of results was seen for the *different-identity* conditions. There were significantly bigger responses to the *within-expression* ($t_{(16)} = 2.49$, $p = 0.02$) and *between-expression* ($t_{(16)} = 3.17$, $p = 0.01$) conditions compared to the *same-expression* condition. However there was no significant difference in the response to the *within-expression* and *between-expression* conditions ($t_{(16)} = 0.22$). This equivalent sensitivity to both

within-expression and *between-expression* changes suggests that the pSTS has a continuous representation of expression.

In contrast to the pSTS, the amygdala was only sensitive to between-emotion changes in expression (Figure 3.8). An ANOVA revealed a significant main effect of Expression ($F_{(2,48)} = 22.52$, $p < 0.0001$) but not Identity ($F_{(1,24)} = 4.03$, $p = 0.06$) and there was also no significant interaction between Expression and Identity ($F_{(2,48)} = 1.49$, $p = 0.29$). For the *same-identity conditions*, there was no significant difference between the *same-expression* and *within-expression* conditions ($t_{(20)} = 1.61$, $p = 0.12$). However, there was a significant difference between the *same-expression* and *between-expression* conditions ($t_{(20)} = 4.86$, $p < 0.0001$) and between the *within-expression* and *between-expression* conditions ($t_{(20)} = 4.62$, $p < 0.0001$). There was a similar pattern for the *different-identity* conditions. There was no significant difference between the *same-expression* and *within-expression* conditions ($t_{(20)} = 0.84$), but there was a bigger response to the *between-expression* condition compared to the *same-expression* conditions ($t_{(20)} = 3.58$, $p < 0.0001$). There was also a bigger response to the *between-expression* condition compared to the *within-expression* condition ($t_{(20)} = 2.06$, $p = 0.05$). These results show that the amygdala is more sensitive to changes in expression that cross an emotion category boundary.

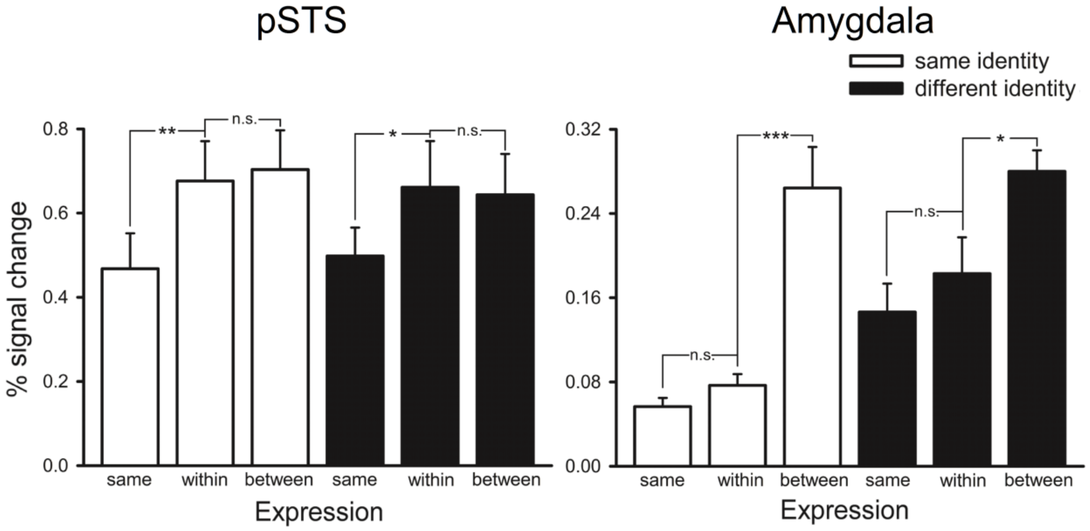

Figure 3.8 *Experiment 2 results. Peak responses to the different conditions in the pSTS and amygdala * p < 0.05, ** p < 0.01, *** p < 0.001.*

In the FFA (Figure 3.9) there was a significant main effects of Expression ($F_{(2,48)}$ = 6.36, p = 0.004) and Identity ($F_{(1,24)}$ = 9.29, p = 0.01). However, there was no significant interaction between Expression and Identity ($F_{(2,48)}$ = 2.54, p = 0.09). For the same-identity conditions, there was no significant difference between the *same-expression* and *within-expression* conditions ($t_{(24)}$ = 1.71, p = 0.10). There was also a no significant difference between the *within-expression* and *between-expression* conditions ($t_{(24)}$ = 1.38, p = 0.06). However, there was a significant difference between the *same-expression* and *between-expression* conditions ($t_{(24)}$ = 2.96, p = 0.01). For the *different-identity* conditions, there were no significant difference between the *same-expression* and either the *within-expression* ($t_{(24)}$ = 1.88, p = 0.07) or *between-expression* conditions ($t_{(24)}$ = 0.65). There was also no difference in response between the *between-expression* and *within-expression* conditions ($t_{(24)}$ = 1.50, p = 0.15).

In the OFA, there was a significant main effect of Expression ($F_{(2,48)}$ = 8.53, p = 0.001) and Identity ($F_{(1,24)}$ = 7.77, p = 0.01). There was also a significant interaction between Expression and Identity ($F_{(2,48)}$ = 4.15, p = 0.02). The interaction was due to differences between the *same-identity* conditions, but not between the *different-identity* conditions (Figure 3.9). For the *same-identity conditions*, there was a significant difference between the *same-expression* and *within-expression* conditions ($t_{(19)}$ = 2.14, p = 0.05) and the *same-expression* and *between-expression* conditions ($t_{(19)}$ = 2.51, p = 0.02). However, there was no significant difference between the *within-expression* and *between-expression* conditions ($t_{(19)}$ = 0.69). In contrast, for the *different-identity* conditions, there were no significant difference between the *same-expression* and either the *within-expression* ($t_{(19)}$ = 1.61, p = 0.22) or *between-expression* conditions ($t_{(19)}$ = 0.26). There was also no difference in response between the *between-expression* and *within-expression* conditions ($t_{(19)}$ = 1.44, p = 0.17).

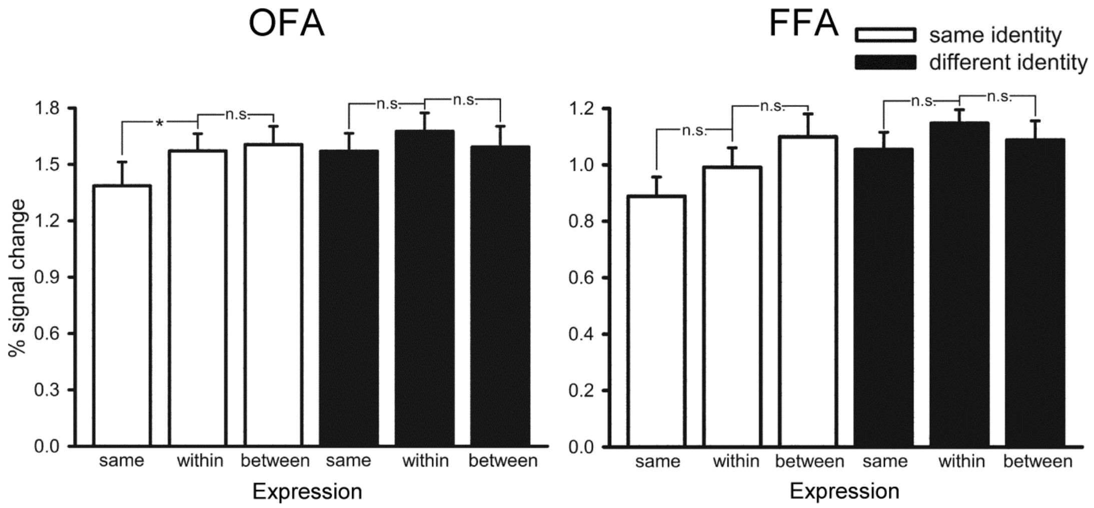

Figure 3.9 *Experiment 2 results - peak responses to the different conditions in the OFA and FFA. * p < 0.05.*

The results from Experiment 2 reveal dissociable representations of expressions in two expression sensitive neural regions. Consistent with a continuous representation of expression, the pSTS was sensitive to any change in the image. In contrast, the representation of expression in the amygdala was more categorical.

3.3.5 Discussion

The aim of this study was to determine how facial expressions of emotion are represented in face-selective regions of the human brain. In this experiment morphs between expressions were used to determine whether the response to expression in the pSTS and the amygdala revealed a categorical or continuous representation. These results clearly show a dissociation between the code used by these regions to represent expression; the pSTS processes facial expressions of emotion using a continuous neural code, whereas the amygdala has a more categorical representation of facial expression.

These findings offer an alternative to the longstanding controversy about whether facial expressions of emotion are processed using a continuous or categorical code. Behavioural findings consistent with a categorical representation of facial expression are evident when participants report discrete rather than continuous changes in the emotion of faces that are morphed between two expressions (Calder et al., 1996;

Etcoff & Magee, 1992). Stronger evidence for a categorical representation is seen in the enhanced discrimination of face images that cross a category boundary compared to images that are closer to the prototype expressions (Calder et al., 1996; Etcoff & Magee, 1992). Nonetheless, a purely categorical perspective is unable to account for the systematic pattern of confusions that can occur when judging facial expressions (Woodworth & Scholsberg, 1954) and it also has difficulty explaining why morphed expressions that are close to the category prototype are easier to recognise than expressions belonging to the same category but more distant from the prototype (Young et al., 1997). So there is evidence to support both continuous and categorical accounts of facial expression perception (Bruce & Young, 2012).

In an attempt to resolve these seemingly conflicting positions, more recent computational models have suggested that a unitary representation could underpin both a continuous and a categorical coding of facial expression (Dailey, Cottrell, Padgett, & Adolphs, 2002; Martinez & Du, 2012). Our results provide a converging perspective by showing that different regions in the face processing network can have either a primarily categorical or a primarily continuous representation of facial expression. Of course, the more categorical response in the amygdala compared to the pSTS does not imply that the amygdala is insensitive to changes in facial expression that do not result in a change to the perceived emotion. Indeed, a perceiver needs to be aware both of the category to which a facial expression belongs (its social meaning) and its intensity, and a number of studies have shown that responses in the amygdala can be modulated by changes in the intensity of an emotion (Morris et al., 1996; Thielscher & Pessoa, 2007). Nevertheless, the key finding here is that there is a dissociation between the way facial expressions of emotion are represented in the pSTS and amygdala.

The importance of understanding how facial expressions of emotion are represented in the brain reflects the significance of attributing meaning to stimuli in the environment. When processing signals that are important for survival, perception needs to be prompt and efficient. Categorical representations of expression are optimal for making appropriate physiological responses to threat. The more categorical representation of facial expressions of emotion in the amygdala is consistent with its role in the detection and processing of stimuli pertinent to survival (Adolphs et al., 1999; Sander et al., 2003). Indeed, neuropsychological studies of patients with amygdala damage have demonstrated impairments in emotion recognition (Adolphs et al., 1994; Anderson & Phelps, 2000; Young et al., 1995), which are often accompanied by an attenuated reaction to potential threats (Feinstein et al., 2011; Sprengelmeyer et al., 1999). Although deficits in emotion recognition following amygdala damage have mostly been reported for the perception of fear, more recent functional neuroimaging studies have provided support for a role of the amygdala in the processing of other emotions (Phan et al., 2002; Sergerie et al., 2008; Winston et al., 2003). In the present study, the morphed stimuli were not restricted to those involving fear, so the categorical response shown in this study provides further support for the involvement of the amygdala in processing a range of facially expressed emotions.

However, not all naturally occurring circumstances demand a categorical response, and there are many everyday examples where a continuous representation might be of more value. For example, although there appear to be a small number of basic emotions that seem to be recognised consistently across participants, there are many facial expressions that do not correspond to one or other of these categories. Even with basic emotions, the expressions can actually be quite variable in ways that can signal subtle but important differences (Rozin et al., 1994). Furthermore, judgements of facial expression can be influenced by the context in which they are seen (Russell & Fehr, 1987). Together, these findings suggest a more flexible continuous representation is also used in judgements of facial expression. The results from this study suggest that the pSTS could provide a neural substrate for this continuous representation. These results are consistent with a previous study that used MVPA to show a continuous representation of expression in the pSTS (Said, Moore, Norman, Haxby, & Todorov, 2010). These findings highlight the role of the pSTS in processing moment-to-moment signals important in social communication (Allison et al., 2000).

An interesting question for further studies concerns how the differences between responses of pSTS and amygdala are represented at the single cell level. The BOLD signal measured in fMRI clearly reflects a population response based on the aggregated activity of large numbers of neurons, and there are many ways in which differences in responses across brain regions might therefore be represented in terms of coding by single cells. A suggestion as to a plausible way in which the population responses shown in the fMRI data might reflect coding by cells in pSTS and amygdala is now offered.

Facial expressions are signalled by a complex pattern of underlying muscle movements that create variable degrees of change in the shapes of facial features such as the eyes or mouth, the opening or closing of the mouth to show teeth, the positioning of upper and lower teeth, and so on. An obvious hypothesis, then, is that pSTS cells are involved in coding this wide variety of individual feature changes, and this would be consistent both with the data presented here showing an overall sensitivity of pSTS to any change in expressive facial features and with other fMRI findings that demonstrate pSTS responsiveness both to mouth movements and to eye movements (Pelphrey, Morris, Michelich, Allison, & McCarthy, 2005). When expressions are perceptually assigned to different emotion categories, however, the underlying feature changes are used simultaneously, so that a particular emotion is recognised through a holistic analysis of a critical combination of expressive features (Calder et al., 2000). Cells in the amygdala would therefore be expected to have this property of being able to respond to more than isolated features and it is known, for example, that the amygdala's response to fearful expressions is based on multiple facial cues since it can be driven by different face regions when parts of the face are masked (Asghar et al., 2008).

The same distinction can clearly be seen in computational models of facial expression perception such as EMPATH (Dailey et al., 2002). EMPATH forms a particularly good example as it is a well-developed 'neural network' model that is able to simulate effects from behavioural studies of facial expression recognition that show continuous or categorical responses in classification accuracy and reaction time. To achieve this, EMPATH has layers of processing units that correspond to early stages of visual analysis (Gabor filters, considered as analogous to V1), to Principal Components (PCs) of variability between facial expressions (as identified by PCA of facial expression images), followed by a final classification stage based on the outputs of the PC-responsive layer. There is a clear parallel between coding expression PCs and the fMR properties reported for pSTS, and between classifying PC output combinations and the fMR response from the amygdala.

As mentioned in Experiment 1, the most efficient way to analysis the variety of information available in a face is thought to involve different neural regions optimally tuned to different properties in the face (Bruce & Young, 1986, 2012; Haxby et al., 2000). As a result models of face processing have proposed that the changeable aspects of the face, for example expression, are processed independently from the invariant features of a face such as identity. Consistent with Experiment 1, this study found that the response in the pSTS and amygdala to expression was largely independent to changes in identity. Although the pSTS has been shown to be influenced by identity (Andrews & Ewbank, 2004; Fox et al., 2009; Winston et al., 2004) the results reported in Experiments 1 and 2 suggest that the neural responses in pSTS and amygdala are primarily driven by changeable aspects of the face, such as expression. In contrast, and again consistent with the findings in Experiment 1, this experiment found that the OFA and FFA were sensitive to both changes in expression and identity. These findings might be seen as consistent with previous studies that suggest the FFA is involved in judgements of identity and expression (see for example Cohen Kadosh et al., 2010; Fox et al., 2009; Ganel et al., 2005). However, it also possible that these results show that the FFA is just sensitive to any structural change in the image.

In conclusion, this experiment found that face-selective regions in the pSTS and amygdala were sensitive to changes in facial expression, independent of changes in facial identity. Using morphed images, the results showed that the pSTS has a continuous representation of facial expression, whereas the amygdala has a more categorical representation of facial expression. The continuous representation used by pSTS is consistent with its hypothesised role in processing changeable aspects of faces that are important in social interactions. In contrast, the responses of the amygdala are consistent with its role in the efficient processing of signals that are important to survival.

3.4 Experiment 3: Dynamic stimuli reveal a selectivity for facial expressions of emotion in the amygdala, but not in other face-selective regions of the human brain

3.4.1 Introduction

The STS and the amygdala are sensitive to range of basic facial expressions (Breiter et al., 1996; Morris et al., 1996; Narumoto et al., 2001, Said et al., 2010). The majority of studies investigating the role of these regions to facial expression have used static images which convey the apex of the emotion. However, facial expressions are naturally dynamic when encountered in the environment (Edwards, 1998) and humans appear sensitive to this information (Ambadar et al., 2005; Kamachi et al., 2001; Wehrle et al., 2000). Sensitivity to the dynamic component of facial expressions is reflected in neuroimaging studies which show modulation of the STS and amygdala to dynamic compared to static facial expressions (LaBar, Crupain, Voyvodic, & McCarthy, 2003; Pitcher et al., 2011)

Experiments 1 and 2 demonstrated that although the response in both the STS and amygdala to facial expression was largely independent of identity, there was a difference in the way these regions represented information about facial expression. The aim of this study was to further explore the representation of expressions in these regions using dynamic stimuli. Movies were generated that displayed a dynamic change in expression by morphing between a neutral and a prototypical expression (either anger, disgust, fear, happiness or sadness). The four conditions used in Experiment 1 were implemented in this study. However, instead of the stimuli being static images of the apex of the emotion, the stimuli were instead movies that were played from a neutral to a prototype expression in the following four conditions: (1) *same-expression change, same-identity* (2) *same-expression change, different-identity* (3) *different-expression change, same-identity* (4) *different-expression change, different-identity*. So, within a block, five movies were played, in the same expression conditions these movies would show the same change in expression whereas in the different expression conditions each movie would have a different change in expression. Regions sensitive to the expression information should show a greater response to changes in expression compared to changes in identity.

3.4.2 Methods

3.4.2.1 Subjects

All participants were right-handed and had normal or corrected-to-normal vision. Visual stimuli (8˚ x 8˚) were back-projected onto a screen located inside the magnetic bore, 57 cm from subjects' eyes. All subjects provided written consent and the study was given ethical approval by the York Neuroimaging Centre Ethics Committee. Nineteen participants took part the experiment (14 females; mean age, 23).

3.4.2.2 Localiser scan

A localiser scan (Version 1) was used to identify face-selective neural regions within each individual's brain (see Chapter 2.2.3.1).

3.4.2.3 Experimental Scan

The aim of this experiment was to investigate the sensitivity of the pSTS and amygdala to dynamic changes in expression and static changes in identity. There were four conditions in this experiment which all involved a dynamic change in expression: (1) *same-expression change, same-identity*, (2) *same-expression change, different-identity*, (3) *different-expression change, same-identity*, (4) *different-expression change, different-identity*. The *same-expression change* conditions involved 5 movies all displaying the same change in expression. In the *different-expression change* conditions each of the 5 movies displayed a different emotion change (Figure 3.10). Each movie involved one identity, in the *same-identity* conditions the same identity was shown in each of the 5 movies. In the *different-identity* conditions each of the 5 movies had a different identity. The movie stimuli were presented in blocks, with 5 movies per block. Each movie started at a neutral expression and finished in a prototype expression (movies were generated in Psychomorph by morphing between two different expressions: see Chapter 2.1.6). Each movie was presented for 1160 ms and separated by a gray screen presented for

200 ms. Successive stimulus blocks were separated by a 9 s fixation gray screen. Each condition was presented 10 times in a counterbalanced order, giving a total of 40 blocks. This resulted in total experiment length of 10.5 mins. To ensure participants maintained attention throughout the experiment, participants had to push a button when they detected the presence of a red dot, which was superimposed onto 20 % of the movies.

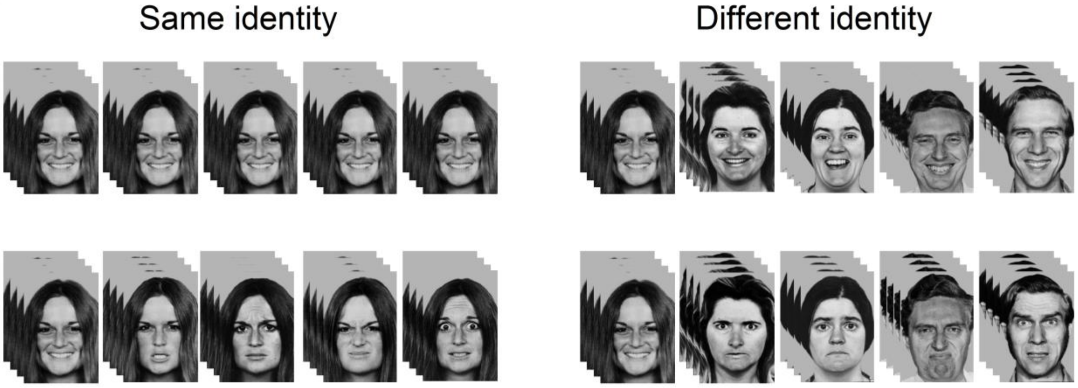

Figure 3.10 *Experiment 3 conditions. Top row: same-expression change, same-identity (left), same-expression change, different-identity (right). Bottom row: different-expression change, same-identity (left), different-expression change, different-identity (right).*

3.4.2.4 Imaging parameters and fMRI analysis

Imagining parameters and the steps involved in the fMRI analysis are reported in Chapter 2. See section 2.2.4 and 2.2.5.

3.4.3 Results

The localiser identified six regions of interest that were more responsive to faces than the non-face stimuli. These regions, left and right OFA, left and right FFA, right amygdala and right pSTS can be seen in Figure 3.11 and the locations of these regions are detailed in Table 3.3. To investigate whether there was an effect of hemisphere a 4 x 2 x 2 ANOVA with the factors Condition (same-expression change, same-identity; same-expression change, different-identity; different-expression change, same-identity; different-expression change, different-identity) Region (OFA, FFA) and Hemisphere (left, right) was conducted. The amygdala and pSTS are omitted from this ANOVA investigating the effect of hemisphere as they were only identified in the right hemisphere. There was no significant Condition x Hemisphere x Region interaction ($F_{(3,\ 36)}$ = 1.17, p = 0.34), nor a significant interaction between Condition x Hemisphere ($F_{(3,\ 36)}$ = 0,45), nor a significant interaction between Region and Hemisphere ($F_{(1,\ 12)}$ = 0.63) and there was no significant effect of Hemisphere ($F_{(3,\ 12)}$ = 2.73, p = 0.13), therefore the responses for the OFA and FFA were combined across hemisphere.

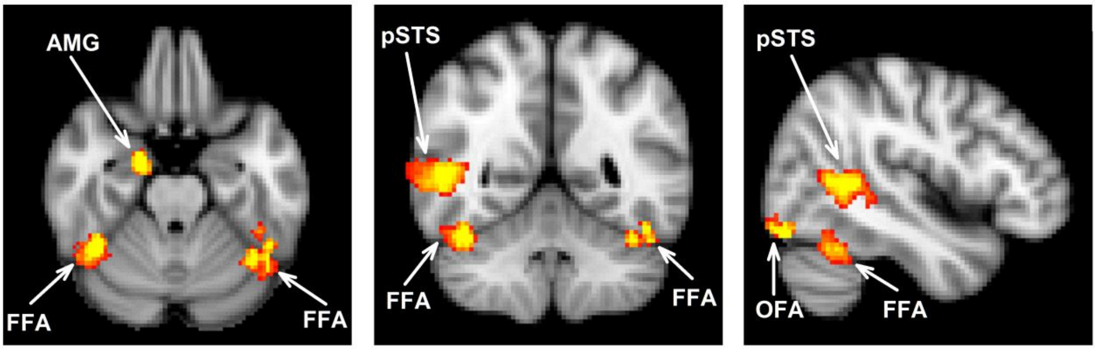

Figure 3.11 *Locations of face-selective regions in Experiment 3. Average region of interest across participants transformed into standard space. All brain images are depicted in radiologic convention, i.e. coronal and axial slices are left/right reversed.*

The data were first analysed using a 4 x 2 x 2 ANOVA with Region (pSTS, amygdala, FFA, OFA) Expression (same, different) and Identity (same, different) as the factors. There were significant effects of Expression ($F_{(1,14)}$ = 7.30, p = 0.02) and Region ($F_{(3,42)}$ = 63.71, $p < 0.0001$) but not Identity ($F_{(1,14)}$ = 4.35, p = 0.06). There was also a significant interaction between Region x Expression ($F_{(3,42)}$ = 3.06, p = 0.04). Therefore to investigate which face-selective regions were sensitive to dynamic expressions the response in each individual ROI is now considered.

Table 3.3 *MNI coordinates (mm) of face-selective regions in Experiment 3. Coordinates for the centre of gravity averaged across all participants in standard space is reported. Standard error is reported in parenthesis.*

Region	*n*	*x*	*y*	z
FFA	19			
L		-41 *(1.0)*	-54 *(1.5)*	-21 *(1.0)*
R		43 *(1.1)*	-55 *(3.2)*	-22 *(1.6)*
OFA	19			
L		-39 *(2.1)*	-84 *(1.5)*	-16 *(0.9)*
R		43 *(1.6)*	-80 *(2.0)*	-14 *(1.2)*
STS	18			
R		53 *(1.7)*	-51 *(2.6)*	4.7 *(1.0)*
Amygdala	16			
R		17	-9	-18

Figure 3.12 shows the peak responses in the posterior part of the right STS. The pSTS was sensitive to any change in the expression. A 2 x 2 ANOVA with the factors Expression (same, different) and Identity (same, different) revealed no significant effect of Expression ($F_{(1,17)} = 0.66$), or Identity ($F_{(1,17)} = 0.20$). There was also no significant interaction between Expression and Identity ($F_{(1,17)} = 1.97$, $p = 0.18$).

In contrast, the amygdala was sensitive to blocks of faces in which the dynamic change in expression was different across the block (Figure 3.12). A 2 x 2 repeated measures ANOVA found a significant main effect of Expression ($F_{(1,15)} = 5.10$, $p = 0.04$) but not Identity ($F_{(1,15)} = 0.23$). There was no significant interaction between Expression and Identity ($F_{(1,15)} = 0.08$). The main effect of Expression was due to the bigger response to the *different-expression change* conditions compared to the *same-*

expression change conditions (different expression: 0.19 %, same expression: 0.05 %).

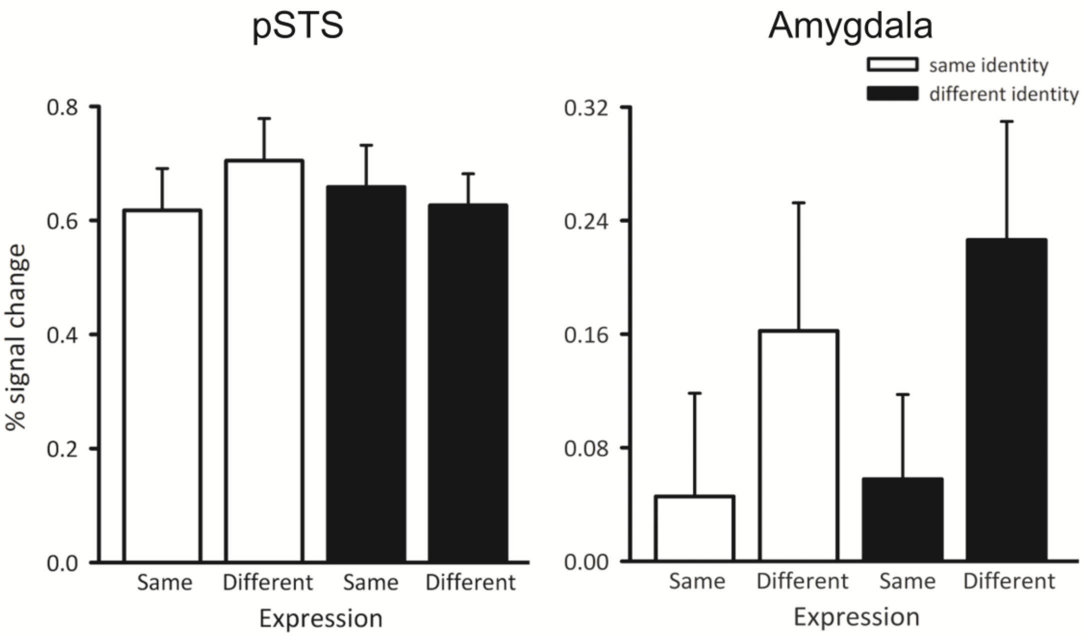

Figure 3.12 *Experiment 3 results. Peak responses to the four conditions in the pSTS and amygdala.*

The results from the FFA show a greater response to the different identity conditions compared to the same (Figure 3.13). A 2 x 2 ANOVA revealed no significant main effect of Expression ($F_{(1,18)}$ = 0.44) but there was a main effect of Identity ($F_{(1,18)}$ = 6.37, p = 0.02). There was no significant interaction between Expression and Identity ($F_{(1,18)}$ = 3.48, p = 0.08). The main effect of Identity was due to a bigger response to the *different-identity* conditions compared to the *same-identity* conditions (different-identity: 1.13 %, same-identity: 1.03 %).

The OFA shows a similar pattern of response to that found in FFA (Figure 3.13). There was no significant effect of Expression ($F_{(1,18)}$ = 0.73) but there was a significant effect of Identity ($F_{(1,18)}$ = 10.15, p = 0.01). There was also a significant interaction between Expression and Identity ($F_{(1,18)}$ = 4.47, p = 0.05). The interaction was due to a significantly bigger response to *different-identity* condition compared to *same-identity* condition for the *same-expression change* ($t_{(18)}$ = 3.31, p = 0.004) but not for the *different-expression change* conditions ($t_{(18)}$ = 1.59, p = 0.39).

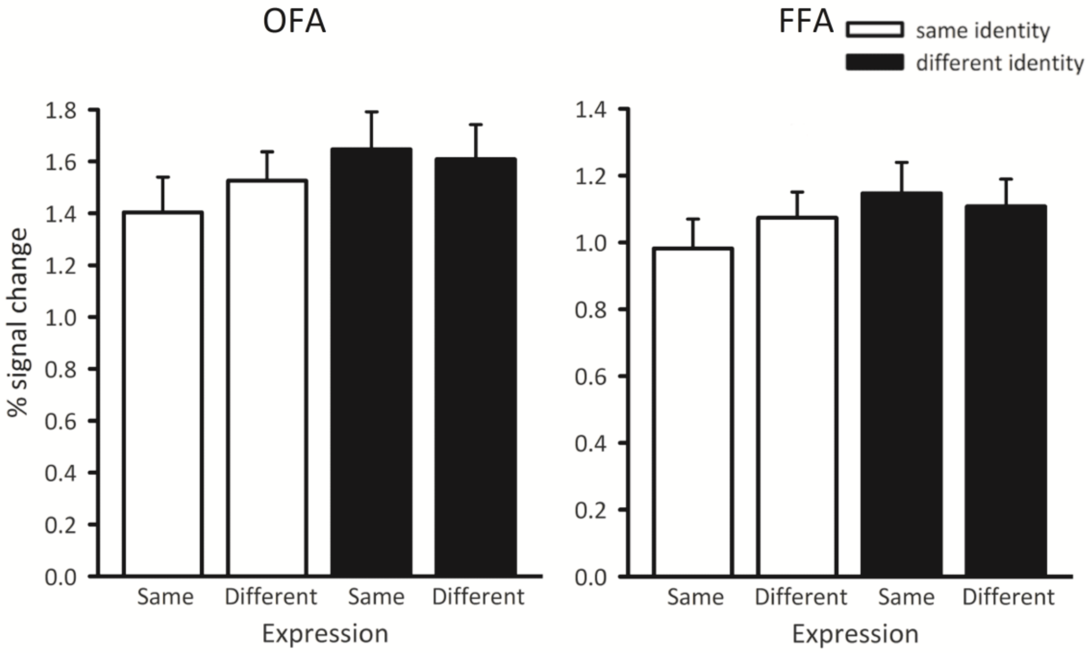

Figure 3.13 *Experiment 3 results. Peak responses from the OFA and FFA to the four conditions in this experiment.*

The results from this experiment reveal that the amygdala was sensitive to the emotion category, with a greater response to blocks of movies that varied in the emotion category compared to blocks of movies displaying the same change in emotion. This is consistent with the more categorical response found in this region in Experiment 2. This is dissociable form the response in the pSTS which did not discriminate between blocks with same change and different changes in expression. This is consistent with a continuous representation which was demonstrated in Experiment 2.

3.4.4 Discussion

The aim of this experiment was to investigate the sensitivity of the pSTS and amygdala to facial expression and identity using more ecologically valid dynamic stimuli that displayed a change in expression. This experiment clearly demonstrates that the amygdala is sensitive to the category of the emotion. The amygdala responded more to dynamic changes in expression that displayed a new emotion category compared to dynamic changes in expression that did not change the emotion. In contrast, the pSTS, responded to any dynamic change in expression regardless of whether there was a change to the emotion category.

These results suggest that the amygdala plays an important role in categorising different facial expressions of emotion. The amygdala is known to be critical in processing biological relevant stimuli (Sander et al., 2003; Whalen, 1998). This is reflected in neuropsychological studies in which lesions in the amygdala result in impairments in interpreting emotion (Adolphs et al., 1994; Anderson & Phelps, 2000; Young et al., 1995) and is often accompanied by attenuated response to threat (Feinstein et al., 2011; Sprengelmeyer et al., 1999). Experiment 2 demonstrated that the amygdala uses more categorical representations of expressions which are optimal for making the appropriate response to physiological threat. The use of dynamic stimuli in this experiment extends this finding by showing that the amygdala is more sensitive to changes in expression that result in a categorical change in the emotion displayed. Previous studies have shown that the response in the amygdala can be modulated by dynamic changes in facial expressions compared with static expressions (LaBar et al., 2003; Pitcher et al., 2011). This experiment demonstrates that the increased response associated with the movie stimuli is not due to increased attention, rather it is change in the emotion present in the movie that drives the response in the amygdala.

However, assigning expressions into discrete categories of emotion is not always an appropriate response to facial expressions. Continuous representations of expressions can account for the variability in how the basic expressions are posed and the influence of context on the interpretation of expression (Rozin et al., 1994; Russel & Fehr, 1987). Experiment 2 demonstrated that the pSTS is a region sensitive to gradations in the intensity of facial expressions. Consistent with this, this experiment found that the pSTS was sensitive to any dynamic change in facial expression. The pSTS did not show a difference between the same expression and different expression conditions. This could be attributed to the nature of the movie stimuli and the conditions used in this experiment. Both the same and different expression conditions involved movies that displayed a change in expression from neutral to a prototype expression. A region that holds a continuous representation of expression should be sensitive to both types of expression changes. Taken together these findings are consistent with the role of the pSTS in processing moment-to-moment signals important in social communication (Allison et al., 2000).

Models of face processing propose that the most efficient way to extract the large amount of information available from a face is to recruit different neural subcomponents optimally tuned to certain face signals. As such, the cues to facial expression and identity are thought be extracted relatively independently (Bruce & Young, 1986, 2012; Haxby et al., 2000). This experiment identified that the response in the amygdala is primarily driven by the changeable aspects of the face. In contrast the FFA is thought to represent cues important for facial identity (Grill-Spector et al., 2004; Roteshtein et al., 2005; Yovel & Kaniwisher, 2005). The results from this study are consistent with the FFA being a region more sensitive to facial identity than expression. However, the response in the FFA does not reflect a fully image-invariant representation of the type often considered necessary for identity recognition. The FFA did not discriminate between the same and different identity when the expression was varied. The FFA, therefore, may extract certain aspects of expression information that are relevant to the identity information (see Cohen Kadosh et al., 2010; Fox et al., 2009; Ganel et al., 2005). However, it is also possible that the FFA could be extracting a mean representation across a movie in order to process the image properties.

In conclusion, using dynamic changes in facial expression, this experiment demonstrated that the amygdala was only sensitive to dynamic changes in expression that resulted in a different category of expression being displayed by each movie in the block of trials. This category dependent representation in the amygdala may reflect the efficient processing needed when processing stimuli pertinent to survival. The response in the amygdala was dissociable form that found in the pSTS which was sensitive to any change in expression; the representation of any change in expression is needed to interpret signals important in social communication.

3.5 Conclusions

The aims of this chapter were to identify face-selective neural regions that were sensitive to facial expression and investigate the neural representation of expressions in those regions. Experiment 1 highlighted two face-selective regions, the pSTS and the amygdala, in which the response to facial expression was largely independent of identity. Experiments 2 and 3 probed the representation of expression in these regions. Both experiments found that these different neural regions can hold primarily categorical or continuous representations of expression. In the amygdala, a more categorical representation of expression was found, with expressions being assigned into discrete categories of emotion. This reflects the proposed role of the amygdala in the efficient processing of signals pertinent to survival. In contrast, the pSTS was sensitive to any change in expression, which suggests a graded representation of expression along continuous dimensions of variation. This is consistent with this region's role in processing information important in social communication.

Chapter 4

The role of shape-based and surface-based information in the processing of facial expression

4.1 Introduction

Chapter 3 explored the neural coding of facial expression and found evidence for both categorical and continuous representations of expression. This chapter asks the related question - what facial information is used to code expression?

Any view of a face involves a pattern of light and dark regions resulting from an interaction between the face's surface pigmentation and ambient illumination conditions. The pigmentation of the face often changes sharply at the boundaries of facial features such the as lips or eyes. The salience of these changes can be seen in the fact that computer edge-detection algorithms can easily extract a basic representation of the position and shapes of key facial features (Bruce & Young, 1998). Because facial expressions are conveyed by a complex pattern of underlying muscle movements that alter the shapes of facial features (Ekman, 1972), this edge-based information may contain many useful cues to expression from the shapes of the mouth, eyes and so on.

As well as defining feature shapes, though, the pattern of light across a face also provides cues to surface-based texture patterns and, via shape from shading, can provide some information about the 3D structure of the face (Bruce & Humphreys, 1994; Bruce & Young, 1998, 2012). Texture information may also have a potential role in interpreting facial expressions because some expressions introduce substantial regional texture changes; for example opening the mouth often creates a dark area, or showing the teeth introduces to a lighter region.

A broad distinction, then, can be made between feature shape (edge-based) cues and larger surface texture patterns (Bruce and Young, 1998). These shape-based and surface-based cues are thought to contribute differently to the perception of facial expression and facial identity. The perception of facial expression is thought to be predominantly based on the shape-based cues whereas the perception of facial identity is more reliant on surface-based cues (Bruce & Langton, 1994; Bruce, Valentine, & Baddeley, 1987; Kemp, Pike, White, & Musselman, 1996; White, 2001; White & Li, 2006).

The importance of shape-based cues in the categorisation of expression is shown by the reasonably accurate recognition of expression from line drawings. Line drawings are free from surface-based information and only provide shape-based cues. The accurate perception of expression from line drawings demonstrates that expression recognition is predominantly driven by the available shape-based information (Kirita & Endo, 1995; Magnussen, Sunde, & Dyrnes, 1994; Mckelvie, 1973). In contrast, recognition of facial identity from line drawings is relatively poor (Rhodes, Brennan, & Carey, 1987). Davies, Ellis and Shepherd (1978) found that recognition of facial identity was most accurate from photographs, followed by detailed line-drawings (which provided some surface-based information) and then outline drawings which were free from surface-based cues. This demonstrates the importance of surface-based, rather than shape-based, cues in the accurate recognition of facial identity.

If judgements of expression predominately rely on shape-based information, then it should prove difficult to discriminate between different facial expressions from faces devoid of shape information. One way to remove the shape-based cues involves blurring or pixilating the face images. This manipulation disrupts the high-spatial frequency information which carries shape-based cues in a face, whilst preserving low-spatial frequencies which convey surface-based cues (Collishaw & Hole, 2000; Goffaux, Hault, Michel, Vuong, & Rossion, 2005; Schyns & Oliva, 1999). White and Li (2006) found that blurring and pixilation of faces impaired performance on expression judgements whilst judgements of identity were unaffected. This implies that the shape-based information carried by high-spatial frequencies were critical for judgements of expression but were relatively unimportant for identity judgements.

Principal Component Analysis (PCA) has been used to explore whether facial expressions can be categorised from shape-based and surface-based information. PCA is a technique which can describe complex stimuli in terms of a limited number of principal components or factors which represent the variability and similarity between different images (Calder, 2011). Calder, Burton, Miller, Young, and Akamatsu, (2001) applied

PCA to four different types of images: (1) full images that were pre-processed to have the same inter-ocular distance and eye position; (2) Shape-free images in which the faces had been averaged to the same face shape; (3) Shape only images in which a small number of anatomical landmarks were defined; and (4) Shape-free + shape images in which the shape cues were added to the shape-free images. They found that facial expressions could be categorised significantly above chance in all of the image conditions but optimal categorisation was achieved in the shape-free + shape condition. However, when the PCA was restricted to the 10 most important components for expression categorisation it was evident that shape cues were relatively more important than texture cues for the categorisation of expression. Together, PCA suggests that shape cues are of primary importance to expression categorisation, but optimal categorisation of facial expressions is achieved when both shape and texture cues are available.

Photo-negation offers a useful way to tease apart the relative contributions of shape and surface-based cues. Because the contrast between the light and dark regions of the face is reversed in a photo negative image, texture patterns and shape from shading cues are largely disrupted. Conversely, the boundaries between light and dark regions that define the shapes of features remain in equivalent positions in a photo positive and its corresponding photo negative image. So photo negation should largely disrupt the use of texture but not of shape-based information. If expression judgements predominantly rely on shape-based cues then the ability to discriminate facial expression should be relatively equivalent in photo positive and negative (Bruce & Young, 1998; Santos & Young, 2008). This was found in a study by White (2001) in which participants' matching of expression was relatively unaffected by photo negation whereas judgments of identity became less accurate and took longer. The adverse affect of photo negation on identity judgements has been shown across a range of experiments and demonstrates the relative importance of surface-based information in judgements of identity (Bruce & Langton, 1994; Bruce et al., 1991; Galper, 1970; Kemp et al., 1996; Russell, Sinha, Biederman, & Nederhouser, 2006)

The aim of this chapter is to further probe the neural coding of expression by investigating the relative contributions of shape-based and surface-based information to the representation of facial expressions. Photo negation was used to invert the surface-based information whilst preserving shape-based cues. There are two experiments in this chapter. Experiment 4 reports the results of two behavioural studies, which probe the recognition of expression in photo negated faces. In Experiment 5, the fMRI response to photo-negated face images was measured in face-selective regions to probe the contribution of shape-based and surface-based information to the neural representation of faces.

4.2 Experiment 4: Photographic negation reveals the importance of shape-based facial cues to the perception of expression

4.2.1 Introduction

Photo negation was adopted in this Experiment in order to investigate the relative contribution of shape-based and surface-based information in the perception of faces. The Experiment aimed to validate the photo negated stimuli for use in an fMRI experiment (Experiment 5) and to further explore the effect of photo negation on judgements of facial expression and identity. In this Experiment participants made same/different expression and identity judgements for contrast positive and negative faces. As the representation of facial expression is thought to be largely dependent on shape-based information that is relatively spared by photo negation, it was predicted that photo negation should have little effect on judgements of expression. Conversely, as facial identity is conveyed mainly by the surface-based information, the accuracy of identity judgements should decrease when stimuli are in contrast negative.

Reversing the contrast of gray scale faces results in the hair becoming a striking white colour and therefore a potentially useful cue when matching identity in contrast negative. This is reflected in a study by Liu and Chaudhuri, (1998) in which they found that recognition from the internal features of a face was more impaired by photo negation than the recognition from the external features. As external features can be useful when judging identity (Ellis, et al., 1979; Young, et al., 1986), having salient external cues might in itself aid identity but not expression recognition. To avoid this potential confound, faces were cropped to remove the external contours of the face.

4.2.2 Method – Study 1

4.2.2.1 Subjects

Twenty two participants (14 female; mean age, 20) took part in this experiment. All participants gave written informed consent and the study was approved by the ethics committee at the Department of Psychology, University of York.

4.2.2.2 Procedure

There were four stimulus conditions in this experiment: (1) *same-expression, same-identity* (2) *same-expression, different-identity* (3) *different-expression, same-identity* (4) *different-expression, different-identity.* These conditions were presented in both contrast positive and negative (Figure 4.1). Face stimuli were Ekman faces placed on a standard gray background and cropped to a standardised ellipse so the external features were removed (Chapter 2.1.7). The ellipse shape was held constant across identities to prevent the shape of the ellipse becoming a cue to the identity of the face. The bridge of the nose of each face was aligned with the fixation cross to prevent images moving around the visual field. Each trial consisted of 2 faces sequentially presented; these could be either male or female but gender was constant across a trial. Each face was presented for 900 ms and separated by a gray screen presented for 300 ms. Trials were separated by a 2.5 s fixation gray screen during which participants had to judge whether the identity or expression was the same/different (2AFC). Each condition was presented 20 times in a counterbalanced order, giving a total of 160 trials. The experiment was run in two parts; in one part participants matched expression, in the other identity. Both parts were identical in terms of the presented stimuli, so any difference in expression and identity judgments were due to the task rather than a spurious effect related to the presentation order of the stimuli or the stimuli themselves. Order of judgements was counterbalanced across participants.

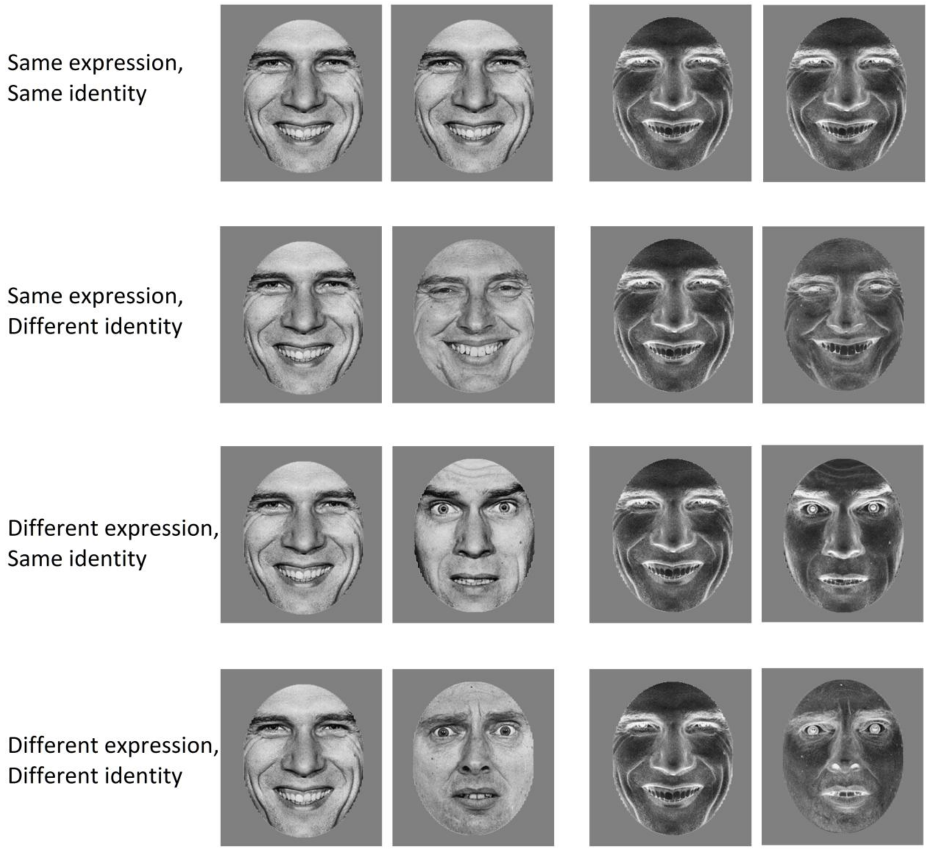

Figure 4.1 *Example trials for each of the conditions in Study 1 (Experiment 4). Each condition was presented in both contrast positive and negative.*

4.2.3 Results – Study 1

Figure 4.2 shows accuracies and reaction times for participants' judgements of identity and expression when stimuli were presented in both positive and negative contrast. Participants' responses were collapsed across conditions to give a single measure of expression and identity judgement performance for the positive and the negative conditions. A full break-down of results can be seen in Figure A.2 and Figure A.3 (Appendix).

The results reveal an increase in error rate when matching faces in negative compared to positive contrast. This effect of contrast negation was greater for judgements of identity than expression. A 2 x 2 ANOVA with the factors Judgement (expression, identity) and Contrast (positive, negative) revealed no significant main effect of Judgement ($F_{(1,15)} = 1.36$, $p = 0.26$), but there was a significant effect of Contrast ($F_{(1,15)} = 76.08$, $p < 0.001$)

and a significant interaction between Judgement x Contrast ($F_{(1,15)} = 8.75$, $p = 0.001$). The interaction was driven by significantly more errors for judgements of identity in contrast negative than positive ($t_{(15)} = 6.97$, $p < 0.001$), but no corresponding difference between judgements of expression in positive and negative contrast ($t_{(15)} = 2.05$, $p = 0.06$).

The reaction time data reveal a similar pattern of results. Although judgements of identity were quicker than expression, there was a greater increase in reaction time when judging identity in contrast negative compared to positive than there was for judgements of expression. A 2 x 2 ANOVA with the factors Judgement (expression, identity) and Contrast (positive, negative) revealed a significant main effect of Judgement ($F_{(1,15)} = 9.75$, $p = 0.01$), a significant main effect of Contrast ($F_{(1,15)} = 13.57$, $p = 0.002$) as well as a significant interaction between Judgement x Contrast ($F_{(1,15)} = 7.27$, $p = 0.02$). The significant interaction was driven by significantly longer reaction times when judging identity in contrast negative compared to contrast positive ($t_{(15)} = 3.80$, $p = 0.002$), but not for judgements of expression ($t_{(15)} = 1.46$, $p = 0.16$).

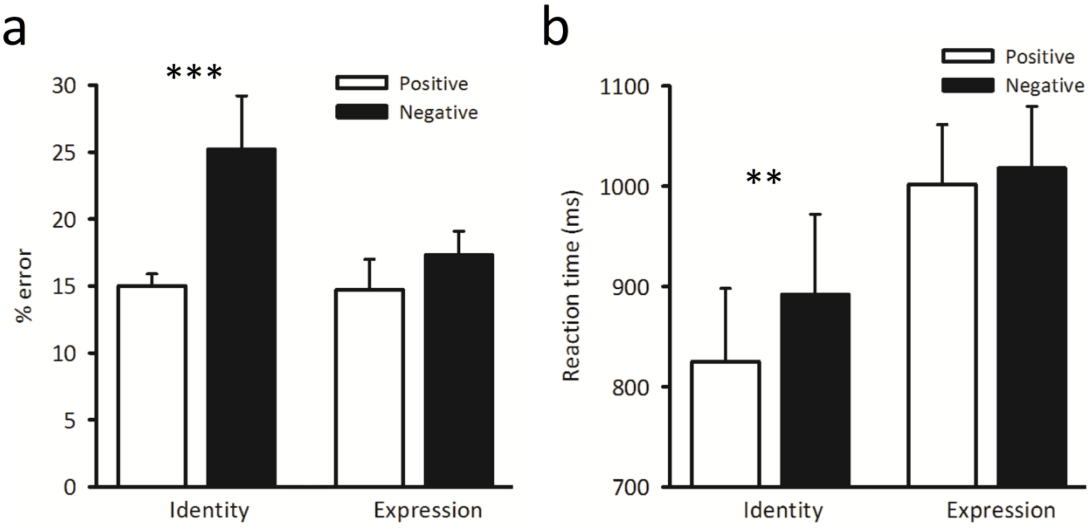

Figure 4.2 *Study 1 (Experiment 4) results. Participant's responses are collapsed across conditions to give a single response for expression and identity in positive and negative contrast. a. Percent error. b. Reaction time for correct trials. ** p < 0.01, *** p < 0.001*

4.2.4 Discussion – Study 1

This experiment aimed to validate the generated photo negated stimuli by asking whether matching expression and identity was impaired in photo negative. As photo negation affects the surface-based information whilst preserving the shape-based cues, it was predicted that matching expressions across two faces should be relatively unaffected by photo negation, whereas judgements of identity should be much less accurate. Using faces with the external features cropped, this study found that photo negation had a greater effect on judgments of identity compared to judgments of expression. These results suggest that shape-based information predominately contributes to the representation of facial expression whereas surface-based information primarily contributes to the representation of facial identity. This is consistent with previous studies that have used photo negation which have found reversing the contrast of faces has an adverse effect on judgements of identity whilst preserving the ability to match facial expressions (Bruce et al., 1991; Bruce & Langton, 1994; Galper, 1970; Kemp et al., 1996; Russell et al., 2006; White, 2001).

A second behavioural study was conducted to further probe the basis of the photo negation effect on judgements of expression and identity by introducing a large image change between successive faces. In Study 2, two faces were presented sequentially, but one face was presented in photo negative and the other in photo positive. The rationale was that despite the large image change across successive faces the shape-based information should remain constant. If expression is predominately based on this information then this manipulation should have little effect on judgements of expression. In contrast, this manipulation will markedly affect the surface-based information important in identity judgements and therefore judgements of identity should be less accurate in photo negative.

4.2.5 Method – Study 2

4.2.5.1 Subjects

Twenty two participants (17 female; mean age, 23) took part in this experiment. All participants gave written informed consent and the study was approved by the ethics committee at the Department of Psychology, University of York.

4.2.5.2 Procedure

There were four stimulus conditions in this experiment: (1) *same-expression, same-identity* (2) *same-expression, different-identity* (3) *different-expression, same-identity* (4) *different-expression, different-identity* (Figure 4.3). Stimuli were the cropped images used in Study 1. Faces were mounted on a gray background and the bridge of the nose was aligned with the fixation cross to prevent images moving around the visual field. Each trial consisted of 2 faces sequentially presented; these could be either male or female but gender was constant across a trial. In each trial one face was presented in contrast positive and the other in contrast negative. The presentation order of contrast positive and negative faces was counterbalanced across conditions. Each face was presented for 900 ms and separated by a gray screen presented for 300 ms. Trials were separated by a 2.5 s fixation gray screen during which participants had to judge whether the identity or expression was the same/different (2AFC). Each condition was presented 32 times in a counterbalanced order, giving a total of 128 trials. The experiment was run in two parts; in one part participants would match expression, in the other identity. Order of judgements was counterbalanced across participants.

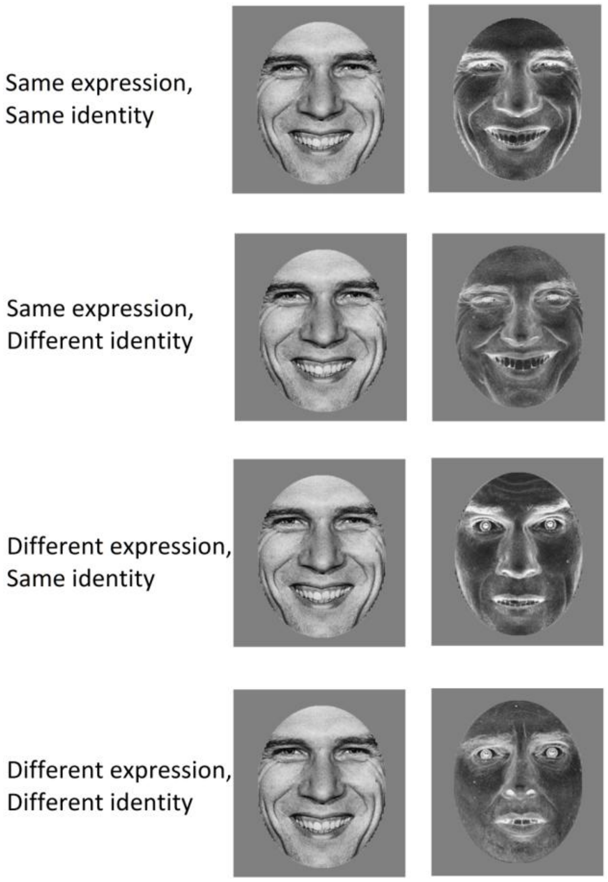

Figure 4.3 *Conditions and example trials used in Study 2 (Experiment 4).*

4.2.6 Results – Study 2

Figure 4.4 shows participants' judgements of identity and expression when stimuli were presented in both positive and negative contrast. Participants' responses were collapsed across the four conditions to give a single measure of performance for expression and identity judgements. A full break-down of results can be seen in Figure A.4 and Figure A.5 (Appendix).

The results revealed a significant increase in error when participants judged identity compared to judgements of expression ($t_{(16)} = 7.49$, $p < 0.001$). A similar pattern of results was found in participants' reaction times; judgements of identity took significantly longer than judgements of expression ($t_{(15)} = 2.22$, $p = 0.04$).

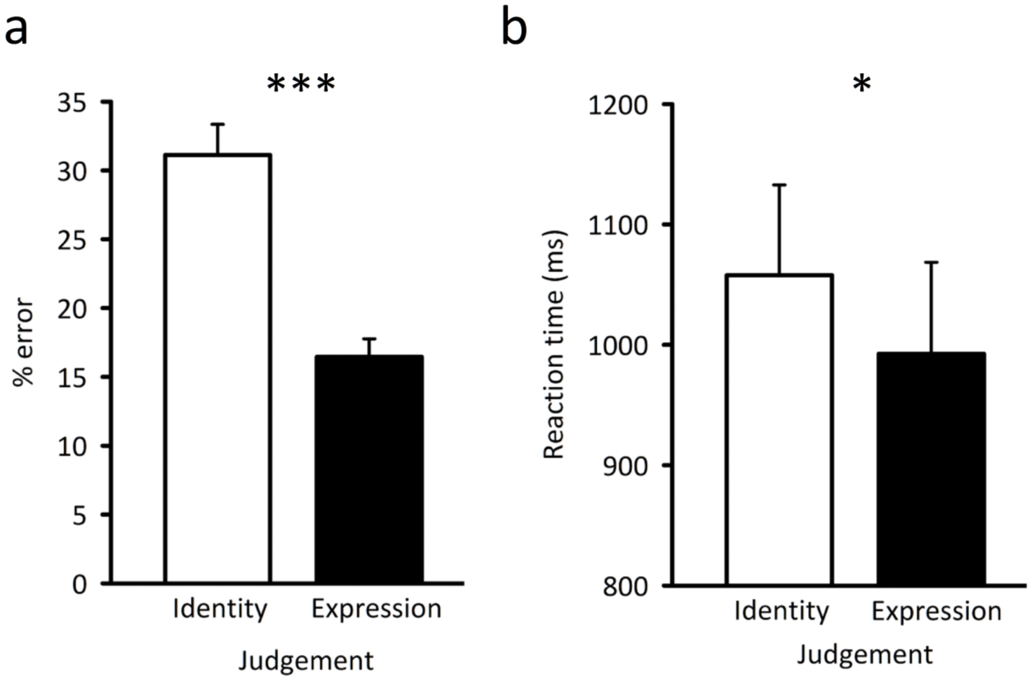

Figure 4.4 *Study 2 (Experiment 4) results. a. Percent error b. Reaction time. Participants' responses were combined across condition to give a single measure of performance for identity judgements and expression judgements. * $p < 0.05$, *** $p < 0.001$.*

4.2.7 Discussion – Study 2

This study aimed to further investigate the use of shape-based information when judging facial expression. Presenting faces in photo positive and negative introduces a large change between successive images in each trial. However, this change predominantly affects the surface-based information whilst the shape-based information is preserved. Despite the large image difference, participants remained relatively accurate when matching expression. The results therefore suggest that shape-based information is primarily used when matching expression. Conversely identity judgements were poor when images were presented in positive and negative, suggesting that these judgements are dependent on the disrupted surface-based information.

4.2.8 General Discussion

These behavioural experiments investigated the information used in judgements of facial expression. Across two studies photographic negation was used to disentangle the relative contribution of shape and surface-based information to the representation of facial expression. Study 1 revealed judgements of identity were significantly impaired by photographic negation; however, photo negation had little effect on judgements of expression. In a second study large image changes were introduced by presenting faces in photo positive and negative in a single trial. Despite the large change in image, judgements of expression were relatively spared. Conversely, judgements of identity were significantly less accurate compared to judgements of expression.

The pattern of light and dark across the face defines the shape of facial features and also provides cues to surface-based texture patterns and some information about the 3D structure of the face (Bruce & Humphreys, 1994; Bruce & Young, 1998, 2012). A broad distinction can be made between these shape and surface based cues and these cues are thought to contribute differently to the perception of facial expression. As facial expressions are conveyed by complex patterns of muscle movements (Ekman, 1972) the shape-based facial information is thought to predominately underlie judgements of expression. Conversely, judgements of facial identity are thought to involve the interpretation of the surface-based information (Bruce et al., 1991; Bruce & Young, 1998, 2012; White, 2001)

Photo negation has been used by previous studies in an attempt to disentangle the contributions of surface-based and shape-based cues to the representation of the face. These studies have mainly focused on identity judgements and found photo negation impairs these judgements (Kemp et al., 1996; Russell et al., 2006; White, 2001). One study has compared the effect of negation on both expression and identity and found that judgements of expression are relatively preserved in contrast negative compared to judgments of identity (White, 2001). The results from the behavioural studies reported here confirm and extend previous findings by demonstrating expression judgments are relatively spared by photo negation compared to identity and this holds true when a large image change is introduced.

The results reported here demonstrate the importance of shape-based information for the discrimination of facial expression. This is consistent with previous studies in which participants are asked to categorise the expression from line drawings of faces. Line drawings are devoid of surface-based cues and only provide shape-based information yet expression can be categorised relatively accurately from these drawings (Kirita & Endo, 1995; Magnussen et al., 1994; Mckelvie, 1973). Conversely judgements of identity are relatively poor from line drawings suggesting accurate identity recognition requires surface-based cues (Davies, et al., 1978; Rhodes et al., 1987). In contrast, the reverse is found when the shape information is degraded, which results in difficulties in the categorisation of expression but not identity (White & Li, 2006).

The importance of shape-based information for the categorisation of facial expression is demonstrated in study which adopted PCA. When restricting their PCA to the 10 most important components for expression categorisation, Calder et al., (2001) found that expression was most accurately categorised from shape-only faces compared to shape-free images, highlighting the role of shape information in expression categorisation. Furthermore, they also found relatively little overlap between the components useful for expression categorisation and the components useful for identity categorisation.

Taken together, the two behavioural experiments reported here found that judgements of expression are predominantly based on shape-based information whereas judgements of identity are more reliant on surface-based cues.

4.3 Experiment 5: Shape-based representations of faces in the pSTS

4.3.1 Introduction

Neuroimaging studies have found a network of regions in the occipital and temporal lobes that respond selectively to faces (Andrews & Ewbank, 2004; Kanwisher et al., 1997). However, it remains unclear as to how shape-based and surface-based information contribute to the neural representation of faces in these regions. Using photo negation and fMRI, this experiment aimed to investigate how shape-based and surface-based cues are represented in face-selective neural regions. To investigate this, the response to the same face was compared to different faces when faces were presented in either contrast positive or negative. It was predicted that the response to the same and different faces should be unaffected by photo negation in neural regions using a shape-based representation of the face. Conversely, contrast negation should have an effect on the neural response to faces in regions with a predominantly surface-based facial representation.

4.3.2 Method

4.3.2.1 Subjects

Twenty five participants took part in this experiment (16 females; mean age, 25 years). All participants were right-handed and had normal or corrected-to-normal vision. Visual stimuli (8˚ x 8˚) were back-projected onto a screen located inside the magnetic bore, 57 cm from participants' eyes. All participants provided written consent and the study was given ethical approval by the York Neuroimaging Centre Ethics Committee.

4.3.2.2 Face localiser scan

A separate face localiser scan (Version 2) was used to independently identify regions in each individual's brain that responded more to faces than non-face stimuli (see Chapter 2.2.3.2 for protocol).

4.3.2.3 Experimental scan

There were six conditions in this experiment: (1) *same-face, positive*; (2) *different-face, positive;* (3) *same-face, negative;* (4) *different-face, negative;* (5) *same-face, positive-*

negative; (6) *different-face, positive-negative* (Figure 4.6). Face stimuli were gray-scale Ekman faces selected from the Young et al FEEST set (2002) (see Chapter 2.1.7). Faces were cropped to remove external features, mounted on gray background and the bridge of the nose was aligned with the fixation cross to prevent images moving around the visual field. Stimuli were presented in blocks, with 6 images per block. In the *different face* conditions 2 different identities, each posing a different expression, were presented during a block of images. Each face was presented for 900 ms and separated by a gray screen presented for 150 ms. Stimulus blocks were separated by a 9 s fixation gray screen. Each condition was presented 8 times in a counterbalanced order, giving a total of 48 blocks. To ensure participants maintained attention throughout the experiment, participants had to push a button when they detected the presence of a red dot, which was superimposed onto 20 % of the images.

4.3.3 Results

The Localiser scan identified five regions of interest that responded more to faces than non-face stimuli. These regions, left and right FFA, left and right OFA, and the right pSTS are shown in Figure 4.5 and their locations are detailed in Table 4.1. Although face-selective, the left STS could not be consistently localised in each individual and was therefore not included in the analysis. Due to signal drop-out it was only possible to identify the amygdala in a small number of participants, therefore the data for the amygdala are not shown here but can be seen in Figure A.6 (Appendix).

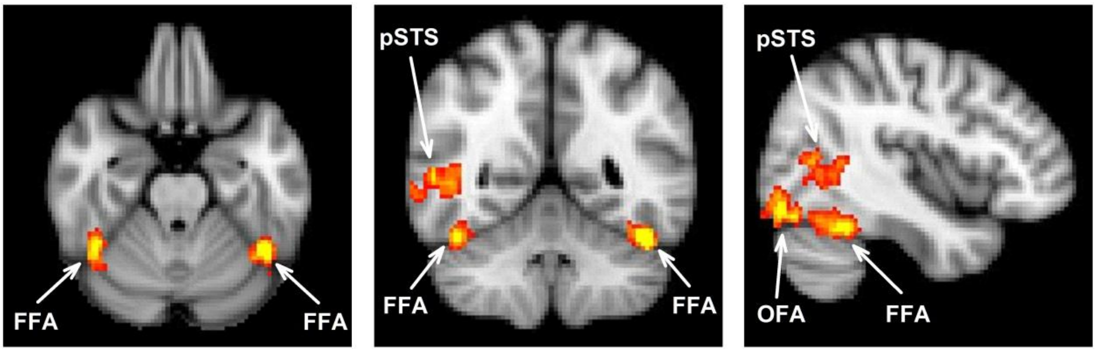

Figure 4.5 *Location of face selective regions in Experiment 5. Average location of each ROI across all participants and transformed into group space. All brain images are depicted in radiologic convention, i.e. coronal and axial slices are left/right reversed.*

A 2 x 2 x 3 x 2 ANOVA with the factors Region (FFA, OFA) Face (same, different) Contrast (positive, negative, positive-negative) and Hemisphere (right left) was conducted to determine whether the two hemispheres of the same region of interest responded differently. The pSTS was not included in this part of the analysis as it was only identified in the right hemisphere. The results revealed no significant interaction between Contrast x Region x Face x Hemisphere ($F_{(2,20)} = 1.411$, $p = 0.27$), nor an interaction between Region x Hemisphere ($F_{(1,10)} = .75$) or between Face x Hemisphere ($F_{(1,10)} = 0.07$) or between Contrast x Hemisphere ($F_{(2,20)} = 0.03$) nor was the a significant main effect of Hemisphere ($F_{(1,10)} = 0.001$). As there was no significant effect of hemisphere the timecourses were averaged across hemispheres resulting in three regions of interest; FFA, OFA, and right pSTS.

The peak responses of the face-selective regions were analysed using a 3 x 2 x 3 ANOVA with Region (pSTS, FFA, OFA), Face (same, different) and Contrast (positive, negative, positive-negative) as the factors. There were significant effects of Region ($F_{(2,30)} = 35.40$, $p < 0.0001$), Face ($F_{(1,15)} = 26.07$, $p < 0.0001$) and Contrast ($F_{(2,30)} = 8.242$, $p = 0.001$). There was also a significant interaction between Region x Contrast ($F_{(4,72)} = 4.14$, $p = 0.03$) as well as a significant interaction between Region x Face x Contrast ($F_{(4,60)} = 2.95$, $p = 0.03$). Therefore, to investigate which face-selective regions were sensitive to Contrast, and in what way each was sensitive to the different contrasts, the response in each individual ROI is now considered.

Table 4.1 *MNI coordinates (mm) of face-selective regions in Experiment 5. Coordinates reported are the centre of gravity of each ROI averaged across all participants and transformed into standard space. Standard error is reported in parenthesis.*

Region	*n*	*x*	*y*	z
FFA	23			
L		-43 *(1.2)*	-52 *(1.3)*	-25 *(1.6)*
R		41 *(1.0)*	-54 *(1.8)*	-21 *(1.3)*
OFA	24			
L		-44 *(2.2)*	-79 *(2.5)*	-11 *(1.5)*
R		44 *(1.6)*	-81 *(2.0)*	-14 *(1.2)*
STS	19			
R		54 *(1.9)*	-53 *(2.3)*	5 *(1.6)*

Figure 4.6 shows the response from the pSTS to the different conditions in this Experiment. A 3 x 2 ANOVA with the factors Contrast (positive, negative, positive-negative) and Face (same, different) was conducted to determine the sensitivity of the pSTS to these changes. This revealed a significant effect of Contrast ($F_{(1,18)} = 10.67$, $p = 0.004$) and a significant effect of Face ($F_{(1,18)} = 10.36$, $p = 0.005$). However, there was no significant Contrast x Face interaction ($F_{(1,18)} = 0.16$). The main effect of Face was due to significantly bigger responses to the *different-face* condition compared to the *same-face* condition for the *positive* ($t_{(18)} = 2.64$, $p = 0.017$), *negative* ($t_{(18)} = 3.14$, $p = 0.006$), and *positive-negative* ($t_{(18)} = 2.59$, $p = 0.018$) conditions. The significant main effect of Contrast was driven by a significantly bigger response to the *positive* compared to the *negative* condition for both the *same face* ($t_{(18)} = 2.54$, $p = 0.020$) and the *different face* ($t_{(18)} = 3.44$, $p = 0.003$) conditions.

In the FFA (Figure 4.6), there was a significant main effect of Contrast ($F_{(2,44)} = 17.91$, $p < 0.0001$) and Face ($F_{(1,24)} = 19.39$, $p < 0.0001$). There was also a significant interaction between Contrast x Face ($F_{(2,48)} = 2.54$, $p = 0.03$). There was a significantly bigger response to the *different face* compared to the *same face* in the *positive* ($t_{(22)} = 6.09$, p <

0.0001) but not the *negative* ($t_{(22)} = 1.96$, p = 0.06) or the *positive-negative* ($t_{(22)} = 0.33$) conditions.

In the OFA, there was a significant main effect of Face ($F_{(1,23)} = 13.82$, p = 0.001) and Contrast ($F_{(2,46)} = 11.20$, $p < 0.0001$). There was also a significant interaction between Face x Contrast ($F_{(2,46)} = 3.35$, p = 0.04). The interaction was due to differences between the *same face* and *different face* in the *positive* ($t_{(23)} = 2.70$, p = 0.013) and *negative* ($t_{(23)} = 3.19$, p = 0.004) but not in the *positive-negative* ($t_{(23)} = 0.38$) conditions (Figure 4.6).

These results demonstrate that the pSTS was insensitive to changes in contrast, showing a greater response to the different faces compared to the same face when the faces were presented in either positive, negative or positive-negative contrast. Conversely, the FFA was sensitive to changes in contrast and only showed a bigger response to the different face in the positive conditions. The OFA appears sensitive to the physical difference between face images. This region demonstrated a bigger response to the different compared to the same faces in both contrast positive and negative. However, the OFA shows an equivalent pattern of response to the same and different faces when the contrast varied.

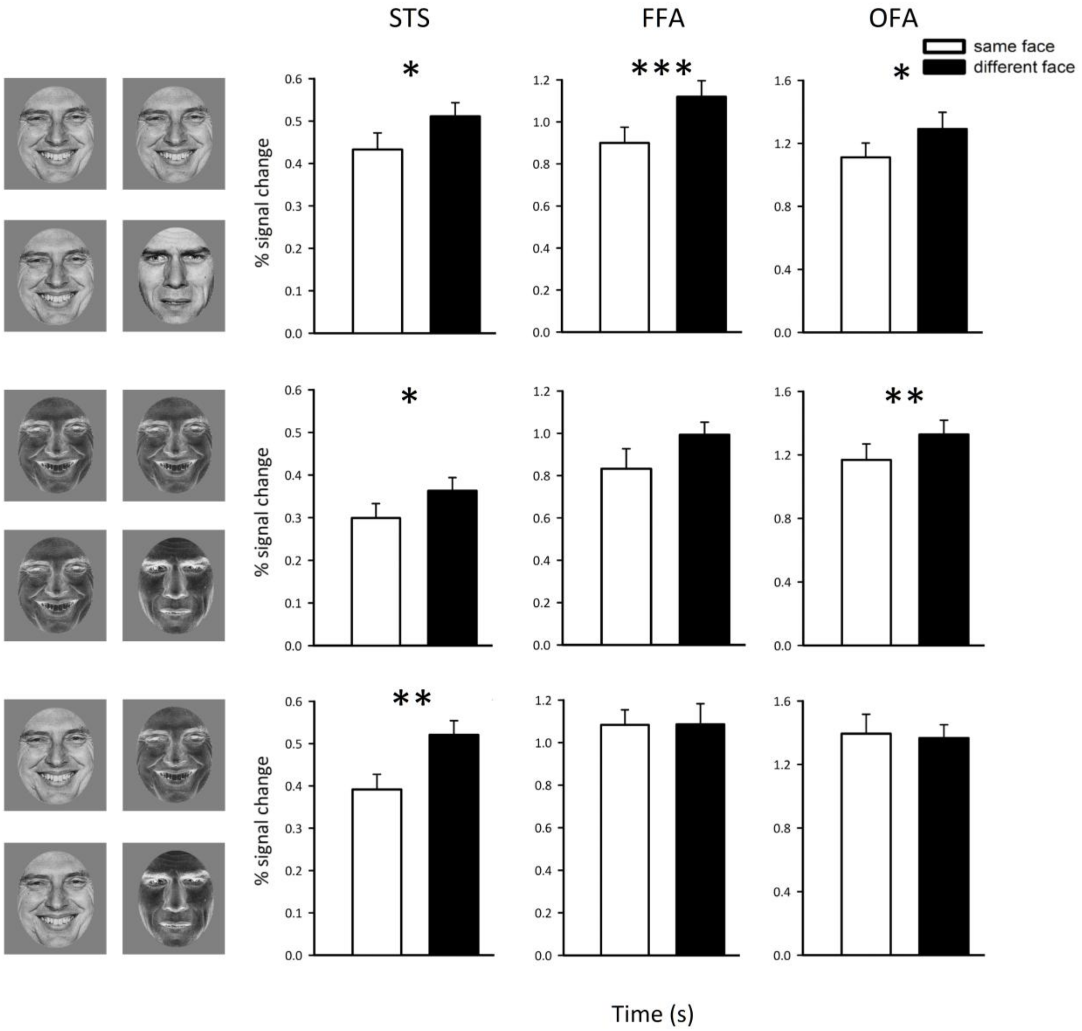

Figure 4.6 *Experiment 5 conditions and results. Example stimuli from the six conditions are shown. From top to bottom: (1) same-face, positive; (2) different-face, positive; (3) same-face, negative; (4) different-face negative; (5) same-face, positive-negative; (6) different-face positive-negative. Peak responses to the different conditions in the pSTS, FFA and OFA. Error bars represent standard error. * p < 0.05, ** p < 0.01, *** p < 0.001.*

4.3.4 Discussion

The aim of this study was to determine how the neural systems underlying the representation of facial identity and expression use different visual information. To address this question, photo-negation was used to reverse the pattern of light and dark across the image. The resulted demonstrated a dissociation between the response in face-selective regions to photo-negative faces. The selectivity of the response in the FFA was significantly attenuated by photo-negation. In contrast, the selectivity of the response in pSTS was not affected by photo-negation.

Facial expressions are conveyed by complex patterns of muscle movements which alter the shape of facial features (Ekman, 1972). As such it is the shape-based information that is thought of important in the categorisation of facial expression. This has been demonstrated in studies in which participants are fairly accurate at judging facial expression when only shape-based information is provided (Kirita & Endo, 1995; Magnussen et al., 1994; Mckelvie, 1973). This was also reflected in Experiment 4 which found that photo negation (which reserves the surface-based information but preserves the shape-based information) had little effect on judgements of facial expression.

Information regarding facial expressions, therefore, appears to be predominantly portrayed by the shape of facial features. It is therefore possible that neural regions sensitive to facial expression will code information from a face primarily from the shape of facial features. The STS is thought to be critical in the processing of facial expression (Haxby et al., 2000) and has been shown to respond more to facial expression than neutral faces (Narumoto et al., 2001), and more to changes in facial expression than facial identity (see Experiment 1, this thesis). This experiment found that the response in this pSTS was insensitive to photo negation, showing a bigger response to different faces across all three contrast conditions. The insensitivity to photo negation, which reverses the surface-based information, suggests that the pSTS uses primarily shape-based information in the neural representation of faces.

An interesting question that remains regards how the pSTS uses the shape-based information. Calder et al., (2000) demonstrated that the expression is categorised based on critical combinations of facial features. Having demonstrated the sensitivity of the pSTS to facial expression (Experiment 1) and that it represent faces using predominately shape-based cues, it would be interesting to extend these results by investigating whether the pSTS combines the different shape-based cues (i.e. the shape of the mouth and the shape of the eyes) in its representation of expression.

Shape-based information, however, is not the only information useful for facial expression categorisation. Texture information conveyed by the pattern of light and dark across the face can also be used to help recognise expression. For example, the nose wrinkle in a disgust face alters the pattern of light and dark across the nose. Texture information is also varied around the mouth and eye regions when different expressions are posed. The importance of texture information was shown by Calder et al., (2001) in which, using PCA, they found expression can be accurately categorised from shape-free texture information. However, despite the potential importance of texture-based information, the results from this experiment suggest that a region sensitive to facial expression, the pSTS, uses primarily shape-based cues to represent faces.

Recognition of facial identity is thought to be reliant on surface-based information. This is reflected in studies showing poor recognition of facial identity from line-drawings which do not provide surface-based information (Davies et al., 1978). Furthermore, photo negation which reverses the surface-based information present in the face results in poor recognition of facial identity (Gilad, Meng, & Sinha, 2009; Kemp et al., 1996; Russell et al., 2006; White, 2001; Experiment 4, this thesis). The FFA is thought to be involved in extracting the invariant features of a face that give rise to a person’s identity (Grill-Spector et al., 2004; Rotshtein et al., 2005; Yovel & Kanwisher, 2005). Here we used an fMR-adaptation paradigm to determine the sensitivity of the FFA to photo-negation. When the faces were all in positive contrast, we found a smaller response to repetitions of the same face image compared to the response to faces that were different in identity and expression. However, when the faces were all in negative contrast or alternated between positive and negative contrast, significant adaptation was no longer evident. Not only was there a reduction in sensitivity to differences between faces with negative contrast images, we also found that there was a lower response to negative contrast images. This finding is similar to other studies that have shown a reduced response in the FFA to negative contrast images (George et al., 1999; Nasr and Tootell, 2012; Yue et al., 2013). These neuroimaging results are also consistent with single neuron recordings from neurons in a face-selective region of the macaque inferior temporal lobe, which show a robust selectivity for appropriate contrast relationships in face image (Ohayon et al., 2012). Together, these results show that the neural representation in inferior temporal face regions is sensitive to patterns of contrast across the surface of the face. In other words, inferior temporal regions are sensitive to a factor that affects the ability to recognise identity behaviourally.

Interestingly, the results from this experiment reveal a dissociation between the types of facial information used by different face selective neural regions. The pSTS uses a more shape-based coding of faces whereas the representation of faces in the FFA is predominately driven by the surface-based information. These results therefore implicate the pSTS as a region sensitive to information that is predominantly used to categorise facial expression, whereas the FFA is sensitive to the cues that are used to categorise identity. This dissociation reflects that proposed in a prominent model of face processing (Haxby et al., 2000) in which the STS is thought to be part of a neural pathway responsible for the interpretation of the changeable aspects of a face such as expression. In contrast, the parallel route which involves the FFA is thought to be involved in processing the invariant features of a face such as the facial identity. These results offer further evidence for a distinction between these pathways showing that different information underpins the representations of faces along these pathways.

In conclusion, this study found that the pSTS was invariant to the contrast of the face images, showing a greater response to different face images compared to same face images regardless of the photo contrast format. As shape-based information is relatively preserved in photo-negative faces these results suggest that the pSTS primarily encodes facial information using shape-based information derived from the configuration of facial features. In contrast, the FFA only showed a bigger response to different compared to same face images when faces were presented in contrast positive, which suggests that it encodes facial information using surface-based information.

4.4 Conclusions

In Chapter 3 evidence for both categorical and continuous neural representations of facial expression was found. This chapter aimed to further explore the representation of facial expression by asking whether predominately shape-based or surface-based information is used in the neural representation of facial expression. In Experiment 4, photo negation was used to determine whether reversing the contrast of a face would have an impact on the ability to recognise facial expression. This experiment found that inverting the contrast of the face had relatively little effect on judgements of expression and thus suggesting that expression categorisation is based predominately on shape-based information. In Experiment 5, the use of fMRI and photo negation revealed that the contrast of the face had relatively little effect on the neural representation of faces in the pSTS. This suggests that, in the pSTS, faces are coded using predominately shaped-based information. As facial expressions are thought to be reliant on shaped-based information, these results are consistent with the role of the pSTS in processing facial expression that was identified in Experiment 1.

Chapter 5

The role of identity in the processing of facial expression

5.1 Introduction

Movements of the head or shifts in expression cause corresponding changes to the size and shape of the face image in an observer's eye. To be useful for recognition, the visual system should ignore these sources of variation and process invariant visual cues that indicate the identity of the face. However, the visual system must also use this information to detect changes that are important for social communication. Models of human face perception suggest that human observers deal with this problem using separate functional pathways, with the pathway involved in the visual analysis of identity being partially or fully independent of the pathway involved in processing changeable aspects of faces (Bruce & Young, 1986, 2012; Haxby et al., 2000).

Neuroimaging studies have identified a number of face-selective regions, which appear to provide support for the idea of separable visual pathways in face perception (Allison, McCarthy, Nobre, Puce, & Belger, 1994; Kanwisher et al., 1997): an occipital face area (OFA), a fusiform face area (FFA), and posterior superior temporal sulcus region (pSTS). The OFA is thought to be involved in the early perception of facial features and has a projection to both the pSTS and the FFA. The connection between the OFA and pSTS is thought to be important in processing dynamic changes in the face, such as changes in expression and gaze, which are important for social interactions (Calder et al., 2007; Engell & Haxby, 2007; Pelphrey, et al., 2004; Puce et al., 1998). In contrast, the connection between the OFA and FFA is considered to be involved in the representation of invariant facial characteristics that are important for recognition (Grill-Spector et al., 2004; Rotshtein et al., 2005).

Influenced by models of face perception, studies over the past decade have concentrated on the functional roles of each of these face-selective pathways (Andrews & Ewbank, 2004; Barton, Press, Keenan, & O'Connor, 2002; Fox et al., 2009; Hoffman & Haxby, 2000; Winston et al., 2004). Consequently, it has remained unclear whether there is any interaction between regions involved in perception of facial identity and expression.

The aim of this chapter was to probe the interaction of the pathways involved in processing the invariant and changeable aspects of faces. Experiment 6 asks whether the neural response to sequences of faces that change in expression and viewpoint direction differs according to whether the face images were of the same person or from different people. In Experiment 7, functional connectivity was used to determine whether there is an interaction between the different functional pathways involved in the perception of faces.

5.2 Experiment 6: Neural responses to expression and gaze in the posterior superior temporal sulcus interact with facial identity.

5.2.1 Introduction

Neural models of human face perception propose parallel pathways. One pathway (including pSTS) is responsible for processing changeable aspects of faces such as gaze and expression, and the other pathway (including the FFA) is responsible for relatively invariant aspects such as identity (Haxby et al., 2000). However, to be socially meaningful, changes in expression and gaze must be tracked across an individual face. This study asks whether the neural response to sequences of faces that change in expression and gaze direction differs according to whether the face images were of the same person or from different people.

5.2.2 Methods

5.2.2.1 Subjects

Data were analyzed from functional localiser scans from 103 different participants (49 females; mean age, 24), run as a standard part of six different fMRI experiment sessions to identify face-selective regions. Taking advantage of the large number of participants run on the same paradigm allowed the investigation of the general properties of identity in face processing and delivered substantial statistical power to the functional connectivity analysis (Experiment 7), which measured low-level interactions that might normally be swamped by stimulus-driven activity (main effects). All observers had normal or corrected-to-normal vision. Written consent was obtained for all participants and the study was approved by the York Neuroimaging Centre Ethics Committee.

5.2.2.2 Stimuli and procedure

The experiment used a block design with six different conditions: (1) *same-identity faces*, (2) *different-identity faces*, (3) *bodies*, (4) *inanimate objects*, (5) *places*, and (6) *scrambled images* of the former categories (Figure 5.1). Face images were taken from the Psychological Image Collection at Stirling (PICS; http://pics.psych.stir.ac.uk/). These images varied in viewpoint (frontal, ¾ view, profile) and expression (neutral, happy, speaking) within a block. The changes in viewpoint correspond to changes in gaze direction, which is often signalled in real life by movements of both eyes and head (Bruce & Young, 2012). The face images in each block therefore varied in both expression and gaze direction, but in one face condition the face identity was constant across the images in the block and in the other face condition identity varied across the block (see Figure 5.1A). Both male and female faces were used, but gender was held constant within a block. To determine low-level differences between the image properties in the two face conditions, the absolute difference in gray value across successive images was calculated. The correlation between corresponding pixel values in consecutive images was also calculated. Figure 5.2 shows the magnitude of the low-level change between successive images in the two conditions. There was no significant difference in absolute pixel values across the two conditions ($t = -1.66$, $p = 0.106$). However, there was a small but significantly higher correlation between successive images in the same identity faces condition compared to the different identity faces condition ($t = 3.52$, $p < 0.05$).

Examples of non-face stimulus conditions are shown in Figure 5.1B. Body images were taken from a collection at the University of Bangor (http://www.bangor.ac.uk/~pss811/), and contained clothed male and female headless bodies in a variety of postures. Images of places consisted of a variety of unfamiliar indoor scenes, houses and buildings, city scenes and natural landscapes. Stimuli in the object condition consisted of different inanimate objects including tools, ornaments, and furniture. Fourier-scrambled images were created by randomizing the phase of each two-dimensional frequency component in the original image, while keeping the power of the components constant. Scrambled images were generated from the images used in the other stimulus categories.

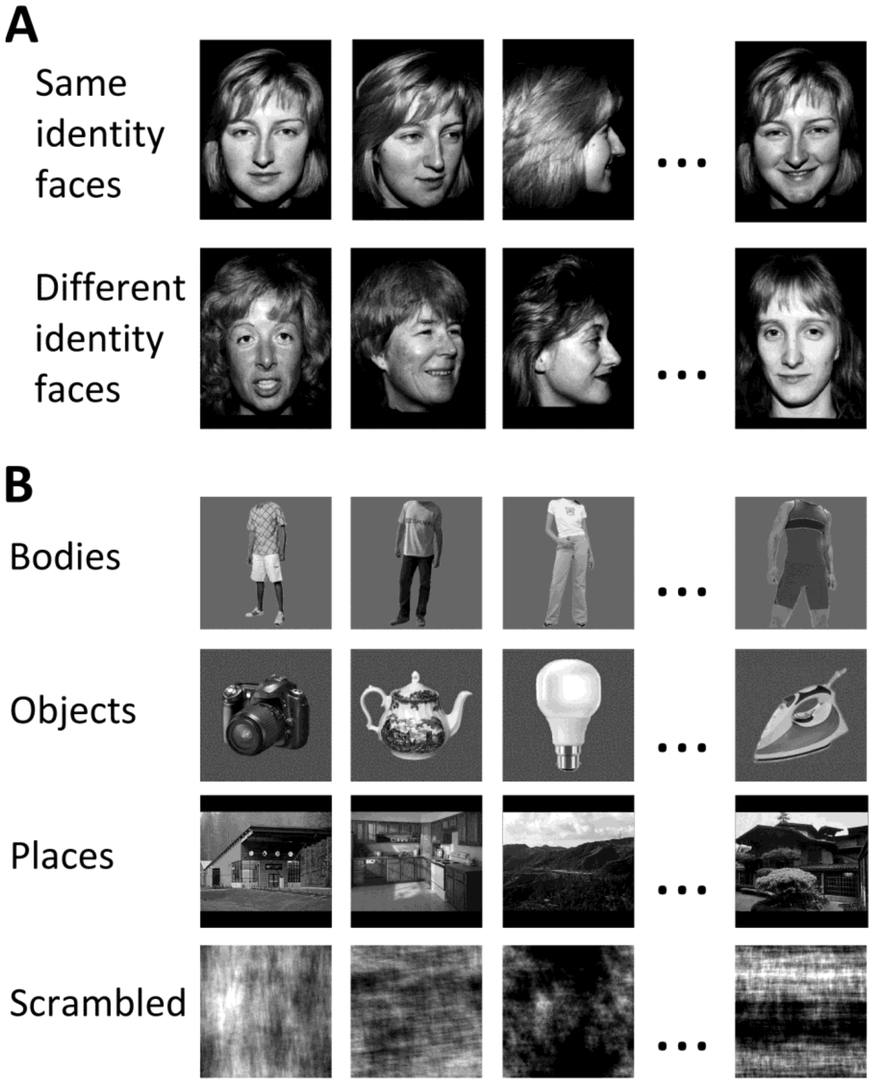

Figure 5.1 *Examples of the stimulus conditions. A. Face images varying in viewpoint/gaze direction and expression were presented with the same identity (1st row) or different identities (2nd row). B. Examples of the non-face stimulus conditions: bodies, objects, places and scrambled images.*

All images (approx. 8º x 8º) were presented in gray scale and were back-projected onto a screen located inside the bore of the scanner, approximately 57 cm from participants' eyes. Each block consisted of 10 images from a single stimulus condition; each image was presented for 700 ms and followed by a 200 ms blank screen, resulting in a total block length of 9 s. Stimulus blocks were separated by a 9 s gray screen with a central fixation cross. Each condition was repeated four times in a counterbalanced design resulting in a total scan length of 7.2 min. All participants viewed the same sequence of blocks and images.

Participants were required to monitor all images for the presence of a red dot that was superimposed on one or two images in each block. Participants were required to respond, with a button press, as soon as they saw the image containing the target. The target could appear in any location on the image, and was counterbalanced across conditions. There were no significant differences in the accuracy or reaction time during any of the experimental conditions. Mean detection accuracy was 96.9 % overall (same identity faces: 96.6 %, different identity faces: 97.5 %, non-face images: 96.9 %). A one-way ANOVA revealed no significant difference in accuracy across conditions ($F_{(2,200)} = 1.01$, $p = 0.37$). Mean reaction time was 441.4 ms overall (same identity faces: 438.7 ms, different identity faces: 441.6 ms, non-face images: 442.1 ms). A one-way ANOVA found no significant effect of condition on reaction times ($F_{(2,200)} = 0.51$, $p = 0.60$).

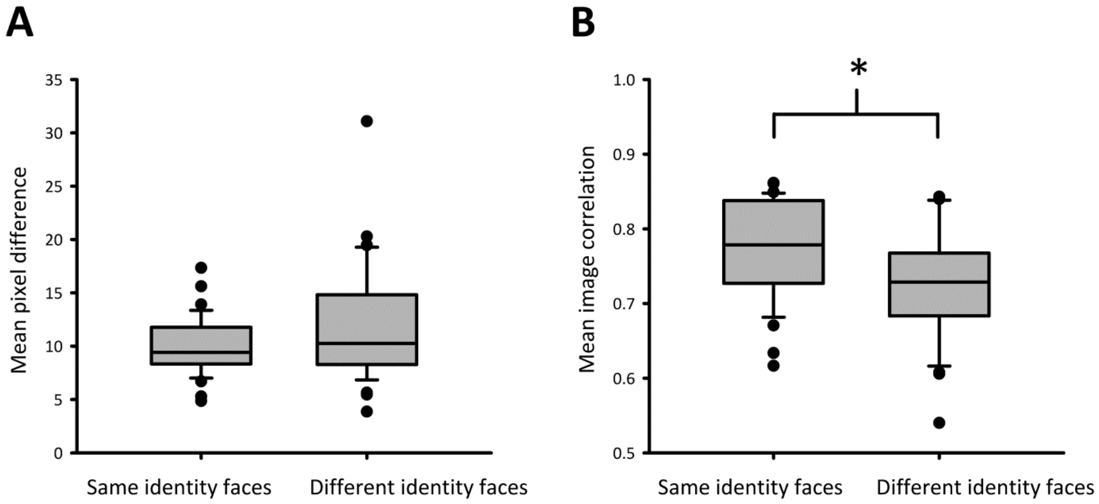

Figure 5.2 *Comparison of images statistics between face conditions. A. Mean pixel difference (RGB 0-255) between consecutive images for the same identity condition and the different identity condition. B. Mean image correlation between consecutive images for the same identity condition and the different identity condition. * p < 0.05.*

5.2.2.3 Imaging parameters

Imaging parameters are reported in Chapter 2.2.5.

5.2.2.4 Whole brain analysis

Statistical analysis of the fMRI data was carried out using FEAT in the FSL toolbox (http://www.fmrib.ox.ac.uk/fsl). The first three volumes (9 s) of each scan were removed to minimize the effects of magnetic saturation, and slice-timing correction was applied.

Motion correction was followed by spatial smoothing (Gaussian, FWHM 6mm) and temporal high-pass filtering (cut off, 0.01 Hz). Regressors for each condition in the GLM were convolved with a gamma hemodynamic response function. Individual participant data was then entered into a higher level group analysis using a mixed effects design (FLAME, http://www.fmrib.ox.ac.uk/fsl) whole brain analysis. To define face-selective regions, *same-identity faces* and the *different-identity faces* were compared to the responses from each of the non-face conditions (bodies, objects, places, scrambled), and the average of these contrasts was taken. To determine the effect of facial identity, the response from *same-identity faces* was compared with the response from *different-identity faces*.

5.2.3 Results

To determine the effect of facial identity, the response to sequences of faces that had the same identity was compared with sequences of faces that contained different identities (see Figure 5.1A). Figure 5.3A shows regions that had significantly different responses to the same identity faces condition compared to the different identity faces condition. Regions that responded more to the same identity faces condition are shown in red and regions that showed a greater response to different identity faces are shown in blue. The data have been thresholded to a value of $p < 0.00001$ (uncorrected) (Z-value > 4.2) in order to highlight regions of interest. A region within the right pSTS showed a significantly greater response to same identity faces compared to different identity faces. In contrast, there was a region in the fusiform gyrus that responded more to different identity faces compared to same identity faces. The coordinates of these regions are shown in Table 5.1 and Table 5.2 The volume of active voxels is also shown for each region, thresholded both at $Z > 4.2$ ($p < 0.00001$, uncorrected) and at $Z > 4.6$ ($p < 0.05$, resel corrected for multiple comparisons).

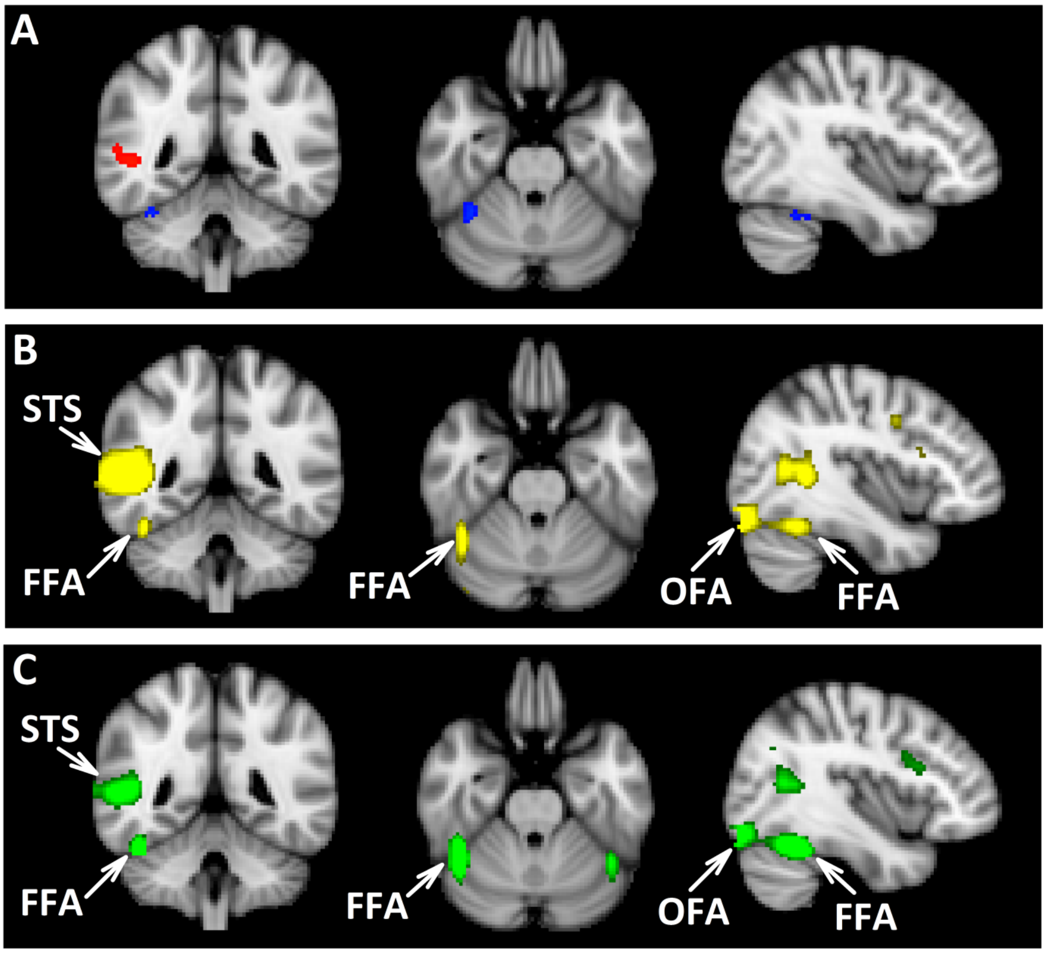

Figure 5.3 *Whole brain analysis (n=103). A. Regions showing greater response to same identity faces compared to different identity faces (red), and to different identity faces compared to same identity faces (blue). B. Activation to same identity faces compared to non-face stimuli (places, objects, bodies and scrambled images). C. Activation to different identity faces compared to non-face stimuli. Face selective regions are labelled: FFA (fusiform face area), OFA (occipital face area) and STS (superior temporal sulcus). All brain images are depicted in radiologic convention, i.e. coronal and axial slices are left/right reversed. The MNI coordinates (mm) of slices shown: x = 40, y = -46, z = -26. Statistical maps were thresholded at Z > 4.2 (p < 0.00001, uncorrected).*

To determine the spatial relationship between the regions shown in Figure 5.3A and face-selective regions such as the pSTS and FFA, each face condition (same and different identity faces) was compared with all the different non-face conditions (Bodies, Objects, Places, Scrambled – Figure 5.1B). Figure 5.3B reveals the location of face-selective regions in the occipital and temporal lobes defined by contrasting the same identity face condition with the non-face conditions (yellow). Figure 5.3C shows the regions defined by contrasting the different identity face condition with the non-face conditions (green). Both of these contrasts reveal a very similar pattern of face-selective regions that includes the bilateral OFA, the bilateral FFA and the right pSTS. It is also clear that the location of the face-selective right pSTS and right FFA in Figure 5.3B and Figure 5.3C (Table 5.3 & Table 5.4) correspond closely with the right pSTS and fusiform gyrus regions shown in Figure 5.3A.

Table 5.1 *Location of regions that show a greater response to same identity faces compared to different identity faces. Coordinates refer to the center of gravity of each group of active voxels. Volume of active voxels is shown thresholded both uncorrected (p < 0.00001) and corrected (p < 0.05) for multiple comparisons.*

Region	*X*	*y*	*z*	Peak Z-score	Volume cm³ Z > 4.2 $p < 0.000013$ (uncorrected)	Volume cm³ Z > 4.6 $p < 0.05$ (corrected)
STS						
L	-42	-65	-3	4.59	0.09	-
R	54	-49	6	5.16	2.55	0.70
IFG						
R	55	18	-3	4.72	0.30	0.02

Table 5.2 *Location of regions that show a greater response to different identity faces compared to same identity faces. Details as in Table 5.1.*

Region	*X*	*y*	*z*	Peak Z-score	Volume cm³ Z > 4.2 $p < 0.000013$ (uncorrected)	Volume cm³ Z > 4.6 $p < 0.05$ (corrected)
medial Occipital						
L	-9	-93	-3	5.64	1.43	0.86
FFA						
R	38	-49	-26	5.25	0.34	0.14
anterior Temporal						
L	-36	-18	-35	4.32	0.02	-
R	34	-16	-36	4.93	0.08	0.02

Table 5.3 *Location of the core face-selective regions defined by the contrast of same identity faces > bodies, objects, places and scrambled images.*

Region	*x*	*y*	z	Peak Z-score	Volume cm³
FFA					
L	-42	-56	-24	4.38	0.05
R	43	-53	-23	6.92	2.07
OFA					
L	-41	-85	-16	5.41	0.93
R	42	-80	-16	7.14	3.06
STS					
L	-49	-55	7	4.94	0.97
R	53	-51	8	8.49	19.90

Table 5.4 *Location of the core face-selective regions defined by the contrast of different identity faces > bodies, objects, places and scrambled images.*

Region	*x*	*y*	z	Peak Z-score	Volume cm³
FFA					
L	-42	-56	-25	5.67	0.77
R	43	-52	-23	8.23	2.98
OFA					
L	-41	-84	-17	5.56	0.92
R	40	-81	-16	7.06	2.23
STS					
L	-	-	-	-	-
R	52	-53	9	6.86	10.44

Other regions that showed a significantly greater response to same identity faces compared to different identity faces include the left pSTS and the right inferior frontal gyrus (IFG) (Figure 5.4, red). The coordinates of these regions are shown in Table 5.1. Regions that responded more to different identity faces compared to same identity faces are shown in Figure 5.4 (blue). In addition to the right FFA, significant responses were found in the anterior temporal lobe and in the medial occipital region (Table 5.2).

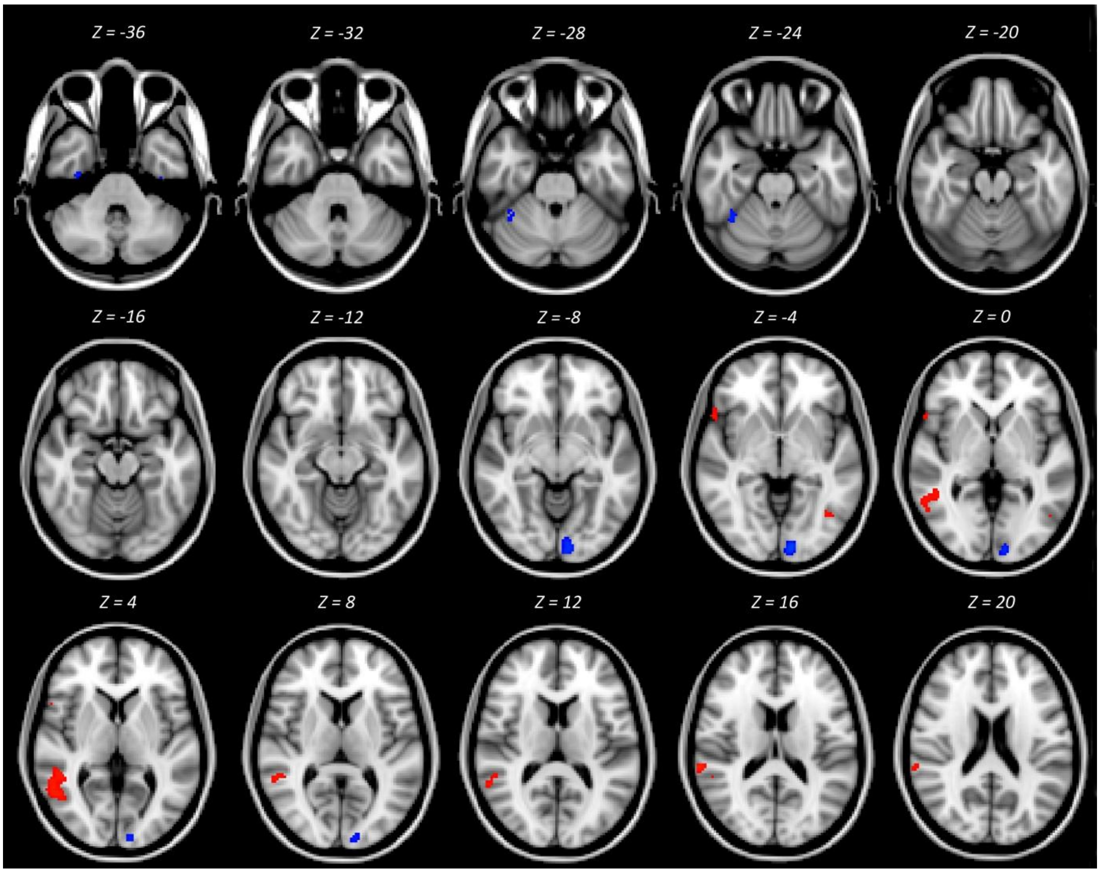

Figure 5.4 *Regions showing a greater response to same identity faces compared to different identity faces (red), and to different identity faces compared to same identity faces (blue). Axial slice number in MNI coordinates (mm) indicated above each slice. Statistical maps were thresholded at* $Z > 4.2$ *(*$p < 0.00001$*, uncorrected).*

5.2.4 Discussion

The posterior STS is a region that is known to respond to changes in facial expression and gaze direction. The aim of this experiment was to establish whether this region is also sensitive to the identity of a face. This experiment found that the face-selective region within the pSTS responded preferentially to sequences of face images that had the same identity compared to sequences of face images with different identities.

Because of the considerable importance attached to different types of facial information, the most efficient way to analyse this information is thought to involve different neural subcomponents that are optimally tuned for particular types of facial signal (Bruce & Young, 1986; Haxby et al., 2000). For example, models of face perception suggest that the analysis of facial identity occurs largely independently of the processing of changeable aspects such as expression. However, not all lines of evidence support a total separation of identity and expression (Bruce & Young, 2012; Calder & Young, 2005). For example, studies have shown that the perceptual effects of adaptation to emotional expression are more pronounced if the adapting and test expressions are from the same person (Campbell & Burke, 2009; Ellamil et al., 2008; Fox & Barton, 2007). These studies fit with other behavioural results that have shown that the ability to judge expression can be influenced by concomitant changes in identity (Schweinberger & Soukup, 1998). A corresponding effect of identity on the processing of facial expression was also shown in later ERP studies (Martens, Leuthold, & Schweinberger, 2010). Further support for the idea that the pathways involved in the perception of identity and expression may not be completely independent can be found in the way the image statistics of the face vary with changes in expression and identity. Principal Components Analysis (PCA) has shown that some of the principal components associated with changes to the face are associated with changes in identity or expression, but others reflect changes in both identity and expression (Calder et al., 2001)

These results show that there is a less than perfect separation between the neural representations used for identity and expression perception. There are two distinct reasons why this might be the case. One possibility might be that the response to facial identity in the face-selective pSTS could reflect that this region can represent the invariant aspects of a face that are necessary for the perception of identity. However, an equally plausible alternative is that the response to facial identity arises through interactions of pSTS with other face-selective regions that are associated with an analysis of invariant aspects of faces (such as identity). These possible explanations are addressed with functional connectivity in Experiment 7.

In conclusion, the analysis of a large dataset revealed a preferential response in the pSTS to a sequence of images that varied in expression and viewpoint when the face identity was the same compared to when it differed.

5.3 Experiment 7: Interaction of the pSTS and other face-selective neural regions when processing the changeable aspects of a face

5.3.1 Introduction

In Experiment 6 the pSTS responded preferentially to faces that varied in expression and viewpoint with the same face identity compared to different face identities. Two possible explanations were offered for this finding: (1) the pSTS is involved in the processing of the invariant features of the face, (2) The pSTS interacts with regions responsible for extracting facial identity.

Neuroimaging studies tend to focus on the functional properties of face-selective regions. Consequently, it has remained unclear whether there is any interaction between regions involved in perception of facial identity and expression. Support for the possibility of interactions between the pSTS and other face-selective regions has come from recent studies that have reported functional connectivity between the pSTS and FFA (Ethofer, Gschwind, & Vuilleumier, 2011; Turk-Browne, Norman-Haignere, & McCarthy, 2010; Zhang, Tian, Liu, Li, & Lee, 2009). Although the exact role of this interaction between the two more heavily-investigated pathways is not known, one possibility is that, to process changeable aspects of faces in a socially meaningful way, it is important to track changes across the same identity. Support for this possibility has come from studies that have shown an increased response to sequences of images with the same facial identity compared to sequences containing different facial identities (Andrews & Ewbank, 2004; Davies-Thompson et al., 2009).

The aim of this experiment, therefore, was to probe the interaction of the pathways involved in processing identity and changeable aspects of faces. Functional connectivity was used to determine whether the preferential response in the pSTS to the same face (Experiment 6) resulted from neural processes within the pSTS face-selective region itself, or whether it was dependent on interactions with other face regions. In the functional connectivity analysis, the stimulus driven activity from the fMR signal was removed and correlated the remaining or residual time-courses between face regions (Davies-Thompson & Andrews, 2012). This can be thought of as an extension of resting state connectivity in which correlations between regions, independent of a response to stimuli, are examined (Biswal, Yetkin, Haughton, & Hyde, 1995; Margulies et al., 2010). The aim of this experiment was to determine how the correlations between face-selective regions change as a function of whether the same or different identity faces were viewed (c.f. Norman-Haignere, McCarthy, Chun, & Turk-Browne, 2012).

5.3.2 Methods

Participants, stimuli, procedure, task and imaging parameters were the same as in Experiment 6.

5.3.2.1 Functional connectivity analysis

To measure the functional connectivity between regions, face-selective regions of interest (ROIs) were identified for each participant using the averaged contrasts of faces > bodies, faces > objects, faces > places, and faces > scrambled, thresholded at $p < 0.001$ (uncorrected). Regions identified included bilateral OFA, bilateral FFA and the right pSTS. Because the left pSTS was found in only a small minority of participants (about 20 %), this region was not included in the functional connectivity analysis. A control region, that was visually responsive but not face-selective, was also defined for each participant by transforming the anatomical ‘Occipital Pole’ region from the Harvard-Oxford Cortical Structural Atlas in the MNI standard brain into the participant’s functional data space (Desikan et al., 2006). This region included the occipital pole of both left and right hemispheres.

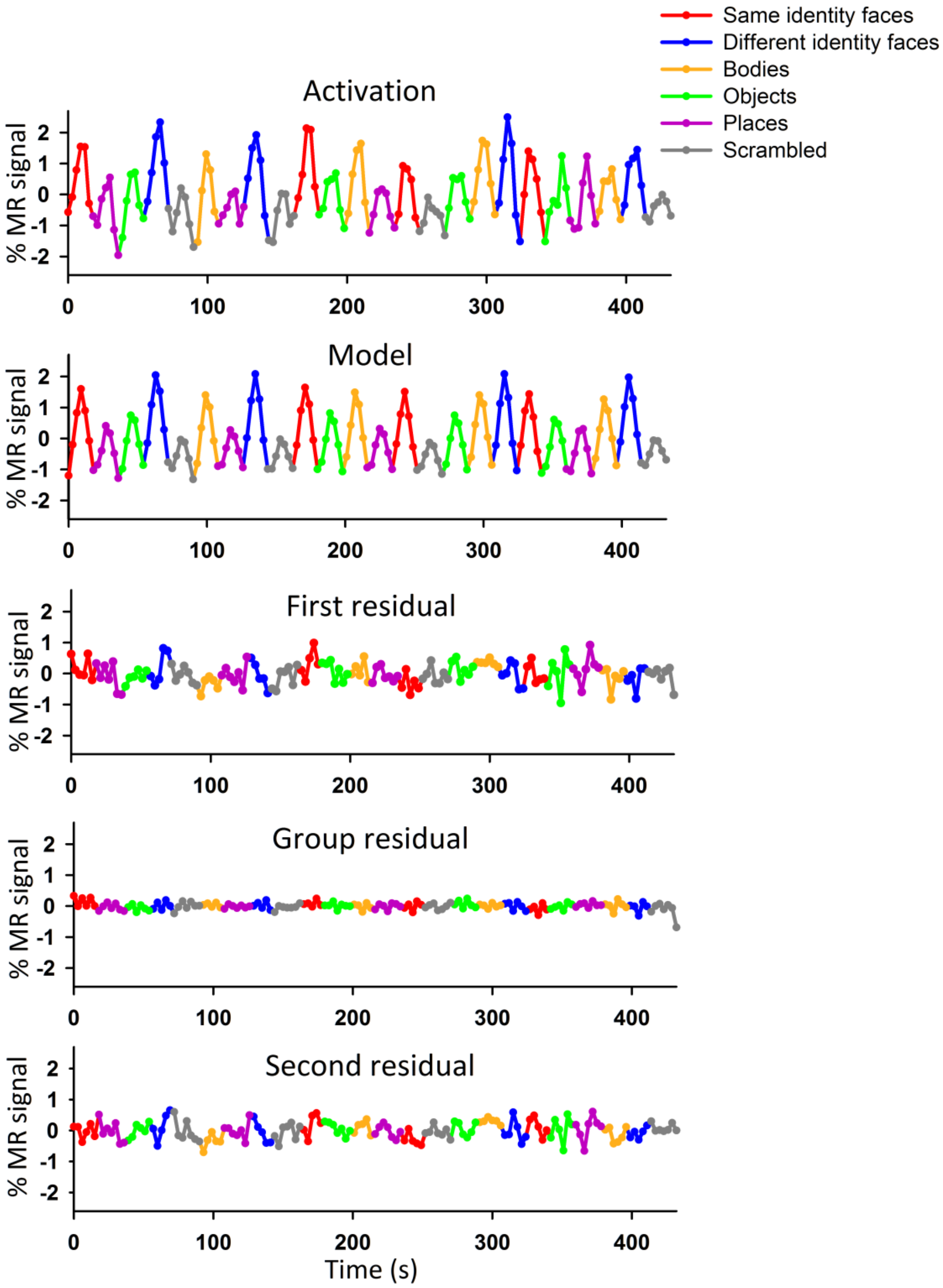

Figure 5.5 *Method for deriving residual activation within a region of interest for a single individual. Activation: Timecourse of activation within a region of interest. Model: General linear model for the timecourse. First residual: The data not explained by fitting the model to the timecourse. Group residual: Average of first residual data across 103 participants. Second residual: Data not explained by fitting the group residual to the first residual.*

In order to assess functional connectivity between regions, first any stimulus-driven activity was removed, as two regions will appear highly correlated if both are driven by the stimulus in parallel through a common input. As such, this analysis with stimulus-driven activity removed is orthogonal to the whole-brain GLM analysis. The stimulus-driven activity was removed through two steps (Figure 5.5): First, the stimulus-driven activation as modelled in the GLM analysis was removed, resulting in a residual time series response for each participant. Second, to capture any remaining stimulus-driven response that might not be fully accounted for by the hemodynamic model, the first residual time series response was averaged across all ROIs (left and right OFA, left and right FFA, right pSTS and OccP) and across all 103 participants. The rationale for combining across regions is that the average time-course of response was very similar across regions (Figure 5.6). The first-level analysis was then repeated with the average first residual response as an additional regressor. This gave rise to a second residual for each participant. The time-points corresponding to the same identity and different identity face blocks were then extracted for the functional connectivity analysis. Correlations between different regions were calculated using the extracted time-points for each participant for the same identity and different identity conditions. Mean correlations and standard error of the mean were calculated across participants for each condition and ROI pair. To test statistically whether the functional connectivity between regions was influenced by facial identity, correlations from each participant were converted to a normal distribution using a Fisher transformation, and then appropriate statistical tests (repeated measures ANOVAs and post-hoc t-tests) were performed.

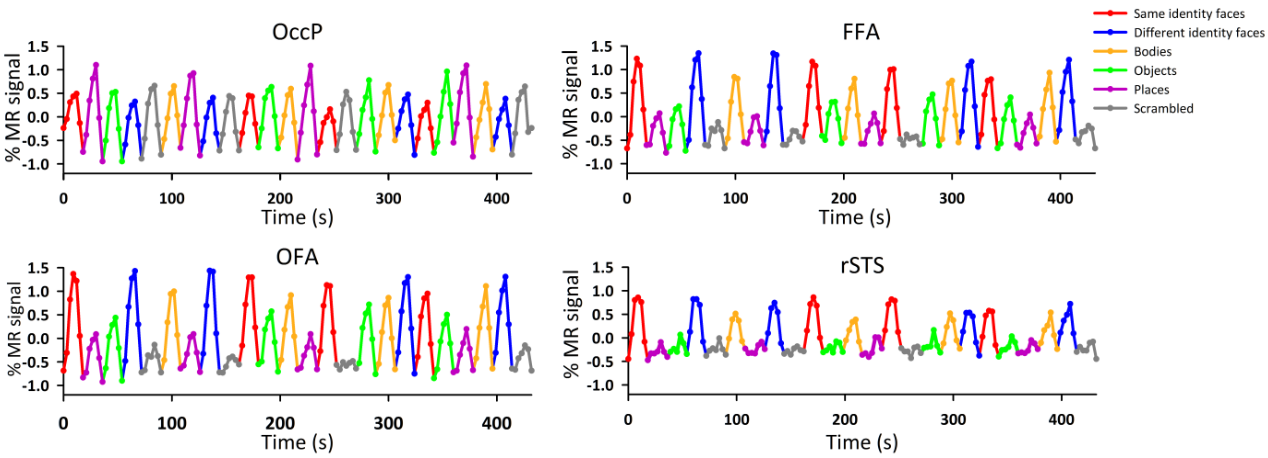

Figure 5.6 *Average timecourse of activation across 103 participants for four regions of interest (OccP: occipital pole, OFA: occipital face area, FFA: fusiform face area, rSTS: right superior temporal sulcus). OFA and FFA have been averaged across left and right hemispheres. Colours represent the different stimulus conditions within each block (see legend).*

5.3.3 Results

The core face-selective regions (left and right OFA, left and right FFA, right pSTS) were identified independently for each participant. The average time-course of response in these regions is shown in Figure 5.6. Next, it was determined whether the higher response to sequences of faces with the same identity in the pSTS was dependent on processing within this region or was dependent on interactions with other face-selective regions. Specifically, the functional connectivity between face-selective regions when viewing same identity faces and different identity faces was compared. To examine the functional connectivity between regions, the stimulus driven activity was removed (Figure 5.5). The residual time-courses were then correlated between pairs of face-selective regions (Figure 5.7). If the selectivity for the identity of faces in the pSTS is dependent on interactions with other face-selective regions, then a change in the correlations when viewing same identity faces compared to different identity faces would be predicted.

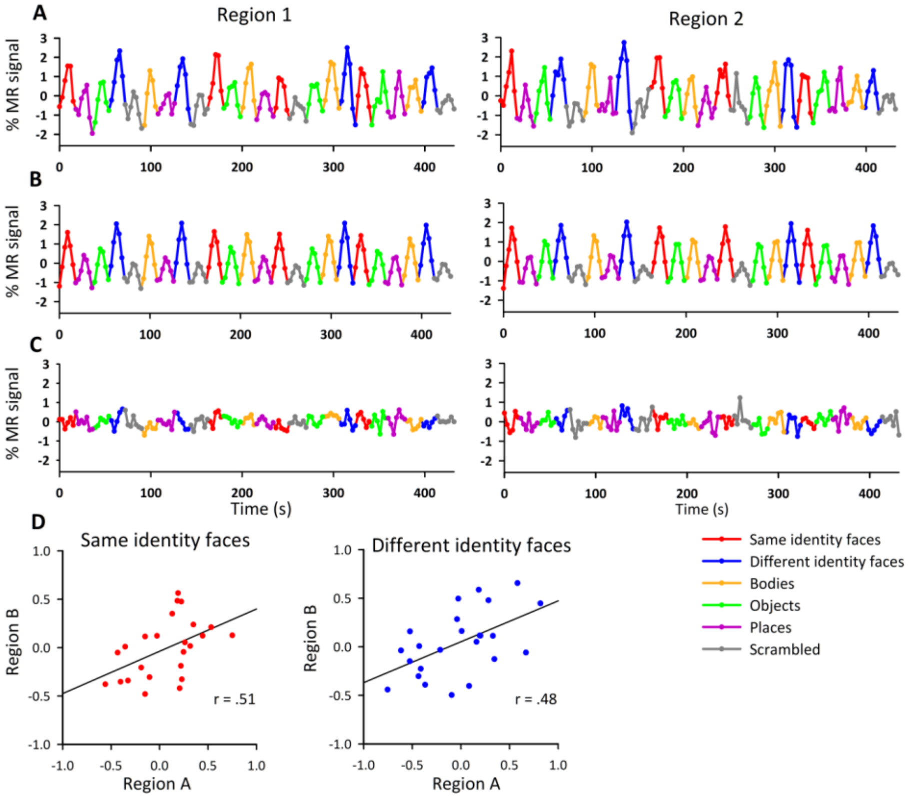

Figure 5.7 *Method for calculating functional connectivity between face-selective regions. A. Timecourse of activation for two regions of interest for a single participant (Left: Region 1, Right: Region 2). B. General linear model for the two regions. C. Residual timecourse of activation not explained by the general linear model. Colours represent the different stimulus conditions within each block (see legend). D. Correlations between region 1 and region 2 for time points within same identity faces (Left) or different identity faces (Right) conditions.*

Figure 5.8 shows the average correlations between the core face selective regions. For the bilateral ROIs (lOFA-lFFA and rOFA-rFFA), there was no significant interaction between hemisphere and condition ($F_{(1,71)} = 0.4$, $p = 0.54$). Therefore, correlations between OFA and FFA have been averaged across hemispheres for each participant. As the pSTS was found in most participants only on the right hemisphere, correlations reported between the pSTS and OFA or FFA refer to the right hemisphere only (i.e., rOFA-rSTS and rFFA-rSTS). The data show that there was an increased correlation between the STS and both the OFA and FFA when viewing same identity faces compared to different identity faces.

A 2 x 3 repeated measures ANOVA with Condition (same-identity faces, different-identity faces) and Connection (OFA-FFA, OFA-pSTS, FFA-pSTS) as factors revealed a significant effect of Condition ($F_{(1,78)} = 6.1$, $p < 0.05$), Connection ($F_{(2,156)} = 41.7$, $p < 0.001$) and a significant interaction between them ($F_{(2,156)} = 3.2$, $p < 0.05$). Post-hoc tests were then used to determine whether there were any significant differences in the correlations between regions for same identity faces or different identity faces. Significantly greater correlations were evident between the OFA and pSTS ($t_{(79)} = 2.7$, $p < 0.01$) and between the FFA and pSTS ($t_{(88)} = 1.9$, $p < 0.05$) for the same identity faces condition compared to the different identity condition. These results imply that the strength of the connections between the pSTS and other face-selective regions is influenced by the identity of the face. In contrast, there was no difference in the correlations for the same identity faces condition compared to the different identity condition between the OFA and FFA ($t_{(92)} = 0.16$, $p = 0.98$).

To validate the functional connectivity analysis, a separate analysis was preformed to ensure that all stimulus-driven activity was removed from the residual time-series before calculating correlations. Rather than calculating correlations between ROIs within participants, correlations in this control analysis were calculated between random pairs of participants e.g. FFA (participant1) – OFA (participant2). Unlike the positive values generated by the within-participant correlations (see Figure 5.8), control correlations across participants were close to 0 [Mean (standard error of the mean): FFA-OFA = -0.02 (0.03), OFA-STS = -0.05 (0.02), FFA-STS = -0.05 (0.02)].

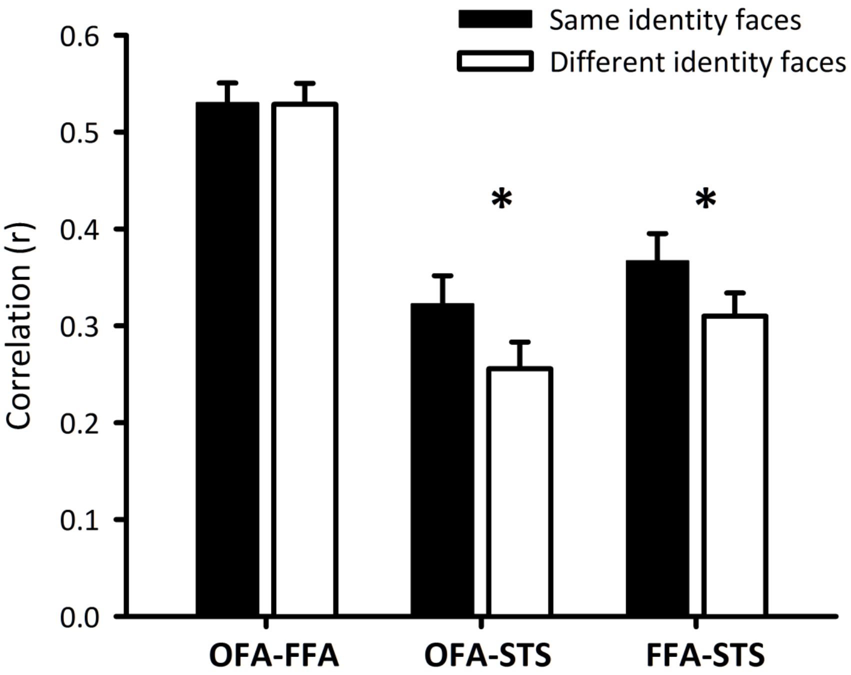

Figure 5.8 *Mean correlations (across participants) of residual activity between three face-selective regions of interest (OFA, FFA and STS) during the same identity faces and different identity faces conditions. Errors represent SEM across participants. * $p < 0.05$.*

To examine whether the effect of identity on connectivity with the pSTS is specific to face-selective regions, the correlations between a control region, the occipital pole (OccP), and each of the face-selective regions, OFA, FFA and STS was calculated. Figure 5.6 shows the mean time course of response in the OccP across all participants. The proportion of voxels in OccP that responded more to face compared to non-face conditions was 2.7 % at $p < 0.05$, uncorrected or 0 % at $p < 0.05$, corrected. This shows that the OccP was not responding selectively to faces. Figure 5.9 shows the average correlation values between the OccP and each face-selective region. A 2 x 3 repeated measures ANOVA with Condition (same-identity faces, different-identity faces) and Connection (OFA-OccP, FFA-OccP, STS-OccP) as factors revealed no significant effect of Condition ($F_{(1,85)} = 0.12$, $p = 0.73$), and no significant interaction between Condition and Connection ($F_{(2,170)} = 0.15$, $p = 0.86$).

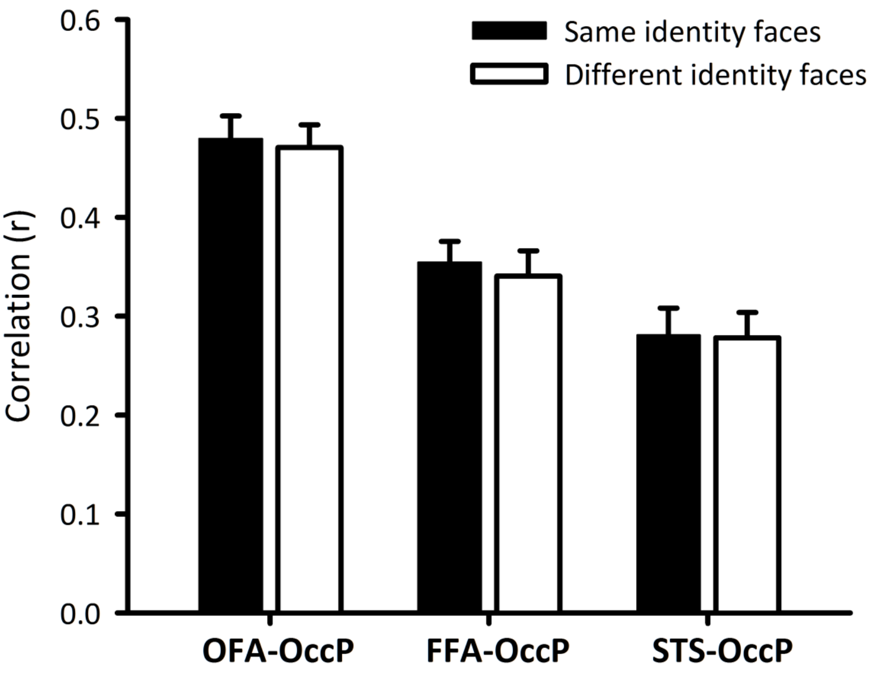

Figure 5.9 *Mean correlations (across participants) of residual activity between three face-selective regions of interest (OFA, FFA and STS) and the occipital pole region (OccP) during the same identity faces and different identity faces conditions. Errors represent SEM across participants.*

Finally, the face-selectivity of functional connectivity between regions was examined by comparing correlation values between regions during face (same identity, different identity) and non-face blocks (bodies, > objects, places and scrambled images). The correlations between the OFA-FFA were significantly higher during face blocks compared to non-face blocks in the right hemisphere ($t_{(83)} = 2.20$, $p = 0.03$), but not in the left hemisphere ($t_{(80)} = -0.09$, $p = 0.93$). Although correlations were not significantly higher during face blocks compared to non-face blocks for rOFA-rSTS ($t_{(79)} = 0.12$, $p = 0.90$) or the rFFA-rSTS ($t_{(88)} = 1.46$, $p = 0.15$), correlations were significantly higher during same identify face blocks compared to non-face blocks for both rFFA-rSTS ($t_{(88)} = 2.86$, $p = 0.005$) and rOFA-rSTS ($t_{(79)} = 2.07$, $p = 0.042$).

5.3.4 Discussion

This experiment used functional connectivity to investigate whether the preferential response in the pSTS to the same identity face (Experiment 6) was dependent on processing within this region, or whether it involved interactions between the pSTS and other face-selective regions. The analysis revealed increased functional connectivity between the FFA and pSTS when changes in facial expression and gaze occur across the same identity compared to when these changes in expression and gaze occur across different identities.

The selectivity of the increased connectivity with the OFA and FFA with the pSTS is shown by the absence of change in connectivity between the OFA and FFA and by the lack of any difference in the correlations between the face-selective regions and an early visual region. Moreover, the increased connectivity between the pSTS and the OFA/FFA for same identity faces was evident even though responses in the OFA and FFA were greater to different identity faces. This suggests that stronger connectivity is not necessarily dependent on the magnitude of the response at both ends of the connection (e.g. FFA and pSTS), but is instead due to the synergistic response between two regions carrying specific facial information relevant to social communication. The selectivity of this connection is further demonstrated by the fact that the correlations between the OFA/FFA and the STS were not increased to the different identity faces compared to the non-face conditions (see also, Davies-Thompson & Andrews, 2012). Indeed, these changes in patterns of functional connectivity could reflect a more general mechanism for category-selective interactions within the brain (see Norman-Haignere et al., 2012). Together, the results indicate that the response to identity in the pSTS is influenced by other face-selective regions that are involved in processing invariant aspects of faces that are important for the perception of facial identity.

To be socially meaningful, changes in expression and gaze direction must often be tracked across an individual whose invariant features (identity) remain constant. The preferential response in the pSTS to sequences of faces which vary in expression and gaze, but not in identity, is therefore consistent with the role of this region in social communication (Allison et al., 2000; Perrett, Xiao, Barraclough, Keysers, & Oram, 2009). However, other studies using event-related fMR-adaptation paradigms have reported an opposite pattern of results with greater responses to different identity faces compared to same identity faces (Fox et al., 2009; Winston et al., 2004). An explanation for the difference could be that the face images used in this study are presented in a sequence that show gradual changes in expression and viewpoint over time consistent with the changes that typically occur during social interactions. Moreover, these results are consistent with previous studies that have shown an increased response in the pSTS to sequences of images with the same facial identity (Andrews & Ewbank, 2004; Davies-Thompson et al., 2009) and with studies have found that response in the pSTS is greater to dynamic sequences of faces (Lee et al., 2010; Pitcher et al., 2011). In addition to the pSTS, the right inferior frontal gyrus was more active to same identity faces compared to different identity faces. Previous studies have shown that the rIFG is involved in face processing and that there is a functional connection between the rSTS and rIFG (Chan & Downing, 2011; Davies-Thompson & Andrews, 2012; Ethofer et al., 2011; Gschwind, Pourtois, Schwartz, Van De Ville, & Vuilleumier, 2012). Interestingly, the response of this region to changes in facial expression has been shown to be attenuated in autism (Dapretto et al., 2006). These findings suggest that signals that are important for social communication may be relayed to the frontal lobe from the pSTS.

Models of face perception suggest that the FFA is involved in processing the invariant features of a face that are important for recognition (Haxby et al., 2000). There was a greater response to sequences of different identity faces compared to same identity faces in the FFA. One explanation for the difference in response could be that the neurons in the FFA are invariant to changes in expression and gaze, but are sensitive to changes in identity. Thus, the same identity faces activate an overlapping population of neurons in the FFA that adapts with repetitive presentations (Grill-Spector, Henson, & Martin, 2006). In contrast, the different identity faces activate non-overlapping populations of neurons that do not adapt and consequently give rise to a greater response. Indeed, if the increased functional connectivity between the FFA and pSTS is conveying important information about identity, it would be necessary for the FFA to discriminate between the same and different identity conditions.

Interestingly, there was also significantly greater activation to different compared to same identity faces in the anterior temporal lobes (Figure 5.4, Table 5.2). These regions are known to contain image invariant representations of complex objects, including faces (Freiwald & Tsao, 2010; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Although these results are consistent with the idea that these regions contain image invariant representations of facial identity, this experiment found a region in the medial occipital lobe typically associated with processing low-level visual features that was also more responsive to different than same identity faces. This suggests that a greater variability in the image statistics across successive images (Figure 5.2) could have resulted in a greater response to the different identity faces condition in early visual areas that is then relayed to face-selective regions in the inferior temporal lobe.

In conclusion, this experiment offers a novel perspective on how the pSTS interacts with other face-selective regions when processing the changeable aspects of a face. This experiment found increases in functional connectivity between the pSTS and face-selective regions implicated in processing facial identity when viewing changes in expression and viewpoint across the same identity. This may reflect the importance of tracking the changeable aspects of the face across the same person.

5.4 Conclusions

This chapter aimed to investigate the interaction between facial expression and identity in the functional pathways that process this information. The results from Experiments 6 and 7 suggest that there is a less than perfect separation between the processing of facial identity and expression. Experiment 6 found that pSTS responded preferentially to sequences of faces that changed in expression and gaze, but did not change in identity, compared with similar sequences that changed in identity. Two distinct explanations of these findings were offered: (1) the pSTS is involved in processing the invariant features of a face. (2) The preferential response for the same identity was due to interactions of the pSTS with other face selective regions. Experiment 7 addressed these possible explanations using functional connectivity. This experiment found increases in functional connectivity between the pSTS and other face-selective regions, such as the FFA, that are implicated in processing facial identity. These results are consistent with the general claim that pSTS is involved in representing changeable aspects of faces (Haxby et al., 2000), but also offer a novel perspective on the neural processing in the pSTS, in which neurons in this region are particularly interested in changeable aspects of the same face. This may reflect the critical social importance of monitoring changes in a particular individual's gaze and expression, and demonstrate through connectivity analyses a potential mechanism through which this can happen.

General Discussion

Faces provide an abundance of cues that are useful for social interactions. Faces can tell us who a person is, what they are looking at and how they are feeling. Over the last 25 years there has been a wealth of research investigating how humans and monkeys perceive the different information available from a face (for a review see Schweinberger & Burton, 2011). The majority of this research has focused on how a face is recognised (Burton, Bruce & Hancock, 1999; Burton, Jenkins, Hancock & White, 2005) and where in the brain facial identity is processed and interpreted (for example, Grill-Spector et al., 2004; Rotsthein et al., 2005). In contrast, the understanding of how facial expressions are processed has been relatively understudied. Facial expressions are important, they inform us of how someone is feeling and they indicate potential threats within the immediate environment (Darwin, 1872; Ekman, 1975). The primary aim of this thesis was to investigate the neural representation of facial expression. Specifically, this thesis asked the following questions:

- Is facial expression extracted independently of facial identity?
- What regions of the brain are sensitive to facial expression?
- How is facial expression represented in these regions?
- What cues are used to extract the expression information?

To explore these issues, face stimuli were carefully selected to ensure that expressions of the same emotion were posed in a consistent way. Selecting stimuli in this way improves upon previous studies that often assume facial expressions of emotions are consistently posed. However, there can be significant variability in the muscle movements used to pose the same facial expression (Darwin, 1872; Ekman & Frissen, 1978). Stimuli used in Experiments 1-5 were taken from a well validated set of face images, from which five actors posing five different expressions were selected (POFA, Ekman & Friesen, 1978; selected from the FEEST set, 2002). So that facial expressions were well recognised and consistent exemplars of each emotion, stimuli selection was based on stringent criteria. Expressions were not only selected based on the visual similarity between examples of each emotion, but also based on the similarity of the muscles used to pose each expression (through FACS coding). The recognisability of the stimuli was confirmed in behavioural experiments reported in Chapter 2. Selecting stimuli in this way adds credibility to the results reported here by ensuring that expressions were well recognised and valid examples of each emotion.

6.1 What is the influence of facial identity on the processing of facial expression?

The most efficient way to process facial expression is thought to involve a dedicated neural subsystem optimally tuned to this type of facial signal. A neural pathway which involves the STS and amygdala is thought responsible for processing the changeable aspects of a face such as expression. This pathway is thought to code facial expressions relatively independently of the more invariant features of a face for example identity (Haxby et al., 2000). This reflects an earlier cognitive model which also proposed that facial expression is represented independently of the identity of the face (Bruce & Young, 1986). This approach has found support from initial cognitive and neuropsychological studies (Calder et al., 2000; Campbell et al., 1996; Young et al., 1986, 1993).

Despite the prominence of these models, very few studies have used neuroimaging techniques to directly compare, in the same experimental paradigm, the response in the STS and amygdala to facial expression and identity. Those studies which have, often report conflicting results and it remains unclear to what extent expression is represented independently of identity (Fox et al., 2009; Winston et al., 2004). In Chapter 3, three experiments directly compared the response in face-selective neural regions to facial expression and identity. Experiments 1 and 2 found that the response in the pSTS and amygdala was driven by changes in facial expression and was largely invariant to changes in facial identity. These results add further evidence in support of the above models of face processing. By including changes in both facial expression and identity these studies comprehensively demonstrate that the responses in the pSTS and amygdala are predominantly driven by facial expressions.

Further support for the involvement of the pSTS in the processing of facial expression relatively independently from facial identity comes from the type of facial information processed in this region. Experiments 4 and 5 used photo negation to reverse the pattern of light and dark across face images. This manipulation is thought to relatively preserve the shape-based information critical in processing facial expression whilst adversely affecting the surface-based information important in judgments of facial identity (White, 2001). This was reflected in Experiment 4 which demonstrated the perception expression was unaffected by photo-negation whereas judgements of identity were less accurate when faces were in contrast negative. Critically, Experiment 5 revealed that processing within the pSTS was also unaffected by photo-negation suggesting that this region is particularly sensitive to shape-based information that gives rise to the accurate interpretation of facial expression. This sensitivity to the shape, rather than the surface-based information in the pSTS provides strong evidence for the processing of facial expression within this region. Furthermore, these results provide support for Haxby et al., (2000) model of face processing by suggesting the pSTS is sensitive to expression information relatively independently of identity information.

However, although responses in the pSTS may be predominantly driven by expression, they need not be completely insensitive to identity. Physiological responses to and interpretation of a person’s facial expression are dependent on who they are (Ekman, 1975). In order, therefore, for facial expressions to be socially meaningful, information about who is posing the expression is also required. As such, at some point in the face processing network information about the invariant features of a face will inform the processing of facial expression. This is reflected in more recent behavioural work in which identity information influences judgements of facial expression (Campbell & Burke, 2009; Ellamil et al., 2008; Fox & Barton, 2007; Schweinberger & Soukup, 1998; Schweinberger et al., 1999).

In Chapter 5 evidence was found for an influence of identity on the neural representation of facial expression. Using a large data set, Experiment 6 found that the response in the pSTS was modulated by facial identity, with a greater response to changes in the variant features of a face across the same face identity compared to different identities. The cause of the identity processing was addressed in Experiment 7. Using connectivity analysis, Experiment 7 revealed that the modulation of pSTS to the same identity was driven by its connections with other face-selective neural regions, rather than processing within this region itself.

Taken together, the results reported in this thesis offer a resolution to whether facial expression is extracted by a dedicated neural subsystem. The idea of extracting facial expression entirely independently of the more invariant facial cues may be an oversimplification. Although the response in the pSTS seems to be predominantly driven by the changeable aspects of faces, this region was found to interact with other face-selective neural regions in a way consistent with a role in monitoring the social meaning of the expression information. As a region important in processing the social component of dynamic changes in the face (Allison et al., 2000), it appears the STS can track the meaning of the changes largely based on the invariant features of a face (i.e. identity). This is an interesting alternative to the seemingly conflicting positions of fully independent vs. fully interactive processing of facial expression and identity.

An interesting approach for future research would be to separately investigate the effect of facial identity on changes in facial expression and viewpoint. Experiments 6 and 7 investigate the effect of identity on the changeable aspects of the face and therefore changes in facial expression and viewpoint were combined. As the neural representation of expression and gaze is thought to have discrete but overlapping representations (Engel & Haxby, 2007) it would be interesting to explore whether both these representations are influenced by identity information.

6.2 How are facial expressions represented in the brain?

Two regions of the brain are thought to be of considerable importance in representing changes in facial expression; the pSTS and amygdala (Breiter et al., 1996; Haxby et al., 2000; Morris et al., 1996; Narumoto et al., 2001; Said et al., 2010). This was reflected in Experiments 1 and 2, in which the neural responses of face-selective regions to changes in facial expression and identity were compared. This comparison revealed that the pSTS and amygdala were selective to changes in facial expression.

Having identified these two neural regions as critical to the representation of facial expression, Experiments 2 and 3 aimed to explore how facial expression information is represented in these regions. The pSTS and amygdala have seldom been compared in the same experimental paradigm and as such it remains relatively unknown as to the relative contribution of these regions to the representation of facial expression. Historically, facial expressions have been thought to be represented either as belonging to discrete categories of emotion (Darwin, 1872; Ekman, 1999) or as continuous representations varying along graded dimensions (Russell, 1980; Woodworth & Scholsberg, 1954). There is evidence to support both approaches and despite this being a long-debated topic (see Ekman (1994) and Russell (1994) for reviews) little is known, at the neural level, as to whether the brain uses categorical and continuous representations of facial expressions.

This issue was addressed by directly comparing the responses in the pSTS and amygdala to categorical and continuous changes in facial expression (Experiments 2 and 3). Experiment 2 used morphed expression continua and found a dissociation between the representation of expression in the pSTS and amygdala. The pSTS was sensitive to any change in facial expression, which is consistent with a continuous representation of expression. In contrast, the amygdala was only sensitive to changes in facial expression that resulted in a change in the emotion category, which reflects a categorical representation of expression. This was confirmed in Experiment 3, which used more ecologically valid dynamic stimuli. By showing that different regions of the brain can have a primarily categorical or continuous representation of facial expression, these results offer a resolution to the controversy over how facial expressions are represented in the brain.

Interestingly, the representations of facial expression used by the pSTS and amygdala appear optimal for the type of information these regions are thought to extract. The amygdala is thought of as critical in processing information pertinent to survival (Sander et al., 2003; Whalen, 1998) and categorical responses to facial expression are optimal for making prompt and efficient responses to physiological threat. Conversely, a categorical response is not always appropriate and in everyday communication there is a need to discriminate different intensities of the same emotion. A continuous representation allows for these discriminations and reflects the proposed role of the pSTS in extracting the social component of the more changeable aspects of the face (Allison et al., 2000; Pelphrey et al., 2003).

The key finding here is that there is a dissociation between how the pSTS and amygdala represent facial expression. These two regions have been incorporated into a neural network dedicated for the processing of facial expression (Haxby et al., 2000). However, rather than these regions having a single unitary representation of facial expression, the results from Experiments 2 and 3 demonstrate that different neural regions can have primarily categorical or continuous representation of expression. Intuitively, the representations of expression that are used in these regions are optimal for the processing thought to occur within these regions.

6.3 What information present in the face is critical for facial expression perception?

A final question investigated in this thesis concerned the type of facial signal that is used to interpret facial expression. Information from a face can be broadly organised into two types: (1) surface-based information carried by the pattern of light and dark across the face, and (2) shape-based information conveying the configuration and relationship between the edges of facial features (Bruce & Humphreys, 1994; Bruce & Young, 2012). As facial expressions are recognised by critical combinations of facial features conveyed by complex movements of muscles (Calder et al., 2000), it is the shape-based information that is thought of as of primary importance when representing the facial expression (Kirita & Endo, 1995; Magnussen et al., 1994; Mckelvie et al., 1973; White & Li, 2006).

By using photo negation to disrupt the surface-based information whilst relatively preserving shape-based information, Experiment 4 found that judgements of expression were relatively unaffected by photo negation, suggesting that judgements of expression were based predominately on the relatively spared shape information. Although the results suggest that the shape cues are of primary importance when making judgements of expression, some expression cues are none the less conveyed by the surface-based information. For example opening the mouth in surprise or showing the teeth in a grin.

Experiment 5 provides a significant contribution to the literature by being one of the first reported studies to investigate the effect of photo negation on the neural representation of faces in the pSTS. This study found little effect of photo negation on the processing of faces in the pSTS, therefore suggesting that the pSTS predominantly uses shape information when representing faces. As facial expression is thought to be primarily conveyed by shape-based information the findings from Experiment 5 reflect the role of the pSTS in processing facial expression which was highlighted in Experiments 1-3.

An interesting way to extend these findings would be to investigate the effect of altering the shape-based information on processing in the pSTS. By blurring the face images to remove the high-spatial frequency edge-information, White and Li (2006) found that judgements of expression became significantly worse. It would be predicted that degrading the edge-based information would have an effect on the representation of faces in the pSTS. This could potentially provide further evidence of the use of shape-based information in this region.

6.4 Conclusions

This thesis aimed to provide a significant contribution to the understanding of the neural representation of facial expression. The most widely used neural model of face processing proposes a dedicated neural subsystem for the processing of facial expression which includes the STS and amygdala (Haxby et al., 2000). By directly comparing the responses in these regions to facial expression and identity, this thesis demonstrated that both the pSTS and amygdala are primarily involved in the representation of facial expression. These regions are usually thought to extract this expression information relatively independently from the more invariant features of a face. However, this independent processing approach appears to be an oversimplification. Instead, the pSTS through its connections with other face-selective neural regions, is indeed sensitive to changes in facial identity. This region shows a preference for the same facial identity, and it is postulated that this information is used in order to interpret the social meaning of facial expressions. Although the pSTS and amygdala have been previously implicated in the processing of facial expression, little has been reported as to how these regions contribute to the neural representation of expression. This thesis, therefore, provides a significant contribution to the literature by demonstrating a dissociation between the representation of facial expression in these regions; the response in the pSTS reflects a continuous representation of expression, whereas there was a more categorical response in the amygdala. Finally, it appears that shape-based information from a face is of primary importance to the representation of faces in the pSTS.

Appendices

7.1 Supplementary Figures

7.1.1 Chapter 2

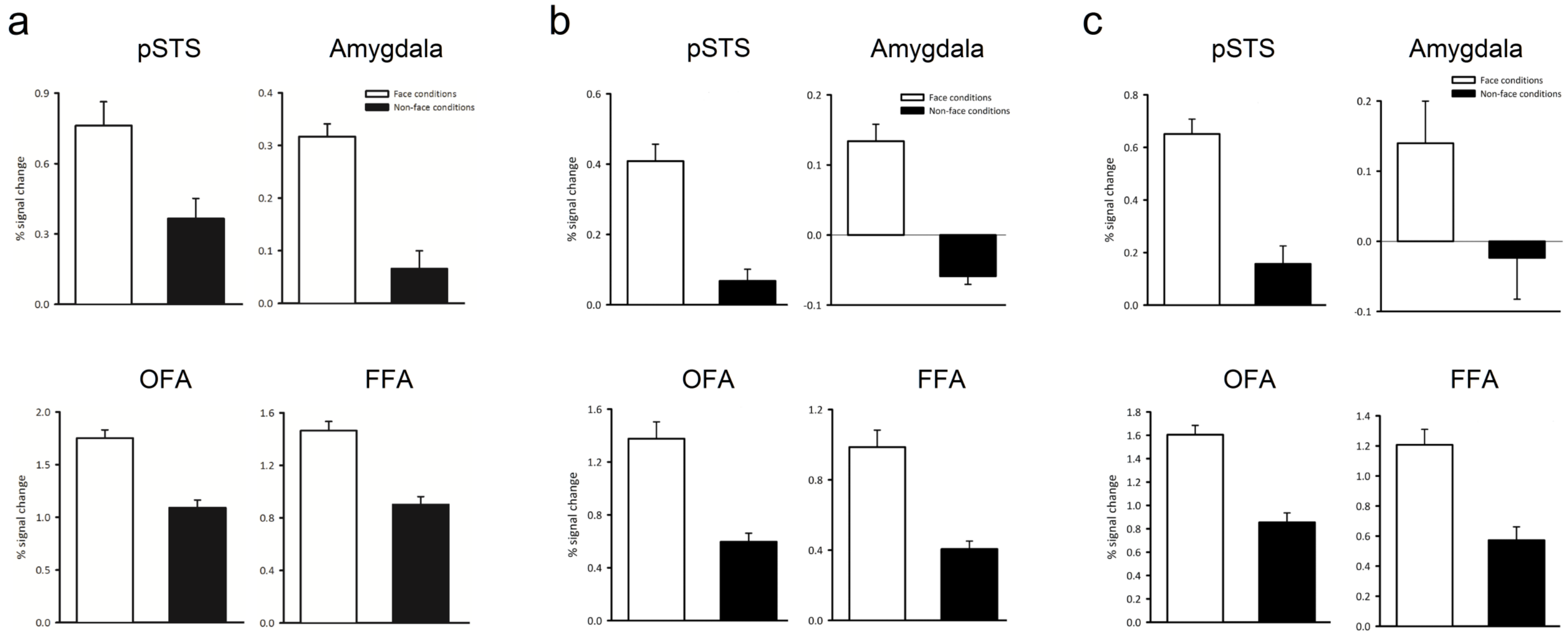

Figure A.1 *Responses in face-selective regions to blocks of faces and non-face objects during the localiser scan. a. Experiment 1. b. Experiment 2. c. Experiment 3.*

7.1.2 Chapter 4

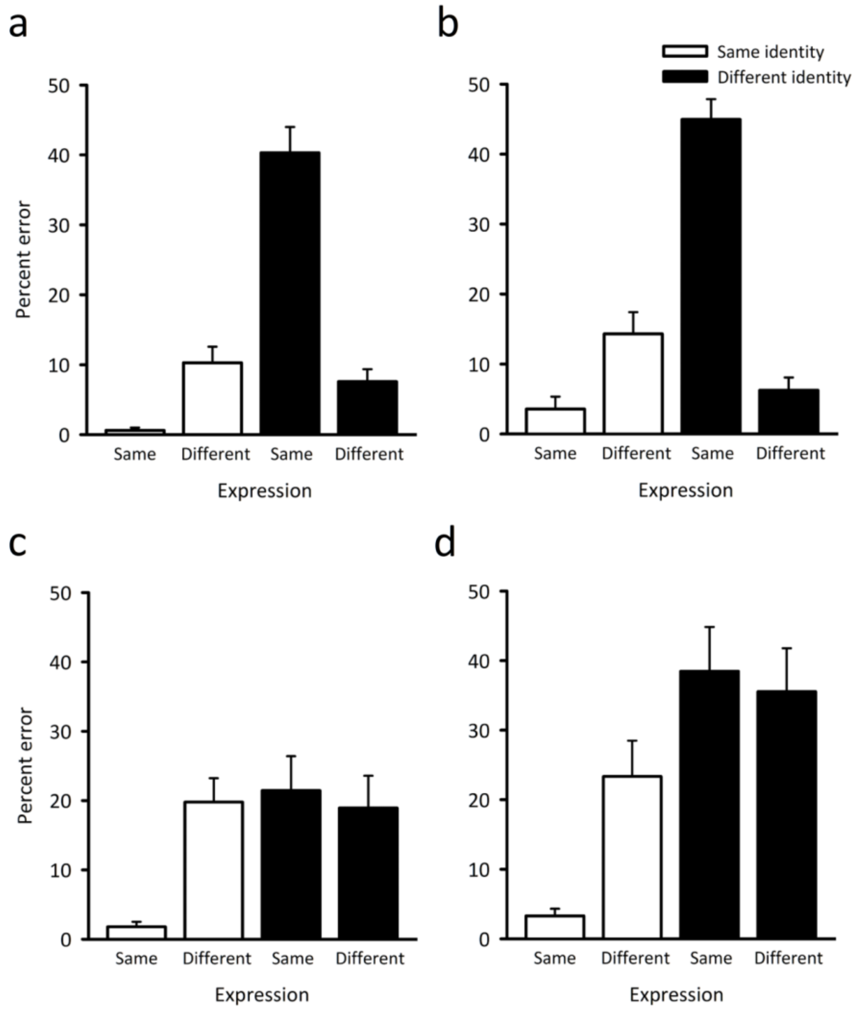

Figure A.2 *Experiment 4, Study 1 results. Percent error averaged across all participants for all four conditions. a. Judgements of expression, positive contrast; b. Judgements of expression, negative contrast; c. Judgements of identity, positive contrast; d. Judgements of identity, negative contrast. Error bars represent standard error*

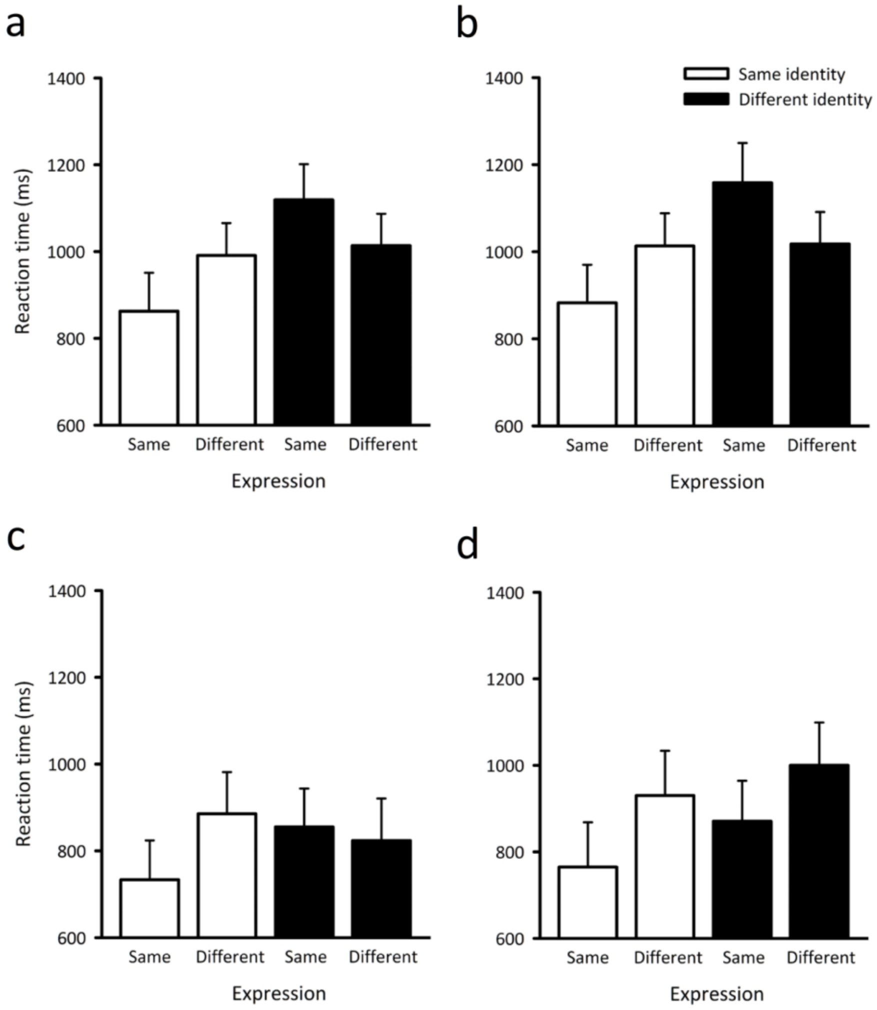

Figure A.3 *Experiment 4, Study 1 results. Reaction time (ms) for the correct responses, averaged across participants for the four conditions. a. Judgements of expression, positive contrast; b. Judgements of expression, negative contrast; c. Judgements of identity, positive contrast; d. Judgements of identity, negative contrast. Error bars represent standard error*

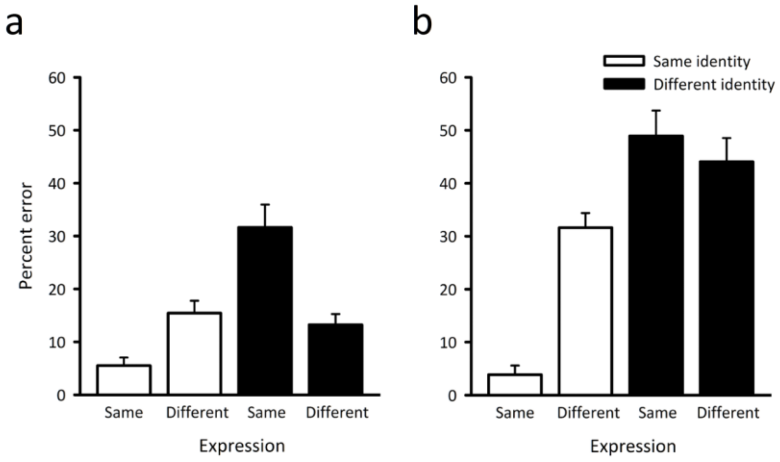

Figure A.4 *Experiment 4, Study 2 results*. *Percent error averaged across all participants for all four conditions. a. Judgments of expression. b. Judgements of identity. Error bars represent standard error.*

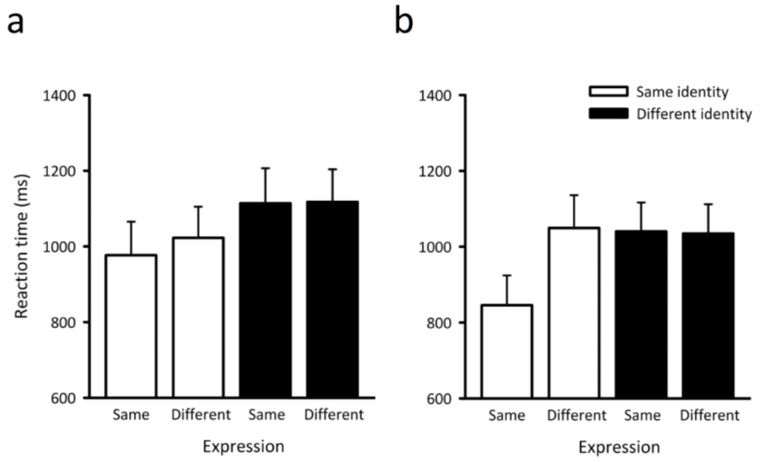

Figure A.5 *Experiment 4, Study 2 results. Reaction time (ms) for the correct responses, averaged across all participants for all four conditions. a. Judgements of expression; b. Judgements of identity, Error bars represent standard error.*

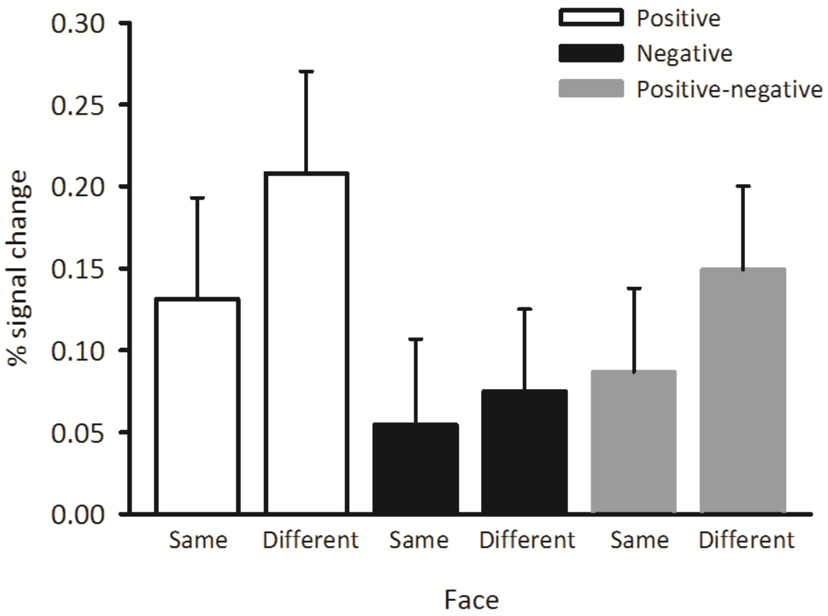

Figure A.6 *Experiment 5 results. Peak responses to all six conditions in the amygdala. Error bars represent standard error of the mean.*

7.2 Supplementary Tables

7.2.1 Chapter 2

Table B.1 *Action units (muscle groups) used by the five actors selected as stimuli when posing each of the five expressions as reported in the FEEST set. The numbers refers to the muscles used to pose the expression. Letters preceding numbers refer to the part of the muscle used. Letters proceeding numbers refer to the intensity of the expression. Actors are denoted by the codes assigned in the FEEST set*

Actor	Anger	Disgust	Fear	Happiness	Sadness
F5	4C, 5C, T23B	7D, 9C, 10A, 17B	1C, 2C, 5E, R20B, 26	6C, 12C 26	1B, 4A
F6	4D, 5B, 23C, 25, 38A	4D, 7B, 9D, 10A, 25	L1A, R1A, L2A, R2B, 4A, 5A, 26	1A, 2A, 6C, 12D, 26	1B, 17B, 24A
F8	4C, 7C, 16, 25, 29	4C, 7C, 9D	1B, 2B, 5C, 25	6C, 12C 25	1C, 4C
M1	4D, 5B, 10B, 25	4A, 7B, 10C, 17D	4B, 5C, 11A, 25	6D, 12D, 25	1B, 4C, 17A, 25
M6	4E, 5A, 16, 26	4D, 7B, 9C, 25	1B, 2B, 4B, 5B, 20A, 25	6E, 12D, 25	1A, 4C, 6A, 25, 64A

Table B.2 *Results from a 6 AFC recognition experiment reported in the FEEST set (2002). % recognition rate for the five actors and five expressions selected as stimuli in this thesis. Actor codes refer to the codes assigned in the FEEST set.*

Actor	Anger	Disgust	Fear	Happiness	Sadness
F5	100	100	88	100	88
F6	100	83	84	92	94
F8	100	94	79	100	92
M1	83	97	92	100	97
M6	100	97	88	100	79

Table B.3 *Recognition accuracy when stimuli were presented for 500 ms. The numbers reported refer to % correct of each actor and expression (see Chapter 2.1.3).*

Actor	Anger	Disgust	Fear	Happiness	Sadness	Mean
F5	90	90	90	100	100	94
F6	95	50	100	100	95	88
F8	80	80	95	100	95	90
M1	80	100	85	95	95	92
M6	95	80	85	100	55	83
Mean	88	80	91	99	88	90

Table B.4 *Recognition accuracy when stimuli were presented for 1000 ms. The numbers reported refer to % correct of each actor and expression (see Chapter 2.1.3).*

Actor	Anger	Disgust	Fear	Happiness	Sadness	Mean
F5	90	95	100	100	100	97
F6	95	65	95	100	100	91
F8	100	95	90	100	95	97
M1	70	100	80	100	100	91
M6	100	90	85	100	80	93
Mean	91	89	92	100	95	93.8

Table B.5 *Recognition accuracy when stimuli were presented for 1500 ms. The numbers reported refer to % correct of each actor and expression (see Chapter 2.1.3).*

Actor	Anger	Disgust	Fear	Happiness	Sadness	Mean
F5	100	95	100	100	95	98
F6	100	75	90	100	95	92
F8	90	100	90	100	95	95
M1	80	100	80	100	85	89
M6	95	75	95	100	85	90
Mean	93	89	91	100	91	92.8

Table B.6 *Results from the behavioural experiment to compare the recognition rate of static and dynamic stimuli (Chapter 2.1.6). Recognition rate (%) for the movie stimuli.*

Actor	Anger	Disgust	Fear	Happiness	Sadness	Mean
F5	98	88	95	95	93	94
F6	98	45	90	98	98	86
F8	98	85	73	100	83	88
M1	80	98	73	100	68	84
M6	88	45	85	100	40	72
Mean	92	72	83	98.5	76	84.3

Table B.7 *Results from the behavioural experiment to compare the recognition rate of static and dynamic stimuli (Chapter 2.1.6). Recognition rate (%) for the static stimuli.*

Actor	Anger	Disgust	Fear	Happiness	Sadness	Mean
F5	70	97.5	87.5	100	95	90
F6	97.5	62.5	90	100	92.5	88.5
F8	92.5	87.5	77.5	97.5	85	88
M1	72.5	95	85	95	57.5	81
M6	77.5	70	85	92.5	27.5	70.5
Mean	82	82.5	85	97	71.5	83.6

7.2.2 *Chapter 3*

Table B.8 *Results for the expression-categorisation experiment (Chapter 3.3.2.1). % recognition rate for each expression intensity for each expression continuum tested.*

Continua		1 %	33 %	66%	99 %	Continuum mean
Fear-happy						
	F5	97	98	98	100	98
	F8	90	83	85	100	90
Disgust-happy						
	F8	98	90	84	90	91
	M1	99	96	82	95	93
Disgust-sad						
	M1	93	88	86	95	91
	M6	89	83	95	80	87
Disgust-fear						
	F8	89	85	89	90	88
	M6	88	93	82	98	90
Happy-anger						
	F6	95	84	79	98	89
	M6	95	86	95	98	94
Sad-happy						
	F8	100	90	75	93	90
	M1	93	84	87	92	89
Fear-anger						
	F5	90	65	98	91	86
	F6	98	73	85	93	87

References

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*, 669–672.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. R. (1995). Fear and the human amygdala. *The Journal of Neuroscience*, *15*, 5879–5891.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–473.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., et al. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*, 1111–1117.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68–72.

Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex*, *4*, 544–554.

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

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*, 403–410.

Anderson, A. K., & Phelps, E. A. (2000). Expression without recognition: contributions of the human amygdala to emotional communication. *Psychological Science*, *11*, 106–111.

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin’s vase–face illusion. *NeuroImage*, *17*(2), 890–901.

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

Andrews, T. J., & Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *NeuroImage*, *21*, 91–98.

Asghar, A. U. R., Chiu, Y.C., Hallam, G., Liu, S., Mole, H., Wright, H., & Young, A. W. (2008). An amygdala response to fearful faces with covered eyes. *Neuropsychologia*, *46*, 2364–2370.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., et al. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*, 724–732.

Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science*, *21*, 595–599.

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

Baudouin, J., & Humphreys, G. (2006). Compensatory strategies in processing facial emotions: Evidence from prosopagnosia. *Neuropsychologia*, *44*, 1361–1369.

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*, 91–102.

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

Benton, C. P. (2009). Effect of photographic negation on face expression aftereffects. *Perception*, *38*, 1267–1274.

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*, 56–71.

Bimler, D., & Kirkland, J. (2001). Categorical perception of facial expressions of emotion: Evidence from multidimensional scaling. *Cognition & Emotion*, *15*, 633–658.

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*, 537-41

Blair, R. J. R., Morris, J. S., Frith, C. D., Perrett, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, *122*, 883–893.

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

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, *17*, 875–887.

Broks, P., Young, A. W., Maratos, E. J., Coffey, P. J., Calder, A. J., Isaac, C. L., Mayes, A. R., et al. (1998). Face processing impairments after encephalitis: amygdala damage and recognition of fear. *Neuropsychologia*, *36*, 59–70.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*, 369–384.

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception*, *15*, 387–397.

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

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

Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A., & Linney, A. (1991). Recognising facial surfaces. *Perception*, *20*, 755 – 769.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., et al. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception*, *22*, 131 – 152.

Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. *Visual Cognition*, *1*, 141–180.

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

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

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

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., & Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain and Cognition, 2*, 257–284.

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. *Proceedings: Biological Sciences*, *259*, 137–143.

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

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

Calder, A .J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, *3,* 81–118.

Calder A. J., Young A. W., Rowland D., Perrett D. I., Hodges J. R., & Etcoff N. L. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, *13,* 699–745.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 527–551.

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

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

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., & Henson, R. N. A. (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Current Biology*, *17,* 20–25.

Calder, A. J. (2011). Does facial identity and facial expression involve seperate visual routes? In A. J. Calder, G, Rhodes, M. H. Johnston, & J. V. Haxby (Eds.) *The Oxford handbook of face perception* (pp. 427–448). New York: Oxford University Press Inc.

Campbell, R., Brooks, B., de Haan, E., Roberts, T. (1996). Dissociating face processing skills: Decision about lip-read speech, expression, and identity. *The Quarterly Journal of Experimental Psychology Section A*, *49*, 295–314.

Campbell, J., & Burke, D. (2009). Evidence that identity-dependent and identity-independent neural populations are recruited in the perception of five basic emotional facial expressions. *Vision Research*, *49*, 1532–1540.

Chan, A. W., & Downing, P. E. (2011). Faces and eyes in human lateral prefrontal cortex. *Frontiers in Human Neuroscience*, *5*. doi:10.3389/fnhum.2011.00051

Cohen Kadosh, K., Henson, R. N. A., Cohen Kadosh, R., Johnson, M. H., & Dick, F. (2010). Task-dependent activation of face-sensitive cortex: an fMRI adaptation study. *Journal of Cognitive Neuroscience*, *22*, 903–917.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, *29*, 893 – 909.

Cook, S. W. (1939). The judgment of intelligence from photographs. *The Journal of Abnormal and Social Psychology*, *34*, 384–389.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*, 1158–1173.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*, 28–30.

Darwin, C. (1872). *The expressions of the emotions in man and animals*. Great Britain: John Murray.

Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, *63*(2), 180–187.

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*, 1627–1635.

Davies-Thompson, J., & Andrews, T. J. (2012). Intra- and inter-hemispheric connectivity between face-selective regions in the human brain. *Journal of Neurophysiology*.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968–980.

Edwards, K. (1998). The Face of Time: Temporal Cues in Facial Expressions of Emotion. *Psychological Science*, *9*, 270–276.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. *Nebraska symposium on motivation* (pp. 207–283). Lincoln: University of Nebraska press.

Ekman, P. (1973). *Darwin and facial expressions*. New York: Academic Press.

Ekman, P. (1975). *Unmasking the face*. New Jersey: Prentice-Hall Inc.

Ekman, P. (1980). *The face of man*. USA: Garland Publishing Inc.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell’s mistaken critique. *Psychological Bulletin*, *115*, 268–287.

Ekman, P. (1999). Facial expressions. *The handbook of cognition and emotion*. Sussex, UK: John Wiley & Sons.

Ekman, P., & Friesen, V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, Californa: Consulting Psychologists Press.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–235.

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*, 131–146.

Ellamil, M., Susskind, J. M., & Anderson, A. K. (2008). Examinations of identity invariance in facial expression adaptation. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(3), 273–281.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*, 431–439.

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

Etcoff, N. L. (1984). Selective attention to facial identity and facial emotion. *Neuropsychologia*, *22*, 281–295.

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

Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in Brain Research*, *156*, 345–361.

Ethofer, T., Gschwind, M., & Vuilleumier, P. (2011). Processing social aspects of human gaze: A combined fMRI-DTI study. *NeuroImage*, *55*, 411–419.

Fairhall, S. L., & Ishai, A. (2007). Effective Connectivity within the distributed cortical network for face perception. *Cerebral Cortex*, *17,* 2400–2406.

Feinstein, J. S., Adolphs, R., Damasio, A., & Tranel, D. (2011). The Human amygdala and the induction and experience of fear. *Current Biology*, *21*, 34–38.

Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research, 1127*, 80–89.

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

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

Gallagher, H. L., & Frith, C. D. (2004). Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia*, *42*, 1725–1736.

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

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the “fusiform face area” in processing facial expression. *Neuropsychologia*, *43*, 1645–1654.

George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nature Neuroscience, 2,* 574–580.

Gilad, S., Meng, M., & Sinha, P. (2009). Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences, 106,* 5353–5358.

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research, 16,* 416–424.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34,* 77–86.

Graham, R., Devinsky, O., & Labar, K. S. (2007). Quantifying deficits in the perception of fear and anger in morphed facial expressions after bilateral amygdala damage. *Neuropsychologia, 45,* 42–54.

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

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

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35,* 1167–1175.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex, 22,* 1564–1576.

Hagan, C. C., Woods, W., Johnson, S., Calder, A. J., Green, G. G. R., & Young, A. W. (2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proceedings of the National Academy of Sciences, 106,* 20010–20015.

Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cerebral Cortex, 10,* 69–81.

Harris, A., & Nakayama, K. (2008). Rapid adaptation of the M170 response: Importance of face parts. *Cerebral Cortex, 18,* 467–476.

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*, 203–218.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*, 1634–1640.

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

Hsu, S. M. & Young, A. (2004). Adaptation effects in facial expression recognition. *Visual Cognition*, *11*, 871–899.

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

Hooker, C. I., Paller, K. A., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2003). Brain networks for analyzing eye gaze. *Cognitive Brain Research*, *17*, 406–418.

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*, 173–181.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, *19*, 1543–1548.

Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, *3*, 1–38.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, *30*, 875–887.

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*, 4302–4311.

Kanwisher, N., & Barton, J. J. S. (2011). The functional architecture of the face system: integrating evidence from fMRI and patient studies. In A. J. Calder, G, Rhodes, M. H. Johnston, & J. V. Haxby (Eds.) *The Oxford handbook of face perception* (pp. 111-130). New York: Oxford University Press Inc.

Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Ito, K., Fukuda, H., et al. (1999). The human amygdala plays an important role in gaze monitoring: A PET study. *Brain*, *122*, 779–783.

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*, 37–52.

Kesler-West, M. L., Andersen, A. H., Smith, C. D., Avison, M. J., Davis, C. E., Kryscio, R. J., & Blonder, L. X. (2001). Neural substrates of facial emotion processing using fMRI. *Cognitive Brain Research*, *11*(2), 213–226.

Kirita, T., & Endo, M. (1995). Happy face advantage in recognizing facial expressions. *Acta Psychologica*, *89*, 149–163.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *NeuroImage*, *37*, 1445–1456.

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex, 13*, 1023–1033.

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

Lee, L. C., Andrews, T. J., Johnson, S. J., Woods, W., Gouws, A., Green, G. G. R., & Young, A. W. (2010). Neural responses to rigidly moving faces displaying shifts in social attention investigated with fMRI and MEG. *Neuropsychologia*, *48*, 477–490.

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

Liu, C. H., & Chaudhuri, A. (1998). Are there qualitative differences between face processing in photographic positive and negative? *Perception*, *27*, 1107–1122.

Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. *NeuroReport*, *11*, 337–341.

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

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

Margulies, D. S., Böttger, J., Long, X., Lv, Y., Kelly, C., Schäfer, A., Goldhahn, D., et al. (2010). Resting developments: a review of fMRI post-processing methodologies for spontaneous brain activity. *Magma, 23,* 289–307.

Martens, U., Leuthold, H., & Schweinberger, S. R. (2010). Parallel processing in face perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 103–121.

Martinez, A. M., & Du, S. (2012). A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *Journal of Machine Learning Research, 13*, 1589–1608.

Mckelvie, S. (1973). Meaningfulness and meaning of schematic faces. *Perception, & Psychophysics, 14*, 343–348.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*, 812–815.

Mosconi, M. W., Mack, P. B., McCarthy, G., & Pelphrey, K. A. (2005). Taking an “intentional stance” on eye-gaze shifts: a functional neuroimaging study of social perception in children. *NeuroImage, 27,* 247–252.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Cognitive Brain Research, 12,* 225–231.

Nasr, S., & Tootell, R.B.H. (2012). Role of fusiform and anterior temporal cortical areas in facial recognition. *NeuroImage, 63,* 1743-1753.

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research, 12,* 419–446.

Norman-Haignere, S. V., McCarthy, G., Chun, M. M., & Turk-Browne, N. B. (2012). Category-selective background connectivity in ventral visual cortex. *Cerebral Cortex, 22,* 391–402.

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, *13*,135–143.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience, 12*(6), 1013–1023.

Ohayon, S., Freiwald., W, A., Tsao., D. Y. (2012). What makes a cell face selective? The importance of contrast. *Neuron, 74,* 567-581.

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

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, *30*, 10127–10134.

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia*, *41*, 156–170.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, *16*, 1706–1716.

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

Pelphrey, K. A., Morris, J. P., Michelich, C.R., Allison, T., McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: an fmri study of eye, mouth and hand movements. *Cerebral Cortex, 15,* 1866–1876.

Perrett, D.I., Rolls, E. T., & Caan, W. (1979). Temporal lobe cells of the monkey with visual responses selective for faces. *Neuroscience Letters*, *S3,* S358.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*(3), 329–342.

Perrett, D.I., & Rolls, E. (1983). Neural mechanisms underlying the visual analysis of faces. In J. P. Ewert, R. R. Capranica & D. J. Ingle (Eds.) *Advances in Vertebrate Neuroethology* (pp. 543–566). New York: Plenum Press.

Perrett, D. I., Xiao, D., Barraclough, N. E., Keysers, C., & Oram, M. W. (2009). Seeing the future: Natural image sequences produce “anticipatory” neuronal activity and bias perceptual report. *Quarterly Journal of Experimental Psychology*, *62*, 2081–2104.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, *16*, 331–348.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., Williams, S. C. R., et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society B: Biological Sciences*, *265*, 1809–1817.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*, 2356–2363.

Pitcher, D., Walsh, V., Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research, 209*, 481-493.

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*, 2188–2199.

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

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

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979.

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*, 107–113.

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology, 66*, 870–881.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin, 115*, 102–141.

Russell, J. A., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General, 116*, 223-237.

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception*, *32*, 1093–1107.

Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, *38*, 1211–1219.

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

Said, C. P., Baron, S. G., & Todorov, A. (2008). Nonlinear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience*, *21*, 519–528.

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

Said, C. P., Moore, C. D., Norman, K. A., Haxby, J. V., & Todorov, A. (2010). Graded representations of emotional expressions in the left superior temporal sulcus. *Frontiers in Systems Neuroscience, 4*, 6 doi: 10.3389/fnsys.2010.00006.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences*, *14*, 303–316.

Sanghera, M. K., Rolls, E. T., & Roper-Hall, A. (1979). Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Experimental Neurology, 63*, 610–626.

Santos, I. M., & Young, A. W. (2008). Effects of inversion and negation on social inferences from faces. *Perception*, *37*, 1061–1078.

Sato, W., Kubota, Y., Okada, T., Murai, T., Yoshikawa, S., & Sengoku, A. (2002). Seeing happy emotion in fearful and angry faces: qualitative analysis of facial expression recognition in a bilateral amygdala-damaged patient. *Cortex, 38*, 727–742.

Sato, W., Kochiyama, T., & Yoshikawa, S. (2010). Amygdala activity in response to forward versus backward dynamic facial expressions. *Brain Research, 1315,* 92–99.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30,* 1088–1096.

Schienle, A., Stark, R., Walter, B., Blecker, C., Ott, U., Kirsch, P., Sammer, G., et al. (2002). The insula is not specifically involved in disgust processing: an fMRI study. *Neuroreport, 13,* 2023–2026.

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology. Human Perception and Performance*, *24*, 1748–1765.

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: experiments with morphed faces. *Perception & Psychophysics*, *61,* 1102–1115.

Schweinberger, S. R., Burton, A. M. (Eds.) (2011). Person perception 25 years after the Bruce & Young model (1986) [Special issue]. *British Journal of Psychology, 102,* 695-974.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69,* 243–265.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature*, *385*, 254–257.

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *32*, 811–830.

Shuttleworth Jr., E. C., Syring, V., & Allen, N. (1982). Further observations on the nature of prosopagnosia. *Brain and Cognition*, *1,* 307–322.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, *41*, 653–662.

Sprengelmeyer, R., Young, A. W., Schroeder, U., Grossenbacher, P. G., Federlein, J., Büttner, T., & Przuntek, H. (1999). Knowing no fear. *Proceedings of the Royal Society B: Biological Sciences*, *266*, 2451–2456.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, *11*, 843–850.

Thielscher, A., & Pessoa, L. (2007). Neural correlates of perceptual choice and decision making during fear-disgust discrimination. *Journal of Neuroscience*, *27*, 2908–2917.

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature Neuroscience*, *10*, 505–511.

Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Computer graphics in facial perception research. *IEEE Computer Graphics and Applications*, *21*, 42-50.

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cognitive and Affective Neuroscience*, *3*, 303–312.

Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology*, *38*, 690–690.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670–674.

Turk-Browne, N. B., Norman-Haignere, S. V., & McCarthy, G. (2010). Face-specific resting functional connectivity between the fusiform gyrus and posterior superior temporal sulcus. *Frontiers in Human Neuroscience*, *4*, 176.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*, *30*, 829–841.

Webster, M. A., Kaping, D., Mizokami, Y., Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*, 557-561.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, *78*, 105–119.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science, 7*, 177–188.

Whalen, P. J., Kagan, J., Cook, R.G., Davis, F. C., Kim, H., Polis, S., et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science, 306,* 2061–2061.

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

White, M., & Li, J. (2006). Matching faces and expressions in pixelated and blurred photos. *The American Journal of Psychology*, *119*, 21–28.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*, 277–283.

Winston, J. S., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, *20*, 84–97.

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*, 1830–1839.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental Psychology: Revised edition*. New York: Henry Holt.

Wyk, B. C. V., Hudac, C. M., Carter, E. J., Sobel, D. M., & Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychological Science*, *20*, 771–777.

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*, 2800–2807.

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747–759.

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

Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P., & Hanley, J. R. (1995). Face processing impairments after amygdalotomy. *Brain*, *118*, 15–24.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*, *63*, 271–313.

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion - stimuli and tests (FEEST)*. Bury St Edmunds, England: Thames Valley Test Company.

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

Yue, X., Nasr, S., Devaney, K. J., Holt, D. J., Tootell, R. B. H. (2013). FMRI analysis of contrast polarity processing in face-selective cortex in human and monkeys. *NeuroImage* http://dx.doi.org/10.1016/j.neuroimage.2013.02.068

Zhang, H., Tian, J., Liu, J., Li, J., & Lee, K. (2009). Intrinsically organized network for face perception during the resting state. *Neuroscience letters*, *454*, 1–5.