University of Stirling

Psychology
School of Natural Sciences

Alexandra Helen McIntyre

Applying psychology to forensic facial identification: perception and identification of facial composite images and facial image comparison

Thesis submitted for the degree of Doctor of Philosophy
March 2012

Abstract

Eyewitness recognition is acknowledged to be prone to error but there is less understanding of difficulty in discriminating unfamiliar faces. This thesis examined the effects of face perception on identification of facial composites, and on unfamiliar face image comparison. Facial composites depict face memories by reconstructing features and configurations to form a likeness. They are generally reconstructed from an unfamiliar face memory, and will be unavoidably flawed. Identification will require perception of any accurate features, by someone who is familiar with the suspect and performance is typically poor. In typical face perception, face images are processed efficiently as complete units of information. Chapter 2 explored the possibility that holistic processing of inaccurate composite configurations will impair identification of individual features. Composites were split below the eyes and misaligned to impair holistic analysis (cf. Young, Hellawell, & Jay, 1987); identification was significantly enhanced, indicating that perceptual expertise with inaccurate configurations exerts powerful effects that can be reduced by enabling featural analysis.

Facial composite recognition is difficult, which means that perception and judgement will be influence by an affective recognition bias: smiles enhance perceived familiarity, while negative expressions produce the opposite effect. In applied use, facial composites are generally produced from unpleasant memories and will convey negative expression; affective bias will, therefore, be important for facial composite recognition. Chapter 3 explored the effect of positive expression on composite identification: composite expressions were enhanced, and positive affect significantly increased identification. Affective quality rather than expression strength mediated the effect, with subtle manipulations being very effective.

Facial image comparison (FIC) involves discrimination of two or more face images. Accuracy in unfamiliar face matching is typically in the region of 70%, and as discrimination is difficult, may be influenced by affective bias. Chapter 4 explored the smiling face effect in unfamiliar face matching. When multiple items were compared, positive affect did not enhance performance and false positive identification increased. With a delayed matching procedure, identification was not enhanced but in contrast to face recognition and simultaneous matching, positive affect improved rejection of foil images. Distinctive faces are easier to discriminate. Chapter 5 evaluated a systematic caricature transformation as a means to increase distinctiveness and enhance discrimination of unfamiliar faces. Identification of matching face images did not improve, but successful rejection of non-matching items was significantly enhanced.

Chapter 6 used face matching to explore the basis of own race bias in face perception. Other race faces were manipulated to show own race facial variation, and own race faces to show African American facial variation. When multiple face images were matched simultaneously, the transformation impaired performance for all of the images; but when images were individually matched, the transformation improved perception of other race faces and discrimination of own race faces declined. Transformation of Japanese faces to show own race dimensions produced the same pattern of effects but failed to reach significance. The results provide support for both perceptual expertise and featural processing theories of own race bias. Results are interpreted with reference to face perception theories; implications for application and future study are discussed.

Acknowledgements

I am grateful to Peter Hancock who encouraged me to complete a thesis and who provided the support to make it possible. Without his guidance and collaboration the content would have been very much poorer.

Charlie Frowd instigated my love for cognitive research and taught me how to be a researcher. I thank him and hold him responsible in equal measure.

Steve Langton brought clarity, and influenced my treatment and understanding of statistical design. I thank him for always being interested and encouraging me.

Hayley Ness and Catriona Havard taught me more than they could imagine, I thank them for the friendship and for being seriously good role models.

Friends and colleagues have inspired and encouraged me throughout: Kay Faichney, Claire Wilson, Catherine Foster, Gary Dalton, Catherine MacLeod, Kimberley Collins, Karri Gillespie-Smith, Debbie Riby, Robbie Cooper, Anna Law, Jack Griffey, Janelle Renfrew, Nadia Ayal, Vicky Mileva, Mary Cowan, Charlotte Bjuren….

….and some very impressive people took the time to provide encouragement and advice; Vicki Bruce, Graham Davies, Graham Pike, Amina Memon, Tim Valentine, Neil Brewer, Stephen Penrod, and Rod and Marilyn Lindsay. I owe many thanks.

A thesis takes a long time to produce and many people have to work hard to make it possible. This thesis is for my invisible collaborators; my husband Des, Céire and Caitlin, and my Mum and Dad.

Author’s declaration

I declare that the work carried out for this thesis was completed in accordance with the regulations of Stirling University. The research presented in chapters 2, 3, and 4 is my own. The work presented in chapters 5 and 6 was conducted in collaboration with Professor Peter J.B. Hancock. The caricature and race transformations were devised by Peter Hancock. With the exception of pilot work detailed in Chapter 2, all studies are original and have not been submitted for another degree. Any views expressed in this work are those of the author and in no way represent those of Stirling University.

Signed: ………………………………………………

Date: …………………………………………………

Publications

McIntyre, A. H., Hancock, P. J. B., Frowd, C. D., and Langton, S. R. H. Spitting facial composites can improve identification. Experiments 1 and 2 under revision for *Journal of Experimental Psychology: Applied.*

McIntyre, A.H., Hancock, P.J.B., Frowd, C.D. & Langton, S.R.H. The smiling face bias can improve facial composite identification. The work described in Experiments 7, 8, and 9 is under revision for *Journal of experimental Psychology: Applied*. The revised manuscript will include Experiment 10.

McIntyre, A.H., Hancock, P.J.B., & Kittler, J. Improving face matching with caricature. The work described in Experiments 15, 16, and 17, is under revisions for *Applied Cognitive Psychology*.

McIntyre, A.H., Hancock, P.J.B., & Kittler, J. Reducing the own race bias in perceptual face matching. The work described in Experiments 19, 20, and 21 is under revision for *Journal of Experimental Social Psychology.*

Contents

Abstract	i
Acknowledgements	iii
Author's declaration	iv
Chapter 1:	
General Introduction	1
Face perception	4
Familiar and unfamiliar face perception	4
Developing face processing expertise	5
The face inversion effect	6
The composite face effect	7
Face memory	8
Bruce and Young parallel processing model (1986)	8
Valentine's Multi-dimensional face space model (1991)	12
Perception and bias in face memory	14
The own race bias in face perception	14
The smiling face bias in face perception	16
Forensic identification	18
Face recognition and the law	18
Face perception and the law	20
Facial composite recognition	22
Facial composite systems and methodology	22
Experimental methodology	23
Unfamiliar witnesses	25
Featural and holistic processing	26
Familiar recognition and inaccuracy in composite images	27
Recognition bias and facial composites	28
Facial image comparison	29
FIC methodology	30
FIC performance	33
Image quality	34
Viewpoint and lighting	35
Matching people and face images	36
Individual differences and expertise	37
Distinctiveness and discrimination	38
FIC and own race bias	38
FIC and facial expression	40
Thesis structure	41

Contents

Chapter 2:

Holistic face processing and facial composite recognition 43

Featural and configural information 44
Holistic and configural face perception 45
Aim of study 46
Facial composite and photographic stimuli 48
Experiment 1 Configurally enhanced facial composites 49
Experiment 2 Configurally disrupted (split) facial composites 54
Experiment 3: Within participant replication of experiments 61
Experiment 4: Reaction times for complete and split images 66
Experiment 5: The importance of upper features 75
Experiment 6: Presentation of non-holistic composite images 79
Summary and conclusions 82

Chapter 3:

Emotional bias in facial composite construction and recognition 90

Unfamiliar witnesses and contextual memory 91
The smiling face bias and facial composite recognition 92
Aim of study 93
Enhanced expressions in facial composites 94
Experiment 7: Enhancing positive affect in famous facial composites 96
Experiment 8: Enhancing positive affect in personally familiar facial composites 99
Experiment 9: Perception of affective transforms 108
Experiment 10: Identification with different levels of transformation 111
Summary and Conclusions 118

Chapter 4:

Exploring familiarity bias in face matching 125

Discrimination of unfamiliar faces 126
The smiling face bias and facial discrimination 127
Aim of study 128
Experiment 11: Enhancing attention with positive affect 129
Experiment 12: Applying positive affect to multiple item Arrays 134
Experiment 13: Positive affect within sequential face matching arrays 139
Experiment 14: Positive affect in a delayed matching task 143
Summary and conclusions 149

Contents

Chapter 5:

Increasing distinctiveness in unfamiliar face discrimination 155

Distinctiveness and caricature 156
Caricature and multi-dimensional face space 157
Aim of study 159
Creating distinctiveness with caricature 160
Experiment 15: Face matching with 30% caricatures 160
Experiment 16: Face matching with 50% caricatures 167
Experiment 17: Evaluation of caricature levels 171
Experiment 18: Sequential face matching with caricature 179
Summary and conclusions 183

Chapter 6:

Expert face processing and the own race bias 190

Perception of other race faces 191
Aim study 195
Defining and transforming facial characteristics 196
Matching own race Caucasian and African American faces
Experiment 19: Matching original and race transformed faces in simultaneous multiple item arrays 199
Experiment 20: Matching original and race transformed faces in sequential arrays 205
Experiment 21: Matching original and race transformed faces in a combined procedure 210
General Discussion 215
Evaluating ORB in Caucasians with Japanese faces 219
Experiment 22: Matching original and race transformed Japanese faces in simultaneous arrays 218
Experiment 23: Matching original and race transformed Japanese faces in sequential arrays 223
Experiment 24: Matching original and race transformed Japanese faces in combined arrays 227
Summary and Conclusions 230

Contents

Chapter 7:

General Discussion 239

Facial composite perception and identification 240
The split facial composite effect 240
The positive facial composite effect 242
Facial image comparison 247
The smiling face bias and facial discrimination 247
The caricature effect in facial discrimination 248
Own race bias in facial discrimination 251
Imaging methodology 254
Conclusions 256

References 260

1

General Introduction

For more than a hundred years eyewitness memory and unfamiliar face recognition have been the focus of psychological study (Behrman & Davey, 2001; Cattell, 1893; Chance & Goldstein, 1979; Davies & Griffiths, 2008; Davies, Shepherd, & Ellis, 1978; Deffenbacher, Bornstein, McGorty, & Penrod, 2008; Lindsay & Wells, 1985; MacLin, MacLin, & Malpass, 2001); these endeavours have created awareness within the legal system that unfamiliar face recognition is prone to error, and have generated findings that have contributed to the development of formal codes of practice and identification guidelines. In contrast, the apparently simple ability to perceive faces has been overlooked and there has been little formal evaluation of how unfamiliar faces are interpreted or how well they can be discriminated within forensic applications (Bruce, Henderson, Greenwood, Hancock, Burton, & Miller, 1999; Henderson, Bruce, & Burton, 2001; Kemp, Towell, & Pike, 1997; Strathie, 2010). As a consequence, in policing and within the legal system, there is ignorance that even in the most favourable conditions, unfamiliar face discrimination produces high rates of error (Hancock & McIntyre, 2011). This thesis examined the effects of face perception on identification of facial composite images, and on discrimination of unfamiliar faces in facial image comparison procedures.

Facial composites are intended to portray unknown individuals who have been involved in a serious crime. They are produced from memory by unfamiliar witnesses, and identification by people who are familiar with the suspect can be difficult to achieve (Brace, Pike, Kemp, Turner, & Bennett, 2006; Davies, van der Willik, & Morrison, 2000; Frowd, Carson, Ness, Quiston-Surrett, et al., 2005). Since the nineteen-eighties the systems that generate the composites have advanced considerably (Brace, Pike, & Turner, 2008; Davies et al., 2000; Frowd, Carson, Ness, Richardson, et al., 2005; Frowd, Hancock, & Carson, 2004) and interviewing protocols have been developed to facilitate better recall of facial information from memory (e.g. Davies & Milne, 1985; Frowd, Bruce, Smith, & Hancock, 2008; Geiselman et al., 1984), yet identification of composite images remains far from optimal.

The psychology of face perception provides insight as to why facial composites may be difficult for unfamiliar witnesses to produce, for example Brace, Pike, Allen, and Kemp (2006) found that difficulties in remembering and communicating facial information impaired the quality of the composite image, and Frowd, Bruce, McIntyre, and Hancock (2007) showed that external features, such as hairstyles, could distract witnesses from creating a recognisable facial arrangement. What is more, composite procedures that require witnesses to remember individual features are incompatible with natural holistic face processing (Wells & Hasel, 2007). Psychology has also been employed to show that displaying multiple composite images (Brace et al., 2006) and combining them (Bruce, Ness, Hancock, Newman, & Rarity, 2002; Valentine, Davis, Thorner, Solomon, & Gibson, 2010) can enhance familiar face recognition, while caricaturing composite images can improve identification of accurate composite information (Frowd, Bruce, Ross, McIntyre, & Hancock, 2007). The research presented in this thesis also employs the psychology of face perception but it will be the first to draw on how the effects of unfamiliar face perception, which is employed in the construction of facial composite images, can be detrimental to the familiar face perception that is required to achieve successful identification.

Facial image comparison is perhaps the most common facial identification procedure and is 'quite simply', deciding whether two face images portray the same person. However, this straightforward image matching practice consistently produces error rates in region of 30% (e.g. Bruce et al., 1999; Megreya & Burton, 2008). Psychological evaluations have determined that image properties such as viewpoint and lighting (Adini, Moses, & Ullman, 1994; Hill & Bruce, 1996) or changes of expression (Bruce et al., 1999), significantly impair face matching performance, and it has been shown that perceptual discrimination is poorer when matching faces of another race (Megreya & Burton, 2006) or for face images that are more typical (Valentine & Bruce, 1986a). This thesis will also explore the effects of expression, distinctiveness, and race on unfamiliar face matching, but will be the first to show how psychological models of face memory can inspire image transformations that might enhance discrimination within forensic applications.

The introduction which follows will discuss perception and memory of familiar and unfamiliar faces, and will examine two models of how faces may be stored in memory; the own race bias and smiling face bias will describe how the nature of face memory may influence perceptual and identification judgements. Forensic facial identification will then be discussed with particular reference to facial composites and face image comparison and finally an outline of the thesis will be described.

Face Perception

Familiar and unfamiliar face perception

Faces are arguably the most important visual stimuli that we employ to successfully navigate our world, and our capacity to learn and discriminate new ones has no known boundaries or limits. When faces are familiar to us we can discriminate among hundreds of them faster than we can consciously form thoughts about them, or recall their names (Johnston & Bruce, 1990; Young, Ellis, & Flude, 1988; Young, McWeeny, Ellis, & Hay, 1986; Young, McWeeny, Hay, & Ellis, 1986); and yet when faces are unfamiliar, our ability to remember them, recognise them, or differentiate them, is surprisingly poor (Hancock, Bruce, & Burton, 2000).

The contrast between familiar and unfamiliar face perception ability is striking, and is reflected somewhat in the processes and type of information that is employed: if faces are familiar discrimination will be extremely good and recognition will draw predominantly on the stable internal region of the face (i.e. the configural arrangement of the eyes, nose, and mouth), but unfamiliar face recognition, which is much less effective, may be achieved equally well from the internal features as from external information such as hair, face shape, and ears (Ellis, Shepherd, & Davies, 1979; Fletcher, Butavicius, & Lee, 2008; Young, Hay, McWeeny, Flude, & Ellis, 1985); and while other research has found an external feature advantage for unfamiliar faces (Bruce et al., 1999) there is evidence of a shift from an external to internal feature preference as faces become familiar (Bonner, Burton, & Bruce, 2003).

Developing face processing expertise

It is not yet clear how faces are learned and become stored in long term memory, but it is apparent that face processing skill develops rapidly from birth and throughout childhood, reaching adult levels of performance in the second decade of life (Crookes & McKone, 2009; Itier & Taylor, 2004; Kelly et al., 2009, 2007, 2005; Mondloch, Le Grand, & Maurer, 2010; Schwarzer, Zauner, & Jovanovic, 2007). Within the development of face processing expertise is the ability to process face images holistically as complete units of information, as well as sensitivity to the configural relationships within and between the features of the face (Diamond & Carey, 1986; Itier & Taylor, 2004; Maurer, Grand, & Mondloch, 2002; Rhodes, Brake, Taylor, & Tan, 1989).

Holistic processing is regarded as a hallmark of face processing expertise; it is associated with, and is sometimes referred to interchangeably with configural processing. To clarify the terminology; configural processing concerns perception and analysis of the relative sizes and distances between the facial features (Maurer et al., 2002); while holistic processing refers to functional analysis of the complete face image, with perception of featural and configural information interpreted as a single representation. Rossion (2008) suggests that confusion can be avoided if holistic interpretation of the face stimulus is regarded as a perceptual process originating with the observer, while configural properties describe the spatial information about the face that is interpreted most effectively within the holistic process.

The face inversion effect

Inversion inhibits the ability to recognise faces (Sekuler, Gaspar, Gold, & Bennett, 2004; Valentine, 1988; Yin, 1969) and makes holistic interpretation of the complete face image difficult (Farah, Tanaka, & Drain, 1995; Riesenhuber, Jarudi, Gilad, & Sinha, 2004; Tanaka & Farah, 1993; Yovel & Kanwisher, 2004). Most studies have shown that sensitivity to the configural properties is particularly affected and while this might be the main cause of the face inversion effect (Barton, Cherkasova, & O'Connor, 2001; Diamond & Carey, 1986; Leder & Bruce, 2000; Leder, Candrian, Huber, & Bruce, 2001; Maurer et al., 2002), Rossion (2008, 2009) proposes that face inversion disrupts the holistic process and narrows the perceptual field, making it difficult to process the relationships and distances between the facial features. By this view configural analysis is impaired because of loss of holistic processing, and this is supported by findings that show weaker inversion effects for spatial relationships between features that are closer together, or that are closer to the fixation point (Goffaux & Rossion, 2007; Sekunova & Barton, 2008).

It is generally accepted that when faces are inverted holistic analysis is impaired and like object perception, recognition becomes dependent on the individual features (Yin, 1969). Megreya and Burton (2006) found a correlation between unfamiliar face matching and recognition of inverted, but not upright familiar faces, leading them to suggest that unfamiliar faces are not processed as faces, but like objects, in a feature based manner. However, neuropsychological evidence indicates that face matching and face recognition are distinct processes (Malone, Morris, Kay, & Levin, 1982) therefore Megreya and Burton's (2006) results cannot inform as to whether unfamiliar faces are processed in a face-like way when we are not attempting to discriminate between them in a matching task.

The composite face effect

Evidence that unfamiliar faces are in fact processed holistically, is found within composite face paradigms: the composite face effect was established by Young, Hellawell and Hay (1987) and demonstrated that aligning the top half of one famous person's face with the bottom half of a second famous person's face significantly impaired perception of the constituent parts. As this effect disappears when the merged images are inverted, the composite face effect has been widely acknowledged to demonstrate holistic face processing of the novel composite images. Holistic perception of unfamiliar faces is shown when the upper half of two identical unfamiliar face images are aligned with the lower halves of different unfamiliar face images. When the images are viewed together, the identical parts appear different and participants fail to identify them as being from the same person; therefore, holistic processing of these composites images creates the perception of different people and induces discrimination errors (Calder & Young, 2005; Calder, Young, Keane, & Dean, 2000; Hole, 1994; Rossion, 2009). Using a similar discrimination task, Schiltz, Dricot, Goebel, and Rossion, (2010) showed that fMRI activity in the right temporal lobe in the fusiform face area (FFA) indicated that the composite images were responded to as if they were completely novel faces.

Facial mapping is a technique employed by independent expert forensic services that are routinely employed by police forces; it is used to determine whether two face images originate from the same person, and involves vertical alignment of different left and right face halves. Strathie (2010) evaluated this technique in her thesis and also found that the unfamiliar aligned face images were processed in a holistic manner, leading to a substantial increase in false identification errors. As this type of expert testimony is endorsed by the Attorney General (Reference No. 2 of 2002), and carries particular weight in judicial proceedings, her work provides compelling evidence that formal evaluation of unfamiliar face perception and revision of the guidelines is sorely required. The collective findings suggest that processing differences between familiar and unfamiliar faces may be a function of sensitivity to featural or configural variation, rather than to expert face processing strategies per se.

Face Memory

Bruce and Young parallel processing model (1986)

While we do not know how faces are learned or are stored in long term memory, models have been devised to account for face processing effects. The Bruce and Young (1986) model *(a reproduction of the Bruce and Young model is provided in figure 1.1)* describes a sequence of four modular stages of face recognition that operate in parallel with independent structural analysis, facial speech analysis, and expression analysis modules, that is face recognition will operate independently of, but in parallel with, interpretation of facial expression etc.

(e.g. Sergent, Ohta, Macdonald, & Zuck, 1994; Young, McWeeny, Hay, & Ellis, 1986). Face perception begins with structural encoding where a pictorial and structural representation of the face is formed, this will be matched against the memory store of known faces. If there is a good match between the encoded face structure and a stored face representation, activation of a face recognition unit (FRU) will signal familiarity and will access semantic information via activation of a person identity node (PIN); full recognition and recall will be achieved with sufficient activation at the final level, which is name generation.

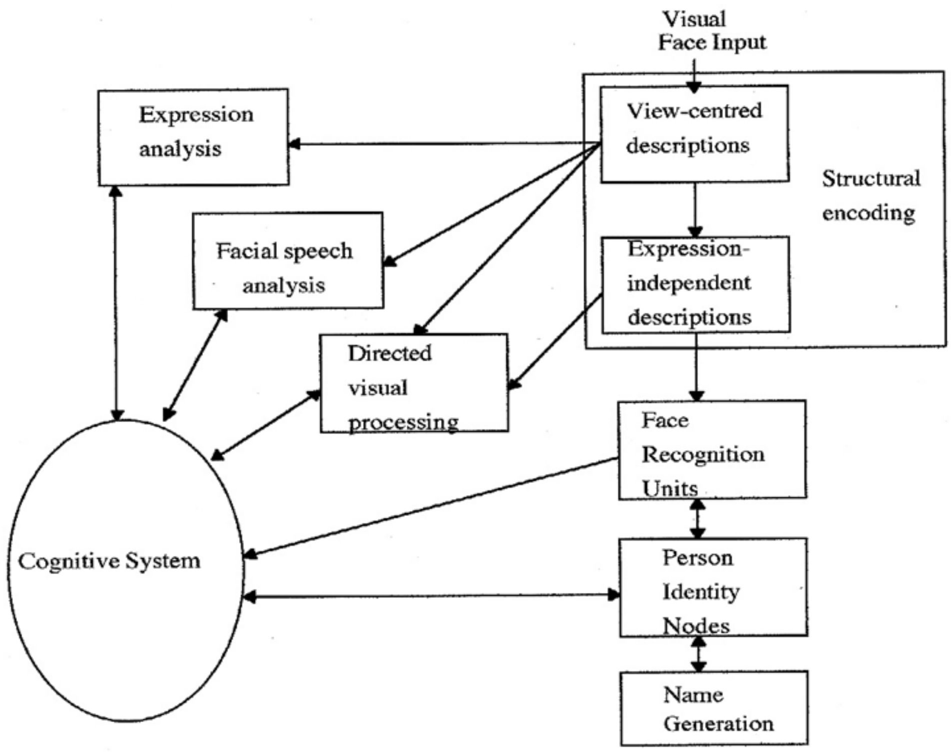


Figure 1.1. A reproduction of Bruce and Young's (1986) face processing model.

The sequential nature of the model is consistent with a diary study of face processing errors: twenty-two participants reported details of 922 face recognition errors over a period of seven weeks; perceived familiarity without retrieval of semantic information, and thus recognition failure, was prevalent, but although recognition without being able to remember a name or some other personal information was also common, the situation where a name could be recalled without first achieving recognition and recalling semantic information never occurred (Young, Hay, & Ellis, 1985).

A sequential process would also entail that familiarity judgements are reached faster than judgements that involve semantic information (Young et al., 1986), while familiar faces should be classified faster by semantic information (e.g. occupation) than by name (Bonner, Burton, Jenkins, McNeill, & Bruce, 2003; Carson, Burton, & Bruce, 2000; Johnston & Bruce, 1990; Young et al., 1988). However, reaction time studies with children (Calderwood & Burton, 2006; Rahman, Sommer, & Olada, 2004; Scanlan & Johnston, 1997) and adults (Bredart, Brennen, Delchambre, McNeill, & Burton, 2005) have also shown that if the people are highly familiar, or if the semantic information that is requested is less well known, naming judgements will be reached faster than semantic decisions. While this suggests that the serial model may fail to accurately represent name and semantic information retrieval, it should be noted that failure to report a specific item of semantic information does not necessitate that no semantic information was recalled; particularly in the case of children, it is not apparent what semantic information or cues would be salient in learning and encoding person information.

Recognition begins with perception and extraction of pictorial and structural information. Pictorial codes are specific to the image properties and context including lighting, pose and expression; structural codes are richer and include information about the three dimensional and invariant properties of faces. These are developed over repeated exposure by combining information over transitions of lighting, expression, viewpoint etc. Unfamiliar face perception allows extraction of a pictorial code and a limited context specific structural representation, but repeated exposure, and hence familiarity, is required to form a more complex presentation that can overcome changes in viewing conditions to enable face, rather than image, recognition.

Within the model, recognition requires a match between the encoded structure and a structure that is stored in memory; for familiar faces that have an elaborate structural representation, recognition can be achieved with even seriously degraded images (Burton, Wilson, Cowan, & Bruce, 1999), but the ability to recognise an unfamiliar face will be dependent upon the encoded stimulus having a close pictorial correspondence to the limited context specific representation that was previously stored. Bruce (1982) demonstrated that although responses were slower, familiar face recognition easily accommodated changes to viewpoint or expression, but that unfamiliar face recognition was significantly impaired by any change of image between study (initial encoding) and test. Indeed, even in a perceptual matching task, structural codes that are dependent upon the image properties make it difficult to form correspondence between different pictures of the same unfamiliar person (Bruce et al., 1999). In order to understand how correspondence between the unknown face images might be achieved at all, it is useful to consider the way in which face images may be stored in memory, and how knowledge of natural facial variation must be employed to compensate for image discrepancies.

Valentine's multi-dimension face space model (1991)

Valentine's (1991) multi-dimensional face space model (MDFS) describes a face memory metaphor that has been extremely influential in face processing research. The concept employs a Euclidian framework within which every face that is encountered will be encoded along an unspecified number of dimensions, and will contribute to the formation of a face memory structure. While the dimensions have never been defined, they are generally held to be any characteristics that are useful for face individuation. Faces within the multi-dimensional model are assumed to have a normal multivariate distribution meaning that they will cluster towards the centre or origin of the space: typical faces are more prolific and will be densely grouped around the central tendency, while more unusual or distinctive faces will occupy sparser positions further away from the origin on at least one dimension. A representation of the MDFS model is shown in Figure 1.2.

Two variations of the model are proposed: in the exemplar based model each face would be coded relative to other face exemplars and perceptual similarity would be determined by Euclidian distance and exemplar density; faces that are close together will look more alike than those that are positioned further apart; within the norm based model, faces would be encoded relative to a norm or prototype face at the centre of the space; individual faces would radiate away from the norm on vectors, and similarity would be a function of distance from the norm and separation from adjacent vectors.

With perceptual learning and encoding of all the encountered faces over time, the multi-dimensional space will become tuned around the dimensions that are most useful for face discrimination (Kelly et al., 2009, 2007, 2005; Nishimura, Maurer, & Gao, 2009). Newly encoded faces would be assimilated with existing face representations, and the stored knowledge should, therefore, provide the basis to interpret a novel signal and form a limited representation from which to perform any similarity or categorical judgments. An example of this would be perceptual discrimination of unfamiliar faces; average faces would be encoded close to densely clustered typical faces with the consequence of discrimination errors, while unusual faces would be encoded some distance from the central tendency and, unless positioned very close to another unusual face, perceptual distinctiveness and a discrimination advantage should be assured.

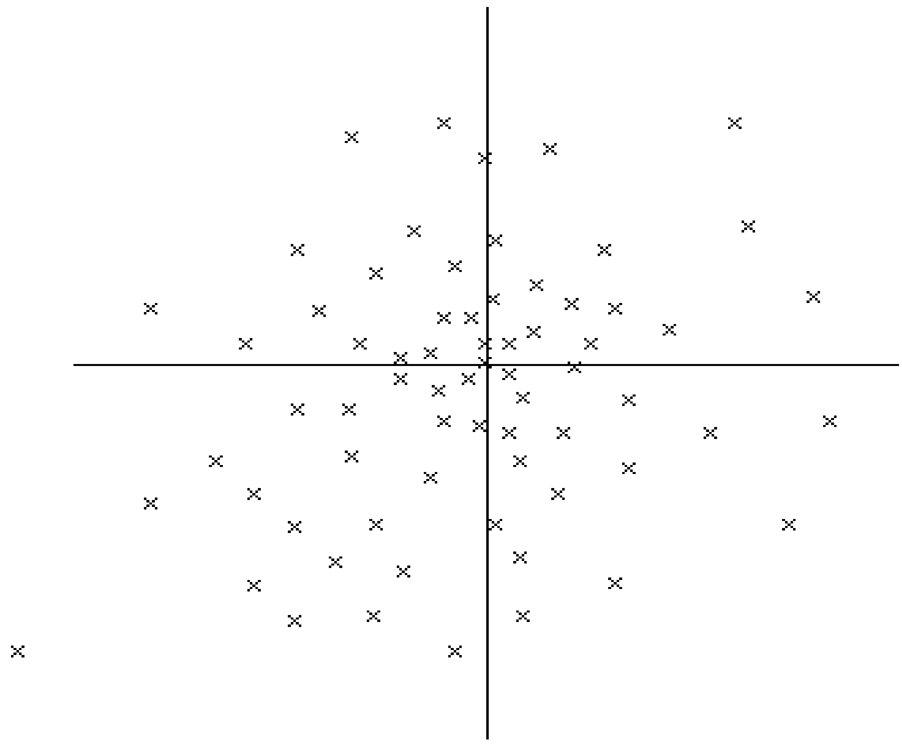


Figure 1.2. A 2 dimensional representation of a multi-dimensional face space. Each x represents a face and the origin represents the central tendency. A normal multivariate distribution means that faces will cluster around the origin of the space with typical images being close to the centre while distinctive faces will occupy more isolated positions further away.

Valentine and Bruce (1986a, 1986b) found that distinctive faces were recognised more easily than typical faces, but that they were also harder to classify as a face. Within MDFS a sense of 'faceness' will be shaped by the facial attributes that have been most frequently encountered; distinctive faces that share fewer of these characteristics will therefore be less face-like but most easy to discriminate relative to the other more typical images. The concept of distinctive faces as less face-like is reinforced by substantially weaker inversion effects relative to typical faces and thus less reliance on holistic processes (Valentine, 1991). The research presented in Chapters 5 and 6 explores face perception within the MDFS metaphor; in Chapter 5 caricature is employed to increase facial distinctiveness by exaggerating the differences between individual face images and a norm face. Within a MDFS this technique would increase the distance between each image and enhance perceptual discrimination. In Chapter 6 the concept of perceptual expertise is evaluated by perceptual discrimination of infrequently encountered other race faces.

Perception and Bias in Face Memory

Own race bias in face perception

Faces of one's own ethnic race are discriminated and remembered more effectively than faces of another race. This is termed the own race bias (ORB), or cross race effect (see Meissner & Brigham, 2001 for a review). ORB is one of the most studied effects in face perception and has prompted a number of potential explanations. The contact hypothesis suggests that exposure to individuals of another race enhances the ability to individuate them (Sporer, Trinkl, & Guberova, 2007; Walker & Hewstone, 2006); while social utility theories suggest that the social minority, or out group faces, are disregarded and hence encoded at a shallower level (Malpass, 1990; Rodin, 1987). One of the most influential theories has been the perceptual expertise hypothesis (MacLin & Malpass, 2001); by this account, limited exposure to other race faces means the appropriate range of facial variation to distinguish them is never learned. This view accords with Valentine's (1991) multi-dimensional space models (MDFS) of face processing. Developmental studies are also consistent with this position; at birth humans have no preference for own race faces but with limited exposure and perceptual learning, tuning towards the race of faces that are encountered, is evident throughout the first year (Kelly et al., 2009, 2007, 2005). Within a face space, other race faces that are encountered infrequently will be distinctive on at least one race specifying dimension; faces of a particular race will, therefore, be encoded together on this dimension but will be positioned some distance from the central tendency, and thus removed from the central locus of perceptual sensitivity (Rhodes & McLean, 1990).

Another theory concerns the way that the faces images may be analysed, and proposes that other race faces do not employ expert holistic face processing but are processed in a featural, bottom up manner. Evidence in support of this theory is found in studies that show weaker inversion effects for other race faces (Rhodes, Brake, Taylor, & Tan, 1989), and in weaker composite face effects when different face parts are combined (Michel, Rossion, Han, Chung, & Caldara, 2006), both of which indicate less reliance on holistic processes. However, MacLin and Malpass (2001) employed the same ambiguous face set with typical Hispanic or African American hairstyles, and found that the race typical hair cue altered perceptual judgements and memory of the exact same facial features and configurations, providing a strong indication that racial categorisation will determine whether or not expert face processing strategies will be employed (see also Michel, Corneille, & Rossion, 2007). Levin (2000) suggested that when race is coded as a feature this will be at the expense of individuating information, but as enhancing depth of processing does not reduce ORB (Chance & Goldstein, 1981; Devine & Malpass, 1985; Sporer, 1991), other race face processing appears to differ in character rather than just quantity. The research in Chapter 7 explores perceptual discrimination of own race and other race faces and the MDFS model is used to conceptualise a method of causing other race faces to vary in the same way as own race faces, thus enabling the study of both configural sensitivity and categorical processes.

The smiling face bias in face perception

The Bruce and Young (1986) model proposes that expression and identity are processed by distinct and parallel pathways *(see figure 1)*, and has been supported by neuropsychological studies showing dissociation of identity and expression analysis in individuals with brain injury (de Haan, Young, & Newcombe, 1992; Hornak, Rolls, & Wade, 1996; Tranel & Damasio, 1985; Tranel, Damasio, & Damasio, 1988); in neurophysiological observations of distinct ERP patterns (Caharel, Courtay, Bernard, Lalonde, & Rebai, 2005), and in cerebral blood flow in different areas of the brain during identification and expression tasks, although other areas were also activated in both to a lesser degree (Sergent et al., 1994). However, smiling expressions have been shown to increase judgements of familiarity for both familiar and unfamiliar faces (Baudouin, Gilibert, Sansone, & Tiberghien, 2000; Davies & Milne, 1982; Gallegos & Tranel, 2005; Kottoor, 1989), while negative expressions had the opposite effect (Lander & Metcalfe, 2007). When faces were smiling, familiar images were also accepted faster and it took longer to reject unfamiliar faces, suggesting that positive affect caused the faces to be assessed more carefully (Endo, Endo, Kirita, & Maruyama, 1992).

Some selective attention studies have found no interaction of expression or identity processing (Calder, Young, Keane, & Dean, 2000; Ganel & Goshen-Gottstein, 2004), while others found that identity interfered with expression discrimination but that expression did not influence identity judgements (Le Gal, 1999; Schweinberger & Soukup, 1998; Schweinberger, Burton, & Kelly, 1999). Ganel and Goshen-Gottstein (2004) propose this occurred because identity decisions in these studies were faster than the expression analysis: they employed highly similar face pairs to make identification more difficult, and found interference from expression to identity judgements for both familiar and unfamiliar faces. This view is consistent with Bruce (1982) who found no effect of expression on recognition of relatives, and Endo, Endo, Kirita and Maruyama (1992) who found a smiling advantage for familiar faces but the opposite effect for relatives. Thus the smiling face recognition bias appears to be restricted to instances of uncertainty which is not found when faces are identified easily.

Collectively these studies suggest some independence of identity and expression analysis, but that under certain conditions they do interact. Capgras syndrome is a striking disorder where sufferers recognise faces but experience the disturbing belief that their loved ones have been replaced by aliens or impostors. Ellis and Young (1990) proposed that the Capgras delusion results from recognition without any confirming reaction from an emotion related secondary route. Thus, interactions with familiar people are expected to be characterised by emotional sequelae. In typical face perception, if there is little room for doubt, identity judgements are rapid and will be made without recourse to affective information, but where a judgement is difficult or must be considered, the evidence suggests that the quality of affective information or response will play an important role.

Within the context of the Bruce and Young model (1986) recognition of an unfamiliar face would rely on some correspondence between the novel encoded face image and a previously stored structural representation. As these representations would not be identical it would take time to overcome image disparity (Bruce, 1982), which would allow affective information from the parallel expression analysis route to become available and to contribute secondary affective information; the evidence suggests that when the affective information has positive valence this will promote perceptions and judgements of familiarity but if the affective signal is negative, it will have the opposite effect. Chapters 3 and 4 explore the influence of positive expression on facial composite identification and on facial image comparison judgements. The following sections will discuss forensic facial identification.

Forensic Identification

Face recognition and the law

Face perception is such an inherent part of daily function, that in spite of occasional memory lapses, we generally assume the process to be straightforward and fairly unremarkable. For most people that assumption will remain largely unchallenged, but for those with the misfortune to experience a crime the real difficulty in identifying or remembering unfamiliar faces will become woefully apparent. The psychology of eyewitness memory has been studied since the late

1800s (Cattell, 1893; Stern, 1939), with Von Schrenck Notzing providing the first expert legal testimony on eyewitness suggestibility in 1896 (Blackburn, 1996) and the first forensic psychology text being published by Hugo Munsterberg (1908). Around that time, a high profile case of mistaken identification in the UK led to the arrest and false conviction of Adolph Beck in 1896, and then again in 1904. When the real perpetrator, Wilhelm Meyer, was arrested, Beck received a full pardon from the King (1904) and a Committee of Inquiry resulted in the creation of the Court of Criminal Appeal (1907).

Continuing concern regarding eyewitness identification (e.g. R v Dougherty, 1973; R v Virag, 1974) prompted the UK government to commission the Devlin report (1976), which called for more psychological research to develop understanding and safer identification procedures. The report recommended that no conviction should be based on eyewitness identification alone *(this requirement was already established in Scottish law)*, and that where a conviction rests substantially on identification evidence, the jury should be cautioned to the high possibility of error. The Court of Appeal formalised the recommendations, creating The Turnbull Guidelines (R v Turnbull, 1977) which state that where a prosecution depends on visual identification, or where eyewitness identification is disputed, the judge must warn the jury that error is common and caution is needed in the interpretation of honest and sincere facial identification; the judge should explain why this is the case and explicitly address factors that are important for identification within the case in question (i.e. familiarity with the suspect, visibility and distance, length of time to view the suspect, and time passed since the suspect was seen).

With the introduction of DNA evidence (e.g. Andrews v Florida, 1987; R v Pitchfork, 1987) the extent of the eyewitness recognition problem is now known (Connors, Lundregan, Miller, & McEwen, 1996; Dwyer, Neufeld, & Scheck, 2000), and recent figures cite misidentification as a major factor in 75% of overturned convictions (The Innocence Project, 2011). A great deal of research has now focussed on unfamiliar face recognition and the potential of eyewitness misidentification to cause miscarriages of justice (e.g. Loftus, 1992; Rattner, 1988) and legal reform and level of awareness is testament to the significant contribution of psychological study to create a safer judicial process. However, not all identification protocols have been the topic of acceptable levels of study and procedural knowledge in face perception and discrimination is still remarkably poor.

Face perception and the law

Code D of the Police and Criminal Evidence Act (PACE, 1984) is a code of practice for the identification of persons by police officers, and provides detailed requirements for the use of identity parades, video line-ups, photograph identification, as well as live confrontations or identification of people within groups. The document is updated to incorporate new knowledge, and is the product of decades of research and legal advancement. However, given the prevalence in contemporary culture of CCTV, photographic identification, and the introduction of facial biometrics, the document provides no guidance on identity verification from facial image comparison or on procedures for matching live people to photographic or video images. This contrast suggests that there is little knowledge of perceptual discrimination of unfamiliar faces at an operational level.

In a recent review of facial image comparison (FIC) for The Home Office Scientific Development Branch (Hancock & McIntyre, 2011), police authorities throughout the UK were canvassed to determine current police practice and knowledge. None of the authorities provided, or were aware of, any FIC training. In day to day policing identity verification (from person to photo or from comparison of facial images) is at the discretion of individual officers who have no specialist knowledge regarding the factors important for accuracy, or of typical levels of error. In terms of FIC awareness, where images were of good quality officers believed that face matching was obvious and could be accomplished with ease, but where images were difficult to match (i.e. image quality is poor, or extensive time has lapsed between image captures), they believed that forensic services should be employed to perform facial mapping techniques and provide independent expert testimony. This state of knowledge is officially endorsed by The Attorney General's Reference No. 2 of 2002, which states that if an image is 'sufficiently clear', jurors may be asked to determine the identity of the defendant, and that 'qualified' facial mapping experts may be called upon to present expert testimony in court proceedings. Given that face matching errors are typically in region of 30% (e.g. Bruce et al., 1999) and that there is no scientific evidence to support facial mapping techniques (Campbell-Tiech, 2005; Edmond, Biber, Kemp, & Porter, 2009; Strathie, 2010), research and collaboration is clearly necessary to develop a level of awareness and understanding of unfamiliar face discrimination that can contribute to the development of formal FIC guidelines.

Within forensic identification, facial composites are a special circumstance where memory and perception of face images must be combined. Facial composites are likenesses that are produced by witnesses following a crime, and are intended to elicit leads to assist an investigation. To be effective a composite must match the ability of the witness to remember and recreate an unfamiliar face, with the attributes that are necessary for familiar face recognition. The following section discusses this topic in detail.

Facial Composite Recognition

Facial composites[1] are produced following a crime, and in an investigation with little or no information leading to a suspect, they are an invaluable way of enabling a witness to communicate the appearance of a perpetrator: the resulting likeness can then be issued to police officers and other forces, and may be released to the media in an attempt to generate information. Typically a cognitive interview will be conducted to enhance recall and obtain a description of the perpetrator (Fisher, Geiselman, & Raymond, 1987; Geiselman, Fisher, MacKinnon, & Holland, 1985); the witness will then work with a forensic artist or a composite system operator to produce a facial composite image.

Facial composite systems and methodology

Current facial composite systems are all computerised but work in one of two ways; the older featural systems, such as ProFIT and E-FIT, require the witness to view a selection of features based on their description, the features are presented within the context of a complete face image, and the items that most resemble their

---

[1] The term facial composite is sometimes employed to describe any face image that is generated by combing parts or elements of more than one face. The term facial composite within this work refers exclusively to forensic likenesses that are generated by witnesses to portray a facial memory. The term should not be confused with the composite face effect reported by Young, Hellawell, & Hay, (1987) to describe processing of separate face halves presented in close alignment.

memory of the person (i.e. the best eyes, nose, mouth, face shape etc.) are adjusted for size and relative position to produce an overall likeness. Artistic enhancement tools within the systems can also be employed to introduce alterations or additions, such as stubble, lines, or modified hairstyles. Newer composite systems, such as EvoFIT and E-FIT-V, are deemed to be holistic in nature as the witnesses are shown whole face images and asked to select the images that most resemble the perpetrator. These systems then employ genetic algorithms to breed together the selected images, until an acceptable facial likeness is achieved. Enhancements can also be applied to the images, most notably and effectively on holistic psychological dimensions, such as masculinity (Frowd, Bruce, McIntyre, & Hancock, 2006) *(for more complete details of construction procedures see Frowd et al., 2004).*

Experimental methodology

Experimental facial composite research endeavours to mimic real life conditions as much as possible. Witness participants are employed to generate composites of strangers; first they are briefly exposed to an unfamiliar target via videos, photographs, or in person, then they will generally return to the lab at another time to construct a facial composite likeness. Delays may range from a few hours to a week, but two days is common. Before composite construction, the witness participant will receive a full cognitive interview, which includes rapport building and context recreation to enhance recall and communication. They will then work with the composite system operator to produce the best possible likeness. When the composites are complete they will be shown to a second cohort of participants who will be familiar with the targets, in an attempt to secure identification.

Faces of relatively well-known people are often used as targets because it allows witness-participants who are unfamiliar with particular celebrity faces (e.g. footballers, or soap stars) to produce a set of composites which are also potentially identifiable by a wide sample of other participants. The aim is to simulate real-world situations where composites are produced by unfamiliar witnesses, but must be identified by a member of the public sufficiently familiar with the target. However, celebrities are not personally familiar to most people, and it is possible that their faces are not processed or stored in the same way as people who are familiar from live interaction (e.g. Endo et al., 1992). The alternative method is to use personally familiar targets and to secure an adequate amount of identification data, i.e. for statistical power an appropriate number of participants must be able to provide naming data for a suitable number of composite items, this is typically achieved by sampling targets and participants from within an occupational setting. Unfortunately while this removes any celebrity confound, it introduces a new one as the participants can resort to a process of elimination rather than actual face recognition; for example if the image has curly hair it must be a certain person. Such a process would obscure or distort inferences about face processing and could be considered to have less ecological validity than the celebrity composite paradigm. In order to address both concerns, the research in Chapter 3 employs both famous target composites and personally familiar target composites with the addition of unfamiliar distracters to reduce the possibility of alternative choice guessing strategies.

Despite the fact that facial composite systems have become rather sophisticated and can assist unfamiliar witnesses to produce images that are independently judged to be good likenesses, they often fail to produce good rates of recognition (Brace et al., 2008; Bruce et al., 2002; Davies et al., 2000; Frowd et al., 2004; Pike, Brace, Turner, & Kynan, 2005) To achieve successful identification the facial composite will need to be processed and recognised as a familiar face, therefore attempts to improve effectiveness should begin with an understanding of familiar and unfamiliar face processing, and an appreciation of how facial composite processing may differ.

Unfamiliar witnesses

Essentially the facial composite is not a picture of the perpetrator, but a portrayal of an unfamiliar face memory. This is an important distinction. The witness will have had limited exposure to an unfamiliar person, usually over a short period of time, and from this they will have been able to form a very limited context specific structural representation. They will then have to recall this information in order to provide a verbal description, and then they will have to recognise facial information during the composite construction process (Brace et al., 2006). The person and face description is an important part of the legal process, and PACE (1984) guidelines require that the initial verbal description is recorded prior to implementation of any identification or facial imaging procedures. However, describing facial information is extremely difficult (Laughery & Fowler, 1980) and there is some indication that subsequent face recognition will be impaired because the verbalisation process 'overshadows' the perceptual memory of the face (Schooler & Engstler-Schooler, 1990). An alternative explanation is that generation of the verbal description is a featural task which then impairs holistic recognition of a complete face image (Dodson, Johnson, & Schooler, 1997; Fallshore & Schooler, 1995). It is highly likely that a witness would attempt to memorise the perpetrator's face by attending to each facial feature in turn, or to the features they consider to be most distinctive or memorable (e.g. Cotton v North Carolina); this might in fact benefit the composite construction procedure because the verbal description would match the face encoding strategy, and also the composite construction process of the featural composite systems.

Featural and holistic processing

Frowd et al. (2008) examined the influence of featural versus holistic encoding strategies: witness participants were asked to form holistic personality judgements, or to memorise the facial features at the time of encoding; they were then required to generate a facial composite two days later using either a featural or a holistic composite system. Inducing holistic personality judgements enhanced the quality of subsequent featural composites, while featural encoding was most effective for the composites produced using the holistic system. Although the effects were counter to predictions, they were interpreted to mean that featural encoding enables featural information to be extracted within a holistic interface, while holistic encoding facilitated recognition of features within the context of the composite construction face of the featural system (e.g. Tanaka & Farah, 1993). In a separate study Frowd et al. (2008) investigated the effectiveness of holistic verbal descriptions, and found that adding a holistic stage to the traditional cognitive interview significantly improved the quality of subsequent composite construction. It would seem, therefore, that ensuring that recall and recognition engage both featural and holistic processes is likely to produce the most identifiable composite images.

Familiar recognition and inaccuracy in composite images

Limitations of unfamiliar witness memory and recall, combined with the constraints of facial composite systems and the difficulty of communicating facial information (Brace et al., 2006), mean that to some extent any composite likeness will be inaccurate or incomplete in terms of the features, the configural arrangement, or most likely both. Where a number of witnesses have created facial composites of the same individual, presenting all of the composite images for identification can improve levels of recognition (Brace et al., 2006), presumably because across the range of composite images there will be correspondence and replication of accurately remembered information. Multiple composites of the same individual can also be successfully combined into a 'super-likeness', by morphing all of the images into one; this will have the effect of averaging out the facial composite elements that differ across images, while retaining those that the different witnesses have agreed upon (Bruce et al., 2002; Valentine et al., 2010).

In most cases however, there will be just one imperfect facial composite likeness; familiar face perception has the benefit of complex structural representations but while this means that face perception can accommodate substantial variation across images (Bruce, 1982; Burton et al., 1999), it also entails extreme sensitivity to feature position and the configural relationships (Haig, 1984) and studies have found that alterations to the configural arrangement of a face can impair identification of the featural information (Hosie, Ellis, & Haig, 1988; Tanaka & Sengco, 1997). Within a facial composite it is likely that the overall configuration will be incorrect and recognition will require identification of accurate facial features; but in keeping with the findings of Young, Hellawell and Hay (1987), the inaccurate composite configuration would be processed in a holistic manner, which would make it difficult to extract or recognise any accurate component information (Hosie et al., 1988; Tanaka & Sengco, 1997; Tanaka & Farah, 1993). The research described in chapter 2 examines this concept and provides evidence that facial composites are perceived holistically as novel faces that don't correspond to any stored representations; therefore important and accurate information within the composite images cannot be recognised. By adopting the split-image technique employed by Young et al. (1987), holistic face processing is impaired to enable enhanced component identification and composite recognition.

Recognition bias and facial composites

Facial composite recognition depends upon a match between the information that the unfamiliar witness can provide and the information that is needed for familiar face recognition. Inaccuracies in the composite image mean that it is difficult to correspond with a stored face representation to achieve identification; but given that some of the information may be correct, it is possible that a feeling of familiarity will be triggered that either doesn't produce specific person information, or a person may come to mind but there is insufficient confidence to offer up a name. In conditions such as these the cognitive task is not straightforward and supplementary information in the form of heuristics (Tversky & Kahneman, 1974), schemas (Rakover, 2002), and affective responses (e.g. Baudouin et al., 2000) may bias judgment and decisions.

When a person is exposed to a crime it will generally be experienced as a negative and highly emotional event and when they subsequently receive a cognitive interview, it will include context reinstatement to enhance recall and the negativity associated with the memories will be reinforced; if the witness then goes on to produce a composite likeness their memory of the face and the resulting likeness will also reflect these negative qualities *(personal communication; Ann Parry Metropolitan Police & Janet Richardson, Forensic Artist)*. The smiling face familiarity bias has consistently been found in old/new recognition tasks, rating tasks, and reaction time data (Baudouin et al., 2000; Davies & Milne, 1982; Endo et al., 1992; Gallegos & Tranel, 2005; Kottoor, 1989). In contrast, it has been shown that negative expressions cause faces to be judged as less familiar (Lander & Metcalfe, 2007). The smiling face bias is apparent in conditions where identification is not easy: facial composites are very difficult to identify because some of the information will be inaccurate; consequently, the affective quality of the image will be important and if the composite is intentionally negative, affective bias will influence judgement away from forming a recognition response. The research in Chapter 3 explores the impact of expression in facial composite recognition and confirms that affective quality does exert a powerful effect that can be reduced with imaging techniques.

Facial image comparison

It would be intuitive to assume that identification errors are caused entirely because we have no stable memory representation of an unfamiliar person; and it would also be reasonable to suppose that if we compare two faces, that much like any other class of object, visual perception would enable us to tell them apart. However, even when there is no memory load and face images must be simultaneously matched or compared, identification errors are very common. Facial image comparison (FIC) is a term that is used within the security communities (e.g. UK Home Office or Federal Bureau of Investigation) to describe the simultaneous perceptual comparison of two or more face images, or the comparison of a person with a facial image; traditionally in the face perception literature this has been simply termed as face matching.

Within daily life it is now common practice to employ photographic identification for everything from gym membership to national passports, and you will only be allowed to board a domestic aeroplane flight if you can provide an official document showing a suitable facial likeness. Because facial comparison has become the most commonly employed security metric (unlike iris or fingerprint recognition, FIC can be accomplished with unwilling targets, and can be conducted covertly or from a distance; as such it is an important part of any security, surveillance, or intelligence operation) FIC is one of the fastest growing areas of national and commercial security, and is the focus of considerable investment by the Home Office and FBI, who recognise a need to understand, educate, and optimise identification procedures that will avoid costly error.

FIC methodology

In experimental work face matching performance is typically assessed using a 'target' or probe image, that must be matched to an item within an array of images that are superficially similar; the size of an array depends on the available materials and the research question, but 8 – 10 items is common. Visually the 'target' will be presented either above, or to the side of the array items, with all of the images available for comparison: this is termed a parallel or simultaneous matching array, and in forensic terms, is like a biometric interface that displays the closest hits for comparison, or to matching of CCTV or surveillance shots to archival images or 'mug books'. The array may contain a second image of the target, together with the appropriate number of distracters or 'foils' (target present), or it may contain only suitable distracter items (target absent): participants will generally be asked to make an identity match selection, or indicate that the target is not represented. Where the target is not present the response may be correct rejection of the array (CR), or a false positive selection may be made (FP): where the target is represented in the array, the correct item may be selected (Hit), the array may be incorrectly rejected (Miss), or a false identification may be made (FID). Inclusion of both target absent and target present arrays mimics the uncertainty of real life and in experimental work the ratio is typically 50:50. However, false identification will be reduced if participants believe that true matches are rare; but as fewer items will be selected overall, there will also be contingent reductions in correct identification and higher miss rates. For this reason, simple measures of accuracy cannot fully capture face matching ability or performance on a given identification task: signal detection analyses of hits and false positive rates can, however, usefully determine additional measures of sensitivity and response bias, while correlation analysis of these measures has also been employed in the study of individual differences (Megreya & Burton, 2006, 2007).

Where it is not necessary to collect an explicit measure of false positive identification, only target present arrays can be employed and the task is considered alternative forced choice (AFC). A selection is usually required and this format is typically employed with smaller 2 or 3 item arrays where an aspect of the array items (e.g. expression) is of principal importance. This procedure has no direct forensic application but enables the study of how image characteristics may influence acceptance of one image over another. The simplest perceptual matching task presents pairs of images that require a response that the images either match (i.e. they are of the same person), or that they do not; this procedure is most similar to the individual face matching commonly undertaken with identification cards and passports. Task difficulty for FIC is therefore a function of the similarity of target and distracter images, and of the probability of selecting an item by chance (e.g. chance performance is 1/number of items in the array; thus in a 2 item AFC task, chance is 50%, while for 10 item arrays it will be just 10%). To reduce the effect of chance responding and enable comparison of the variables that are under study, experimental work most commonly employs multiple item arrays.

In comparison with face recognition, FIC is relatively undeveloped in the experimental literature, and face matching procedures in the laboratory have typically employed simultaneous arrays in a format similar to the line-up identification procedures traditionally used in police work. In memory research this procedure has been associated with a relative judgement strategy, i.e. picking the item that fits best or is closest, rather than a definitive correct match (Lindsay & Wells, 1985; Wells, 1984); the procedure may therefore lower the criterion to accept a likeness, producing more false positive identifications (Lindsay & Bellinger, 1999; Lindsay & Wells, 1985; Smith, Stinson, & Prosser, 2004). A sequential line-up procedure which involves making an unqualified decision about each item in turn, is believed to encourage absolute judgements, and has been found in some studies to reduce false positive identifications (Lindsay & Bellinger, 1999; Lindsay & Wells, 1985). Face matching employs perception without memory load, and is an ideal paradigm to investigate perceptual discrimination and judgement strategies; to investigate the effect of presentation format for FIC, the research in chapters 4, 5, and 6 assesses the effectiveness of both sequential and simultaneous presentation.

FIC performance

In order to determine optimal face matching performance and evaluate degradations caused by changes in viewpoint and facial expression, Bruce et al. (1999) compiled face matching arrays from high quality photographs of 120 male police officers (18 – 35 years) obtained from the Home Office Police Information Technology Organisation (PITO). The professional quality images portrayed full face neutral expressions and were controlled for viewpoint in diffuse studio lighting that avoided shadowing. The target images to be matched to the arrays were stills taken from VHS quality video sequences of 80 of the police officers, which were captured on the same day, and therefore contained no variation in appearance other than method of capture. These showed full face neutral poses, full face smiling poses, and a neutral 30° angled view turned in either clockwise or counter-clockwise direction. The photo arrays were constructed for each target on the basis of similarity ratings provided by 80 participants, such that each target absent array comprised the ten photographs of other officers that were judged to be most similar to the target; the least similar of these was replaced with the target's own photograph for target present conditions (Bruce et al., 1999).

Optimal performance was expected for matches of full face neutral video targets to the full face neutral photo arrays in which only the image media differed; however, in spite of instruction that the target would only be present in half of the trials, correct identification and correct rejection of target absent arrays reached just 70%. As expected, changes in viewpoint with neutral targets shown at a 30° angle reduced performance to 61%, while changes in expression with full face smiling targets produced accuracy levels of just 64%. Familiarity influences the type of information that will be employed in face recognition, such as the pronounced importance of external features for unfamiliar faces (Ellis et al., 1979; Fletcher et al., 2008; Frowd et al., 2007; Young, Hay, McWeeny, Flude et al., 1985). Bruce et al. (1999) confirmed this effect in the face matching task by showing that matching performance on the basis of external face shape and hairstyle was 73%, and fell to just 49% when that information was not available. Given that all of the images were captured on the same day, an external feature bias would have been reasonably successful but with images captured on different occasions, image correspondence would be poorer and error rates would be markedly higher. To exclude the effects of uncertainty and response criterion, a further experiment employed 10 item AFC arrays with the target always represented: participants were instructed to select the item that most resembled the target but while performance was improved, incorrect selections still reached 21%. When the other variables were controlled, the image properties of the photographs and video stills created sufficient variation to significantly impair perceptual matching of the unfamiliar images.

Image quality

The results indicate that unfamiliar face perception is not sufficient to overcome image variation and extract stable facial attributes from different images. With images captured at different times, or with poorer equipment or lighting, performance is likely to be substantially worse. In a study designed to employ realistic CCTV footage of a mock robbery, Henderson, Bruce, and Burton (2001) demonstrated the consequences of poor quality images for accurate face matching.

Participants were shown a series of video stills for each of two robbers and attempted to match each robber to an 8 item array. Correct identification across the arrays was just 20%, with incorrect selections being made more than half of the time. When the CCTV stills were replaced with broadcast quality images performance increased to a level of 64%, which then fell to 43% when the targets were shown wearing hats. As one would expect, ability to match photographs of the robbers to the video images was significantly impaired by occlusion of external features with hats, but the effect of poor CCTV image quality was considerably worse.

Viewpoint and lighting

Perception of unfamiliar faces is dependent upon image properties, and in addition to image clarity, appearance of facial shape and 3-D structure will be dictated by angle of view and reflectance from the available lighting. Adini, Moses, and Ullman (1994) found that images of the same person could appear less similar when lighting and viewpoint changes were made, than images of different people when lighting and view were maintained. Hill and Bruce (1996) studied the effects of lighting and viewpoint on perception of 3-D head models; they found that when viewpoint was held constant, changes to the direction of lighting significantly impaired face matching, while changes of viewpoint could be accommodated fairly easily if the head was lit from above. Lighting is generally assumed to come from overhead (Ramachandran, 1988); perception of faces across viewpoints will be influenced by this and with lighting from other directions, differences will be more difficult to reconcile.

Matching people and face images

In forensic and security settings comparison of pairs of face images is more common than comparison of an array of faces, and while chance performance is higher, one would expect the task to be less demanding and therefore easier. Unfortunately when participants were asked to judge whether pairs of faces were of the same or different people, Megreya and Burton (2006, 2007) reported accuracy levels of less than 80%, while Henderson et al. (2001) reported a false positive rate of 27.5%, and a correct identification rate of just 55%. Performance is also extremely poor when a live person is matched to images in the form of a photo credit card. Kemp, Towell, and Pike (1997) conducted a study in which FIC performance of supermarket cashiers was evaluated when student 'shoppers' presented items for purchase and attempted to 'pay' with one of four photo credit cards: the cards portrayed a good quality photograph of them that corresponded with their current appearance; a good quality photograph of them that did not correspond with their current appearance; a photograph of someone else who resembled them; or a photograph of another person who did not resemble them. The cards displayed an incorrect image half of the time and to ensure 'shoppers' were blind to the condition, were presented in opaque wallets. Cashiers checked the signature and photograph, and called a supervisor if they wished to reject either. Overall accuracy was 67.4%: 9.8% of the correct cards were declined while 63.6% of the similar incorrect cards (i.e. the fraudulent ones) and 34.1% of the dissimilar photo-cards were accepted. Given that the cashiers knew they were being observed, and in view of the importance of FIC at border and airport security points, these results are of particular concern, although experts may generate fewer face matching errors.

Individual differences and expertise

Wilkinson and Evans (2009) compared face matching ability in the general population (n = 61), with two facial imagery experts and reported that as the specialists were more accurate, training and experience produce more reliable expert facial identification. However, the extant literature confirms substantial individual differences in unfamiliar face perception (Bruce et al., 1999; Bruce, Henderson, Newman, & Burton, 2001; Megreya & Burton, 2007), making the result difficult to interpret, as no measures of variability or effect size were reported. It is also unclear how the statistical analysis was handled with such radically different group sizes. It is, however, notable that the experts generated an incorrect identification rate of 25% and that they both wrongly identified the same individual. Given the weight that expert testimony would be given in court, the results make a convincing case for caution rather than endorsement of face matching expertise.

In a separate study, face matching performance of trained experts (n = 14) was again contrasted with an untrained sample (n = 28) (Lee, Wilkinson, Memon, & Houston, 2009). Overall accuracy was 67.3%, but did not differ between the untrained sample and the experts. What is more, when the experts were considered on the basis of length of experience, no difference was observed. While these results suggest that training is unlikely to generate safer face matching, brief familiarisation with images can enhance discrimination (Bruce et al., 2001; Clutterbuck & Johnston, 2004; Megreya & Burton, 2006, 2007), and appreciation of face processing effects can provide insight into the development of safer face matching procedures.

Distinctiveness and discrimination

Within the concept of MDFS any characteristic that causes a face to deviate from the norm is considered to be distinctive, and because a distinctive face representation will be located further away from the origin and from typical face representations, recognition and discrimination will be promoted (Valentine & Bruce, 1986a, 1986b). Sensitivity to unfamiliar faces is poor (Kemp, McManus, & Pigott, 1990; O'Donnell & Bruce, 2001) and face images that are fairly typical will be difficult to discriminate; therefore, techniques that can increase distinctiveness and enhance discrimination are potentially powerful. Caricature can be employed to emphasise distinguishing features and can enhance recognition of familiar faces (Benson & Perrett, 1994; Brennan, 1985; Rhodes, Brennan, & Carey, 1987; Rhodes & Tremewan, 1994; Tanaka & Simon, 1996): the research in Chapter 5 describes a computerised caricature technique that can exaggerate face images relative to an average face, and demonstrates that distinctiveness can be systematically increased to significantly improve unfamiliar face discrimination. Distinctiveness will not, however, enhance discrimination in all cases: other race faces may be perceived to be distinctive and yet discrimination is known to be particularly difficult.

FIC and own race bias

The perceptual expertise theory of ORB suggests that limited exposure to other race faces means the appropriate range of facial variation to distinguish them is never learned (MacLin & Malpass, 2001) and developmental evidence of perceptual tuning around only the race of faces that are encountered (Kelly et al., 2007, 2009, 2007, 2005), suggests that discrimination ability is available at birth but is subsequently lost. Both positions accord with Valentine’s (1991) MDFS model in that the ability to discriminate other race faces is a function of the face memory structure developed around the faces that are encountered. Within a MDFS faces of another race faces that are encountered infrequently will be distinctive on at least one race specifying dimension; faces of this race will therefore be encoded together on this dimension (e.g. African American) but because they are unusual, or distinctive, they will also be positioned some distance from the central tendency area of maximum sensitivity and the type of face images that are more commonly encountered. In this way, although the images are distinctive in comparison with own race faces, their distance from the area of facial sensitivity together with their proximity to each other in the face space, means that they will be particularly hard to discriminate (Rhodes & McLean, 1990).

MacLin and Malpass (2001) have provided evidence that perceptual expertise may not entirely account for other race processing deficits: ambiguous faces were generated to portray features that could reasonably occur in both Hispanic and African American faces; the authors found that these were processed differently depending on whether the hairstyle was typical to the same race as the viewer, or of the other race, indicating that racial categorisation will determine what face processing strategies will be employed. As there is also evidence that other race faces are processed in a less holistic manner (Michel, Caldara, & Rossion, 2006; Michel et al., 2006; Rhodes et al., 1989), it is questionable whether sensitivity to facial variation is implicated at all, or whether a bottom up featural process is employed to compensate for perceived discrimination difficulty. The research presented in Chapter 6 explores perception of own race and other race faces; the MDFS model is used to conceptualise a method of causing the shape of African

American and Japanese faces to vary in the same way as own race Caucasian faces, thus enabling the study of both configural sensitivity and categorical processes.

FIC and facial expression

Matching of unfamiliar faces relies on correspondence between two encoded face images; while this could be accomplished in a featural manner in the same way as object perception (Megreya & Burton, 2006), this would not explain holistic or composite face effects (Calder et al., 2000; Calder & Young, 2005; Hole, 1994; Rossion, 2009), nor would it explain how unfamiliar face matching can be achieved at all when face images have substantial differences in terms of lighting, viewpoint, pigmentation etc. (Henderson et al., 2001). The alternative explanation is that face processing and knowledge is employed, and that stored structural representations will form the basis of all facial interpretation and discrimination. Where face matching has to overcome image or structural disparity, such as where images are taken from a different angle, or show different expressions, judgements are known to take longer and it becomes more difficult to form a correspondence between images of unfamiliar people (Bruce, 1982). This would suggest that the encoded images must be held in working memory until the cognitive system can signal a match, or that correspondence is likely. Differences in facial expression will impair correspondence between unfamiliar images (Bruce, 1982) but as the task is difficult and requires consideration, it is also feasible that the affective quality of the faces will be important, even when the expressions match.

Cognitive strategies and additional sources of information are sometimes recruited when a task is difficult, and the smiling face bias is known to enhance perception and judgements of face familiarity, even when faces are in fact not known (Baudouin et al., 2000; Davies & Milne, 1982; Gallegos & Tranel, 2005; Kottoor, 1989). The Bruce and Young model (1986) *(see figure 1, p9)* does not permit affective information to pass from the cognitive system to the face recognition units (FRUs), or even back to structural encoding; but as identification is difficult, it is feasible that a positive affective response may induce the images to be inspected more carefully (Endo et al., 1992), or may make it more likely that one image will be selected over another. The research in Chapter 4 explores the influence of positive expression on facial image comparison judgements.

Thesis structure

The research described in Chapters 2 and 3 draws on the differences in familiar and unfamiliar face perception to devise methods designed to enhance the identification of facial composite images. Chapter 2 explores the possibility that expert holistic face processing of inaccurate facial composite configurations impairs familiar face recognition of accurate component features. The effect of configural inaccuracy is assessed, and facial composites are split below the eyes and misaligned (cf. Young et al., 1987) to determine whether precluding holistic perception of facial composite images will enhance recognition. Difficult identification decisions will also be influenced by affective cognitive bias such that smiling expressions can enhance perceptions of familiarity, while negative expressions will produce the opposite effect. In Chapter 3 facial imaging techniques are employed to compute a smiling expression transformation which is used to explore the influence of affect on facial composite identification.

The second part of this thesis explores facial image comparison (FIC) of two or more face images. Discrimination of unfamiliar faces is difficult and may implicate an affective identification bias; the smiling expression transformation is employed in Chapter 4 to examine the effect of emotional expression on unfamiliar face matching judgements. Unfamiliar face matching is explored within the multi-dimensional face space metaphor in Chapter 5; to increase facial distinctiveness and enhance discrimination, a systematic caricature transformation is employed to exaggerate the facial differences between an average face image, and the target faces and array images. Chapter 6 investigates the perceptual expertise theory of own race bias within the MDFS construct; this work describes an image manipulation that transforms the shape of African American and Japanese faces toward the Caucasian facial variation of the participants. If perceptual expertise accounts for poor discrimination of other race faces the transformation will enhance perceptual discrimination but if racial categorisation determines processing style performance should not be improved. The thesis concludes with Chapter 7 which summarises the findings of the experimental work with reference to theories of face perception and memory, and discusses the implications for forensic application and recommendations for future study.

2

Holistic face processing and facial composite recognition

The research presented in this chapter examines perception and identification of facial composites and explores the possibility that expert holistic face processing impairs identification because of inaccurate facial composite configurations. The relative importance of featural and configural information for identification of facial composites has not been established; therefore, the effect of configural information was investigated in a series of experiments. Holistic analysis is dominant in typical face processing and will influence perception of all of the facial information. While this may be effective for general face recognition, facial composites are recreated from memory and if the configural arrangement and proportions are inaccurate, such holistic interpretation may impair identification of any accurate facial elements. By adopting the technique of Young, Hellawell, and Hay (1987), this chapter will explore whether presentation of facial composites in a manner that impairs holistic face processing can increase the levels of successful identification.

Featural and Configural Information

Facial composite systems enable a witness to create a likeness of a perpetrator from memory, but while the systems can produce composites that are independently judged to be a good resemblance, they consistently fail to produce good rates of identification (Brace et al., 2006; Bruce et al., 2002; Davies et al., 2000; Frowd et al., 2004). This suggests that important and accurate information within the composite images is not easily recognised.

At the most basic level faces provide items of featural information such as eyes, nose and mouth, but the spatial arrangement and configuration of the features is also processed, and the distinct nature of featural and configural analysis is well documented in studies of the face inversion effect. When faces are regarded in a normal upright orientation both featural and configural information is apparent (Sergent, 1984; Yin, 1969); when faces are inverted, holistic analysis becomes ineffective, making the configural properties difficult to perceive and interpretation of the face image must proceed in a feature by feature manner (Bartlett & Searcy, 1993; Farah, Wilson, Drain, & Tanaka, 1998; Freire, Lee, & Symons, 2000; Leder & Bruce, 1998; Rossion et al., 1999; Rossion & Boremanse, 2008).

In typical face perception both featural and configural information will be necessary for optimum recognition performance. If featural processing is impaired by blurring the image, or configural processes are disrupted by scrambling or inverting the image, recognition will be poorer; if both processes are disrupted, recognition will be near to chance levels (Collishaw & Hole, 2000; Schwaninger, Lobmaier, Wallraven, & Collishaw, 2009). Evidence that the processes are dissociable is provided by neuropsychological reports of a patient who could identify upright intact faces, but not inverted faces, scrambled faces, or face parts

(Moscovitch, Winocur, & Behrmann, 1997); while prosopagnosic patients who cannot recognize upright familiar faces, can sometimes recognize inverted faces (Farah, 1994; Farah, Wilson, Drain, & Tanaka, 1995). In the former case configural representations would appear to be accessible while the components or features were not, while in the latter condition inversion would impair holistic and faulty or impaired configural analysis, and facilitate extraction of the featural information.

Holistic and configural face perception

It is important to note that configural analysis is not synonymous with holistic face processing. Holistic face processing is interpretation of featural and configural information as a single representation (Farah et al., 1995), while configural analysis concerns perception of relative sizes and distances between facial features (Maurer et al., 2002). Some configural relationships are also more important than others; for example, altering the spacing of internal configurations, (e.g. distance between the eyes) will disrupt recognition more than changing the spacing of external features (Hosie et al., 1988). Sensitivity to the configural arrangement of faces (Haig, 1984), which are processed holistically, means that changes to an internal configuration will alter perception of the whole image and the face will appear different.

Rossion (2008) suggests that holistic processing should be regarded as a perceptual process originating with the observer, while configural properties describe spatial information that is interpreted most effectively within the holistic process. The distinction is exemplified by Young et al. (1987): composite photographs were created by aligning the top and bottom halves of photographs of different famous faces; participants could easily name the parts, but identification became difficult when they were aligned with another image. Discrimination improved when the images were inverted, or the parts were misaligned; therefore, the authors concluded that holistic processing of the aligned face halves produced the perception of a novel face, which impaired recognition of the familiar face parts. As the spatial relationships were preserved within each face half, both featural and configural processing of each face half was impaired by holistic analysis of closely aligned images.

Aim of study

Facial composites are created from memory and some of the information that is shown will be inaccurate; given the findings of Tanaka and Sengco (1997), Hosie et al. (1988), and Young et al. (1987), flawed configural representations may be processed holistically as novel face images, which will impair perception and recognition of any accurate composite features. The experiments described in this chapter examined the extent to which inaccurate configural information may impair facial composite identification, and explored techniques designed to inhibit this effect. It was predicted that more accurate facial composite configurations would be identified more easily, while preventing holistic analysis of inaccurate facial composite configurations would facilitate recognition of the original composite images

Experiment 1 was completed as pilot work and examined the importance of configural accuracy for composite recognition: a morphing technique was employed to make the composite configurations more accurate in order to assess how much faulty configurations might impair identification. Experiment 2 was partially completed as pilot work and employed the splitting technique of Young et al. (1987) to evaluate identification of the original facial composites and the configurally enhanced facial composites when holistic processing was impaired. It was predicted that this technique would enhance identification of accurate featural information within the original composite images but would hinder identification of the configurally enhanced images. Experiment 3 evaluates holistic analysis of the facial composite images and the effects of configural inaccuracy in a within participant design.

The nature of the composite face effect within facial composite recognition was explored in the remaining experiments. While excluding holistic analysis of faulty configurations might facilitate identification of accurate composite features, it was also possible that the unusual presentation of the split and misaligned images might enhance identification by engaging prolonged structural analysis. Experiment 4 studied reaction times in a cued matching task for complete and split photographs and facial composites.

O'Donnell and Bruce (2001) reported that upper features are most important for identification of familiar faces, while Young et al. (1987) noted faster reaction times for upper face portions. Experiment 5 examined the type of information that is useful for composite identification. If non-holistic analysis of the facial composites particularly enhances perception of the upper features, which in turn causes an identification improvement, perception of just the upper part of the facial composites will be more effective than presentation of complete composite images. If this is not the case, identification of half of the composite information is likely to be worse than identification of the complete composite. Experiment 6 presents all of the composite information but explores whether any disruption to the overall configuration can enhance identification of the accurate composite elements. Facial composites were presented in a split format that separated the upper and lower portions but did not misalign the face parts. Within this series of experiments it was possible to manipulate and assess reliance on, and interference from, holistic processing and configural information in facial composite recognition.

Facial Composite and Photographic Stimuli

Each experiment employed a set of 32 facial composites of male celebrities. Target identities comprised film stars, television personalities, politicians and sportsmen, and were selected to be identifiable to a wide range of participants. Composites and photographs were drawn from University of Stirling archives: each composite was produced for a previous study using one of three composite systems; E-Fit, PROfit and EvoFIT. Witness participants viewed an unfamiliar target and following a standard cognitive interview worked with an experienced operator to construct the composite likeness. Examples of the original facial composites and target photographs are shown in figure 2.1.

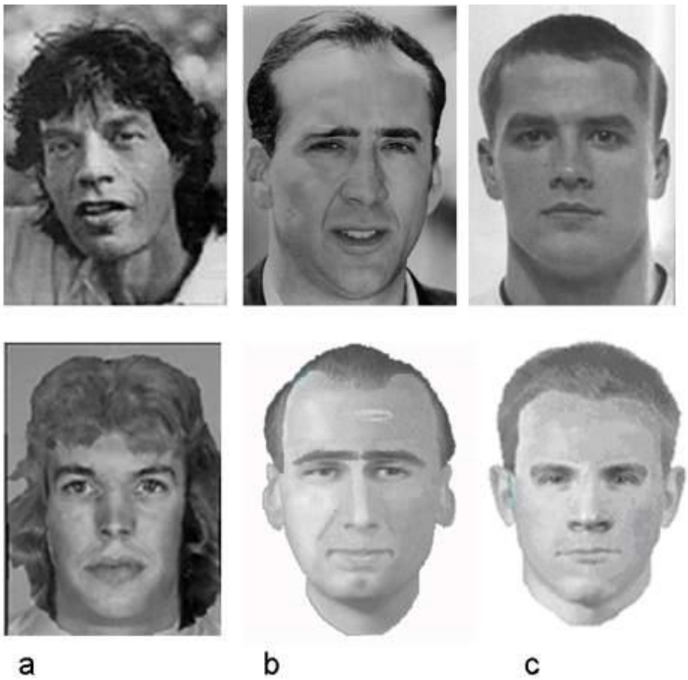


Figure 2.1. Examples of celebrity targets and corresponding original composite stimuli. Top row: (a) Mick Jagger (musician); (b) Nicholas Cage (actor); (c) Michael Owen (footballer). Bottom row: (a) EvoFIT; (b) ProFIT; (c) E-Fit. Facial composites were produced by unfamiliar witness participants.

Experiment 1

Configurally enhanced facial composites

Experiment 1 investigated the importance of configural accuracy for composite recognition. The Psychomorph software (Tiddeman, Burt, & Perrett, 2001) was used to generate a set of 32 configurally enhanced facial composites. Templates were created for each target photograph and original composite by tagging featural and configural landmarks with 179 corresponding data points. To manipulate the composite to more accurately represent the spatial and configural information of the target, the composite image was then morphed to accommodate the photographic template. In essence, the configurally enhanced composites were intended to portray the configural information of the photographs as accurately as possible, within the bounds of the featural information and characteristics selected by the witnesses. It should be noted however, that modifying the configural arrangement of the features might also have improved their shape. This issue will be addressed in the discussion. An example of original and enhanced composite stimuli is shown in figure 2.2.

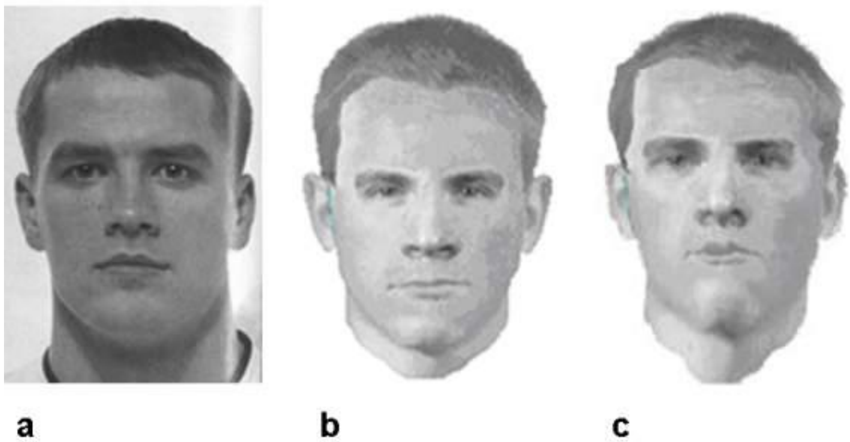


Figure 2.2. Examples of original and configurally enhanced facial composites: (a) The footballer Michael Owen; (b) the original E-Fit facial composite of Michael Owen; (c) the configurally enhanced E-Fit composite. Configurally enhanced composite images were generated by tagging 179 corresponding data points on the target photograph and original composite image; the composite was then warped to portray the configural arrangement of the target face.

Participants

Thirty-two participants (16 male) were recruited from Stirling University campus by opportunity sampling. All had normal or corrected to normal vision, and ages ranged from 18 to 47 years (mean 23.7, *s.d.* 7.1). Participants did not receive payment, although one was awarded course credit.

Materials

Each famous male target was arbitrarily assigned a number (1-32), and the corresponding original composites were randomly divided into two booklets; these were supplemented by the configurally enhanced composites, such that each booklet contained 16 original composites and 16 enhanced composites, with each target represented once. A third booklet was compiled containing the target photographs. The composite images were cropped closely around the head and measured 8 cm in height and subtended a viewing angle of 4.6º at a distance of 1 metre, the target images were cropped to show the head and shoulders and measured 9 cm in height with a viewing angle of 5.2º at 1 metre. All of the images were presented individually in the centre of white A4 paper in landscape orientation.

Procedure

Participants were tested individually and were informed that the images they would be shown were facial composites, such as they would have seen on the "Crimewatch UK" television programme. They were informed that the composites were intended to portray famous males, and were then presented each composite in turn and asked if they could identify the person. In cases where an image could not be named but the participant could provide unequivocal biographical information, responses were accepted as correct. When composite testing was complete, participants were asked to identify the target photographs in order to discard negative responses to the composites of targets with which they were unfamiliar. The presentation order of the composite stimuli was randomized between participants.

Results

The celebrity targets were well known to the participants with a mean identification rate of 92.3%; the target that was known to fewest participants returned an identification rate of 71.9%. Composite naming rates were calculated as a function of the number of targets that were actually known to each participant; therefore, if a participant knew 24 of the targets and could identify 8 of the facial composites, they would return an identification rate of 33.3%, rather than 25% of the full composite set.

The overall composite identification rate was 24.5%: original composites produced an identification rate of 16.6% (*s.e.* 1.6%), while the configurally enhanced composites returned an identification rate of 32.4% (*s.e.* 2.3%). A paired sample *t*-test *(two tailed)* confirmed the advantage for configurally corrected composites to be significant, $t(31) = 6.45$, $p < .001$, $d = 1.4$. There was no correlation between identification of the original and configurally enhanced composites, r .26, $p = .15$. The size of the effect indicates that recognition is grossly impaired by configural inaccuracy in facial composite images. Identification of the original composites was also significantly poorer than the configurally enhanced composite images across the full set of composite items, $t(31) = 4.1$, $p < .001$, but in this case the correlation was significant, $r = .7$, $p < .001$, providing evidence that recognition of a given composite improved when the configural information was enhanced.

Discussion

The aim of Experiment 1 was to establish the importance of configural accuracy for facial composite identification. It was proposed that facial composites would be more identifiable if the configural information was more accurate, and this was facilitated by using target photographs to enhance the accuracy of the composite configurations. The identification rate of configurally enhanced composites was almost double that of original composites (32.4% v 16.6%), indicating that configural accuracy is very important for recognition. The intention was to present enhanced configurations within the bounds of the original composite features, but the morphing procedure may have improved both the configuration and the composite features. Therefore, the observed identification improvement may have been a function of better quality composite features, rather than more accurate configurations enabling perception of the existing featural information.

Experiment 2 explored the potential of a technique employed by Young et al. (1987) to impair holistic analysis of the facial composites and reduce the influence of inaccurate configurations. It was predicted that perception of accurate composite features would be enhanced by splitting and misaligning the top and bottom halves of the composites. With this procedure it was also possible to investigate the nature of enhanced identification observed in Experiment 1: if recognition was facilitated by holistic analysis of the improved configurations, splitting and misaligning the configurally enhanced images should impair perception of the accurate configural information, and identification rates will be comparable with the split original composite images; however, if the morphing technique also improved the features, the identification rate would be better for the split enhanced composite images than for the split original composite images.

Experiment 2

Disrupting holistic processing in facial composite perception

A set of 64 split facial composites was created from the 32 original composites in the manner devised by Young et al. (1987). Adobe Photoshop 5.0 was used to split the composite images horizontally at a point just below the eyes, and each original composite was used to generate two split composite images; one with the lower portion moved left, aligning the nose with the left ear, and one similarly aligned to the right. In essence, split composites portrayed the featural and spatial information selected by the witness, without the potentially novel full face configuration. An equivalent set of split composite images was generated from the configurally enhanced facial composites employed in Experiment 1. Examples of the original and split composite stimuli are shown in figure 2.3.

a b c d

Figure 2.3. Examples of an original and split facial composites: (a) The former Prime Minister, Tony Blair; (b) the original E-Fit composite of Tony Blair; (c) the split composite of Tony Blair; (d) the split configurally enhanced composite of Tony Blair. Split facial composites were generated with Adobe Photoshop 7.0. The direction of facial misalignment was counterbalanced across participants.

Participants

Sixty-four participants (32 male) were recruited from Stirling University by opportunity sampling; all had normal or corrected to normal vision, and ages ranged from 18 to 48 years (mean 26.8, *s.d.* 8.6). None received payment.

Design and procedure

A mixed factor 2 x 2 repeated measures design was employed. The within participant factor was presentation of original facial composites, and split facial composites. The between participant factor was the source of the split composites, which were either generated from the original facial composites or from the configurally enhanced composites. It should be noted that this design is not fully crossed between the split image factor and the configural enhancement factor, as both types of split composites are paired with original facial composites. While a fully crossed design would be more elegant, comparison of configurally enhanced facial composites and split configurally enhanced facial composites would exclude comparison of the effects of each composite alteration with original composites images. The design employed enables direct comparison of the effects of splitting an idealised configural arrangement, and the effects of splitting the original inaccurate configural arrangement, relative to the unaltered original composite images. Between participant samples size would also allow post hoc comparison of effect sizes for different composite alterations across Experiments 1 and 2. Half of the participants were assigned to the original split composite group and half to the configurally enhanced split composite group and testing was conducted in an identical manner to experiment 1.

Materials

Allocation of composites to booklets in Experiment 1 was replicated with two sets of booklets. Split composites generated from original composites completed the first set; the second set contained split composites generated from configurally enhanced composites. Each booklet comprised 16 original composites and 16 split composites, and each target was represented once. Eight of the split composites had the lower portion shifted left, eight were shifted right; the opposite alignment was used for half of the participants. Target photographs were again used to control for familiarity. The split composite images measured 8 cm in height, subtending a viewing angle of 4.6º at a distance of 1 metre; target photographs were 9 cm in height with a viewing angle of 5.2º. All images were presented individually in the centre of white A4 paper (landscape orientation).

Results

The celebrity targets were well known to the participants, with a mean identification rate of 86.8% (*s.d.* 8.7%), the target known to fewest participants obtained an identification rate of 67.2%. Composite naming rates were calculated as a function of the number of targets that were known to each participant, and explicit biographical information could be accepted in place of a name. Mean identification rates are shown in figure 2.4. The split configurally enhanced composites had a mean identification rate of 24.6% (*s.e.* 2%) in comparison with 15.4 % (*s.e.* 2.1%) for original composites. Identification of split original composites was 24.2% (*s.e.* 2%) in comparison with 17.9% (*s.e.* 2.1%) for original composites.

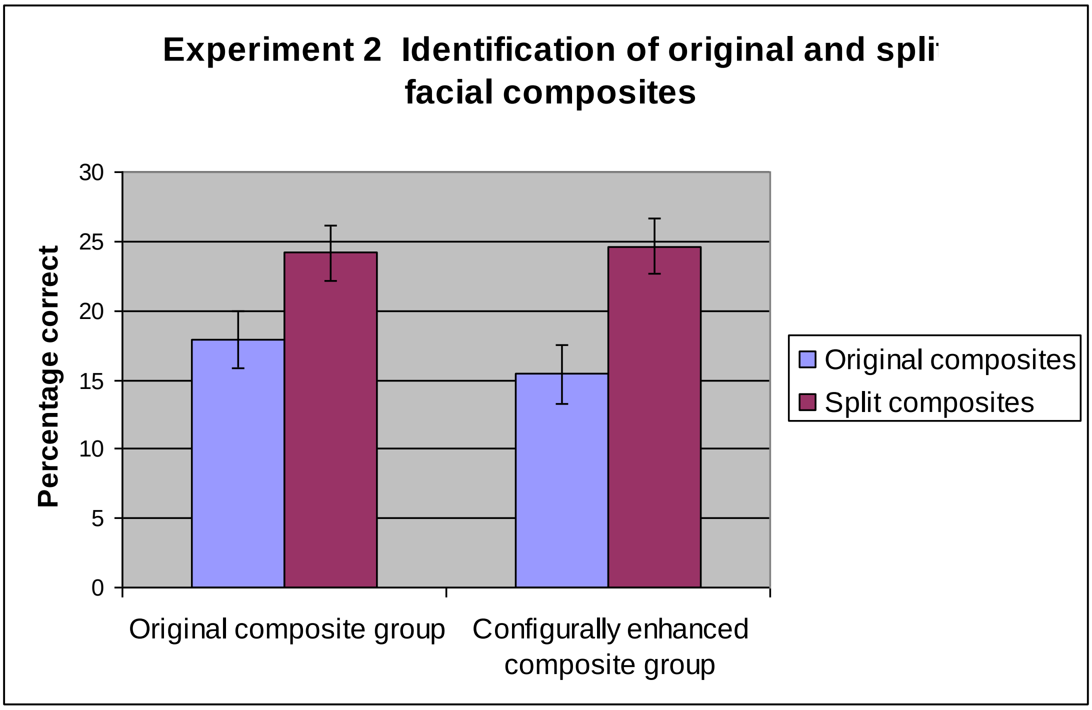


Figure 2.4. Mean composite identification rates and standard errors by group, and by composite type. Identification of the original composite images is comparable across groups. Split composite images were identified better than the original composite images, and performance for the configurally enhanced split images and the original split images do not differ.

A mixed factor repeated measures analysis of variance found a significant main effect of splitting composite images, $F(1,62) = 20.6$, $p < .001$, $\eta_p^2 = .25$, but no main effect of configural enhancement ($p = .7$) and no significant interaction ($p = .4$). Presenting facial composites in a split format significantly increased composite recognition relative to the original composite images and accounted for 25% of the variance in the scores. Within the split image format, configurally enhanced facial composites were not identified any better than the original composite information.

By composite items, a mixed factor repeated measures analysis of variance confirmed a significant main effect of splitting the composite images, $F(1,62) = 10.7$, $p = .002$, $\eta_p^2 = .15$, but no significant main effect of configural enhancement (p

= .8) and no significant interaction (*p* = .4). Presenting facial composites in a split format significantly increased the likelihood that a composite image would be identified and accounted for 15% of the variance in identification rates. Performance for split configurally enhanced composites did not differ from performance with split original composites.

Discussion

Having established in Experiment 1 that inaccurate configural information impairs identification of facial composites, the aim of Experiment 2 was to explore the potential of a split presentation format to reduce the effects of configural inaccuracy. It was predicted that splitting and misaligning the top and bottom halves of facial composites would prevent holistic analysis and enable perception of the facial composite features. In this way, facial composites would be more identifiable when composite information was available without holistic perception of inaccurate full-face configurations. Identification of split composites was significantly better than of original composites, indicating that holistic processing of composite information interferes with identification of the composite elements.

Within this paradigm it was also possible to investigate the nature of the enhanced identification observed in Experiment 1: if recognition was facilitated by holistic analysis of improved configurations, splitting and misaligning configurally enhanced composites and original composites would produce comparable effects; but if morphing also improved composite features, split configurally enhanced composites would be identified better than split original composites. Performance was comparable when the images were split and misaligned, providing a strong indication that featural representations were not significantly enhanced by the morphing procedure, and that split and misaligned presentation successfully impaired holistic analysis of the configural composite information.

Post Hoc comparison of composite alterations

Identification rates from Experiments 1 and 2 were also contrasted with one-way analyses of variance for identification of original composites, and identification of the modified composite images, with participant groups as the factor. Identification of original composites averaged 16.6% (*s.e.* 1.1%), and did not differ between participant groups, p = .7. However, there was a significant difference between groups for identification of the altered composite images, $F(2,95) = 4.9$, p = .01. Post hoc Tukey HSD tests confirmed that identification of configurally enhanced composites was significantly better than identification of split original composites, p = .02, and split configurally enhanced composites, p = .03. recognition of split composites and split configurally enhanced composites was not significantly different, p = 1. The mean identification rates of original and altered facial composite images are shown in figure 2.5.

The results supported the predictions; identification was best when enhanced configural information could be processed holistically, while the poorest identification was observed with holistic processing of the original inaccurate configural information. For the split composites, identification of the original composite information improved, while recognition of enhanced configurations declined, indicating that the manipulation successfully reduced the impact of both flawed and enhanced configural information.

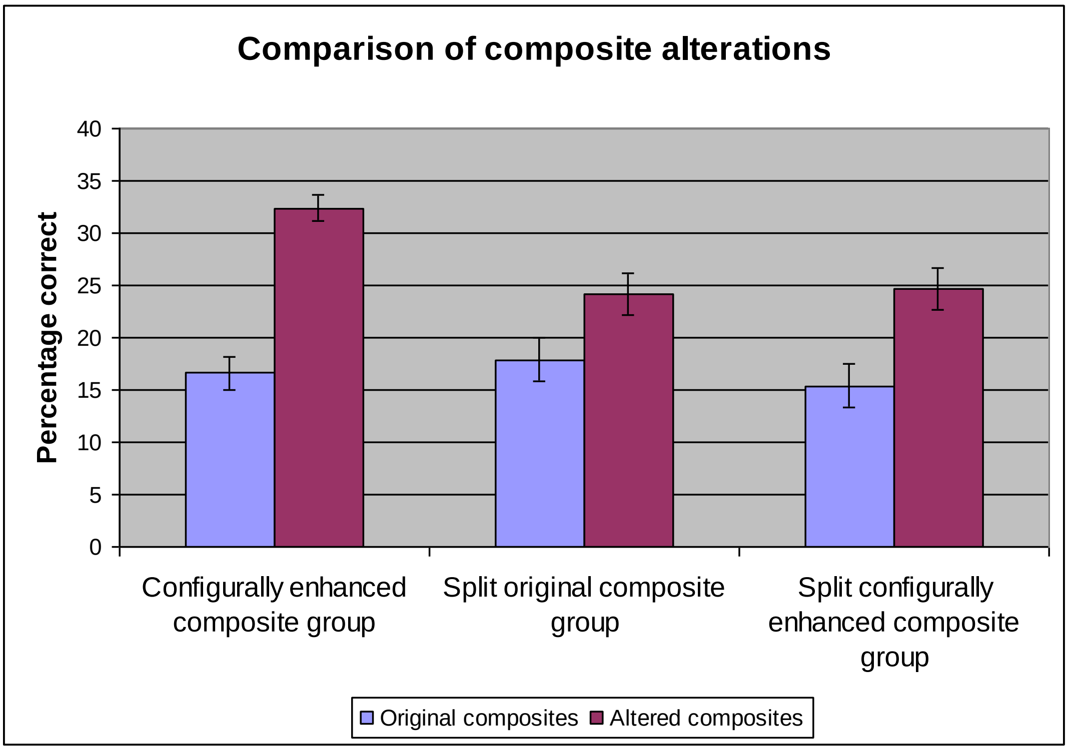


Figure 2.5. Mean composite identification rates and standard errors for original composites, configurally enhanced composites, split original composites, and split configurally enhanced composites. Identification was best when configurally enhanced facial information could be processed holistically. Identification of split original composite images and split configurally enhanced composite images did not differ, indicating that the images were processed in a similar way and that the technique employed to enhance the configural arrangement did not confer any featural advantage. Identification of the original composite images was comparable across groups.

The results of Experiments 1 and 2 provide compelling evidence that facial composites will automatically be processed in a holistic fashion, but that holistic processing of inaccurate facial configurations will impair the ability to extract and recognise the accurate featural elements. The data considered in Experiments 1 and 2 were provided by three separate groups of participants; Experiment 3 therefore replicates these findings with one sample of participants in a within participant experimental design.

Experiment 3

Participants

Seventy-two participants (18 male) were recruited from Stirling University by opportunity sampling; ages ranged from 18 to 50 years (mean 21.4, *s.d.* 5.0). All participants had normal or corrected to normal vision, and none received payment.

Design

A 2 x 2 repeated measures within participant design was employed with factors of image type (original composites and configurally enhanced composites) and presentation format (complete composites and split composites). The design enabled direct comparison of identification rates for the original and configurally enhanced facial composites, together with the effects of splitting original inaccurate configural arrangements and accurate enhanced configural arrangements.

Materials

Four sets of materials were created; each contained 8 original composites, 8 configurally enhanced composites, 8 split composites, and 8 split configurally enhanced composites. Each target was represented once in each test set, and represented once in each format across the sets of materials. E-Prime experimental software was employed with a 17 inch LCD monitor at 1024 x 786 pixels resolution to counterbalance the test sets, and present the images in fully randomised order. Images measured 8 cm in height and were presented in the centre of a white screen at a distance of approximately 70 cm, subtending a viewing angle of 6.5º.

Procedure

Participants were tested individually; it was ensured that each person understood what a facial composite was, and they were informed that they would be shown composite images that were intended to represent famous men. The composites were presented individually in randomised order on the computer monitor, and participants were asked to attempt to identify each one. A correct name or if explicit biographical information was recorded a correct response. Unfortunately, due to coding error identification rates for the target photographs were not obtained and all statistical analyses are based on raw identification rates. Raw identification scores do not control for actual familiarity with the targets, therefore identification rates will be lower than would be observed if negative responses to unfamiliar composite targets were discarded. In this case, if a participant could identify 8 of the composite images, they would obtain an overall accuracy rate of 25% (8/32) irrespective of how many targets were actually known.

Results

As expected, the complete configurally enhanced composites produced the best rates of identification (25.8%, *s.e.* 2.1%), and the poorest results for the original facial composites (17.4%, *s.e.* 1.9%). Identification of split original composites (21.3%, *s.e.* 1.8%) and split configurally enhanced composites (21.4%, *s.e.* 2.1%) was comparable: impairing holistic analysis of accurate configural information is detrimental, but if the configural information is inaccurate, will enable identification of accurate featural information. The mean composite identification rates and standard errors are shown in figure 2.6.

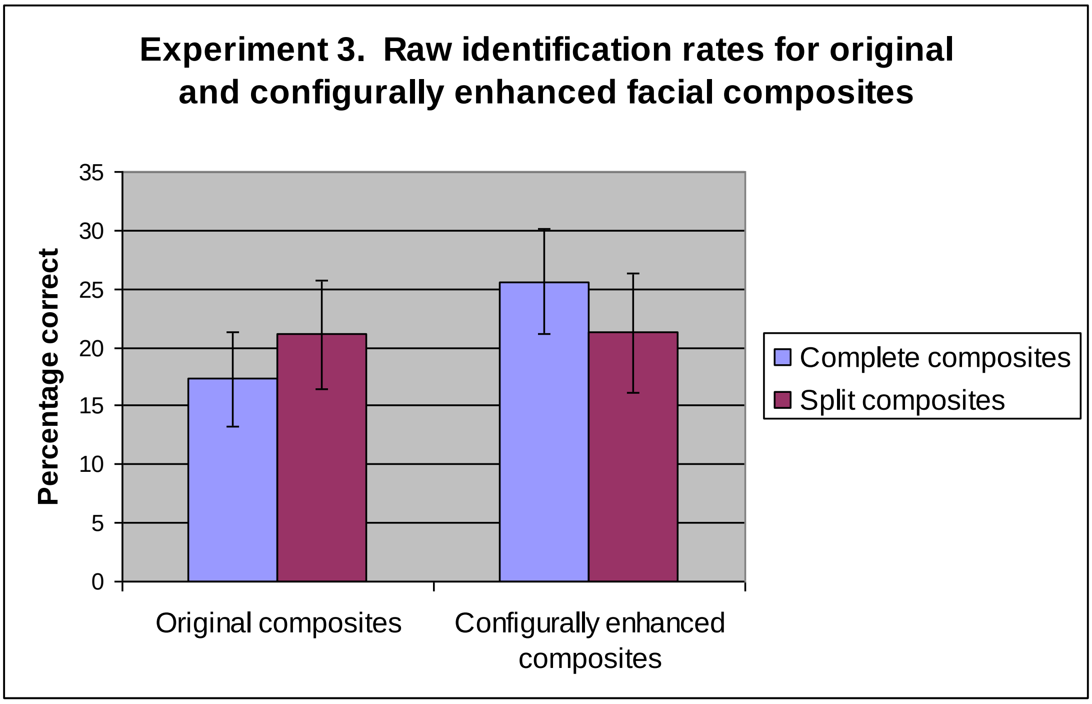


Figure 2.6. Mean composite identification rates and standard errors by composite type and by presentation format. Identification of the original composites was poorer than identification of configurally enhanced images. Splitting the facial composites facilitated identification of the original composites but impaired recognition of the configurally enhanced items.

A repeated measures analysis of variance with composite type and presentation format as factors, confirmed a significant main effect of configural enhancement, $F(1,71) = 4.8$, $p = .03$, $\eta_p^2 =.06$, but no significant main effect of presentation format ($p = .9$), the interaction between configural enhancement and presentation format was marginally non-significant, $F(1,71) = 3.6$, $p = .06$, $\eta_p^2 = .05$. Paired sample *t*-tests showed that the configurally enhanced facial composites were identified significantly better than the original composites, $t(71) = 3.1$, $p = .003$, $d = 0.5$. Identification of the split original composites and the split configurally enhanced composites was comparable ($p = 1$), but splitting failed to produce a significant effect on identification of either the original ($p = .1$) or the configurally enhanced composites ($p = .1$).

A repeated measures analysis of variance by composite items found no significant main effect of presentation format, *p* = .9, and a marginally non-significant effect of configural enhancement, *F*(1,31) = 3.3, *p* = .08, η_p^2 =.10, although the interaction between configural enhancement and presentation format did reach significance, *F*(1,31) = 4.7, *p* = .037, η_p^2 =.13. The configurally enhanced composite images were identified significantly better than the original composite images, *t*(31) = 3.4, *p* = .002, *d* = .0.3, and identification of the split original composites and split configurally enhanced composites was comparable, *p* = 1. While there was a trend for split original composites to be identified better than the original composites, *t*(31) = 1.8, *p* = . 085, *d* = 0.2, the size of this effect is small, and identification of configurally enhanced facial composites and split configurally enhanced composites was not significantly different, *p* = .1.

Discussion

Experiments 1 and 2 used three participant samples to show that facial composites are processed in a holistic fashion, but that inaccurate facial configurations will impair the ability to extract and recognise featural elements from within original composite images. Precluding holistic analysis by splitting and misaligning composite faces was shown to enable recognition of the original composites. Experiment 3 aimed to replicate these findings in one sample of participants but was only partially successful. As predicted, configurally enhanced composites were identified best, and original composites were recognised least well. When the images were split and misaligned identification of the configurally enhanced composites and the original composites was comparable, meaning that recognition of the enhanced configural information was ineffective in the absence of holistic processing.

Although the pattern of results is in line with the predictions, and across items the effect approached significance, a methodological error makes strong conclusions impossible. Facial composite identification relies on activation of a stored memory of a familiar person, as such, identification rates must be qualified by actual familiarity: failure to record target identification in this experiment means the raw identification rates lack the sensitivity or the power to be any more than a guide. However, the split composite effect has since been replicated by Frowd *(personal communication)*: his study employed a different set of facial composites and found a large effect of splitting the original facial composites on identification rates; original facial composite identification was 18.8% (*s.e.* 3.2%), while split composite identification was 34.7% (*s.e*.4.8%), $t(23) = 2.1$, $p = .04$, $d = 0.8$.

As the split composite advantage has now been replicated within participants, in different sets of composite images, and by different investigators, we can conclude that the effect is reliable. Identification of better quality configural information is impaired when configurally enhanced facial images are split, while identification of the original composite information improves. It is proposed that this effect is caused because holistic analysis of good configural information is easier and more effective, but where configural information is inaccurate, holistic interpretation will be misleading and identification of accurate composite elements will be difficult. If this premise is correct it may also be possible to observe the effects in reaction time data, such that for accurate configural information holistic interpretation would be quicker, but for inaccurate configural information an accurate identification response should be achieved faster when holistic analysis is impaired. Alternatively, it is possible that holistic analysis will always be faster and that the split images might have enhanced identification by engaging prolonged structural analysis, in which case split composites will always produce longer reaction times than complete composite images. Experiment 4 employs reaction time as a metric to explore the holistic and non-holistic interpretation of inaccurate and accurate facial configurations.

Experiment 4

Reaction times for complete and split facial images

When Young et al. (1987) aligned the top and bottom halves of different famous faces and asked participants to identify the parts, they found response times were significantly shorter when the face parts were misaligned (1082 ms v 1289.5 ms). They concluded that the new configuration created by aligning different face parts made identification of the components more difficult. One would therefore expect identification of correct configurations to be faster with complete faces than misaligned faces, while identification of inaccurate composite configurations would be faster when the images are misaligned. Experiment 4 employed a cued matching task to explore reaction times for correct and inaccurate configural information in photographs and facial composites; when holistic analysis was possible from intact images, and when it was precluded by splitting the facial images.

Participants

Sixty-three participants (36 male) were recruited from Stirling University by opportunity sampling; ages ranged from 17 to 61 years (mean 21.9, *s.d.* 7.7). All participants had normal or corrected to normal vision, and none received payment.

Design

A 2 x 2 x 2 mixed factor repeated measures design was employed with a between participant factor of image type (composites and photographs), and within participant factors of presentation format (complete face images and split face images), and cue (accurate name cue and inaccurate name cue). The design enabled direct comparison of holistic and non-holistic face processing of photographs and facial composites of the same target celebrities; within the photographs the configural properties would be veridical and thus accurate, within facial composites the facial configurations and some of the features would be incorrect.

The dependent variables were the proportion of accurate responses that the images matched, or did not match the preceding name cue, and the reaction times for accurate responses. By splitting the images, holistic analysis would be impaired and identification would be achieved by recognition of features: it was predicted that for the inaccurate facial composite configurations splitting the image would be beneficial, leading to faster reaction times, but that for the photographs, inability to utilise accurate configural information would impair performance. Identification of target photographs was also expected to be much faster than identification of facial composites and would therefore be analysed separately.

Materials

Eight sets of materials were created, four comprising 32 facial composites; four comprised photographs of the target celebrities. Each set contained 16 complete items, plus 16 split and misaligned items. Half of these images were preceded by the correct name cue, and half preceded by the name of an unrelated celebrity. All image media, image presentation format, and cue conditions were counterbalanced across the test sets, such that each target was represented once in each test set and once in each format. The images and name cues were presented using E-Prime with a 17 inch LCD monitor (1024 x 786 pixels resolution). Name cues were shown in black Times New Roman 18 point font in the centre of a white screen; the face images were presented in greyscale, were cropped around the head and measured 8 cm in height. They were shown in the centre of a white screen, and subtending a viewing angle of 6.5º at a distance of 70cm.

Procedure

Participants were randomly assigned to the facial composite or photograph condition and those in the facial composite condition were briefed regarding the purpose and nature of facial composite images. The participants were informed that they would be shown a series of celebrity names followed by photographs, or facial composites, of famous men; their task was to indicate whether the name cue matched the face image, or whether the face images were of (photograph), or were intended to represent (facial composite), different people.

The test sets were counter-balanced by participant, and the 32 trials were presented in randomised order. If the participant believed that the name cue and the face image matched they were to press m on the keyboard, if they did not believe the name cue and face image matched they were to press z. The keys were reversed for half of the participants and a guide to the appropriate key press was always visible. Each trial began with the cue screen showing a celebrity name for 2000 ms, followed by the facial composite or photograph, which remained on screen for 5000 ms or until a response was given. If the celebrity was unfamiliar, no response was provided and the next trial would begin following the 5000 ms duration. The trial procedure is illustrated in figure 2.7.

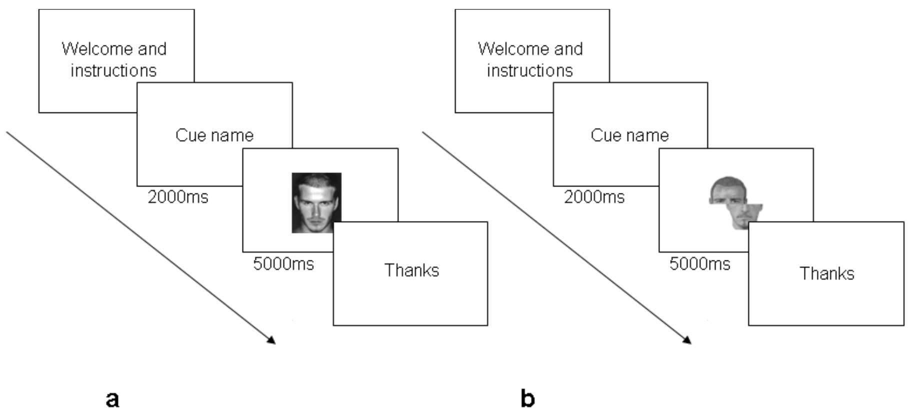


Figure 2.7. Examples of the cued naming reaction time procedure employed in experiment 4. The experiment commenced with a welcome and instruction screen, trials were activated manually by key press when the participant was ready to begin; for each of 32 trials a cue name was shown for 2000ms, followed by (a) a photograph of the target, or (b) a facial composite intended to portray the target, for a duration of 5000ms, or until a response is given. The faces were presented as (a) complete images, or as (b) split and misaligned images.

Results

Accuracy rates and average median reaction times for correct responses are shown with inverse efficiency scores in table 2.1. Accuracy for the photographs is comparable across conditions, although response times indicate that for inaccurate cues, participants took longer to respond to the split images. For the composites accuracy levels indicate that it was easier to reject inaccurate name cues than to accept a composite image as a good likeness for an accurate cue. Repeated measures analyses of variance with factors of presentation format (complete image / split image) and name cue (correct name / incorrect name) were conducted separately for photographs and facial composites.

Target photographs	Complete Image		Split image	
	Accurate cues	Inaccurate cues	Accurate cues	Inaccurate cues
Reaction time (ms)	737 *(179)*	825 *(169)*	730 *(135)*	931 *(175)*
Accuracy (%)	89.7 *(11.6)*	89.5 *(12.4)*	90.4 *(12.6)*	89.1 *(10.4)*
Inverse efficiency score (Reaction time / proportion correct)	0.84 (0.3)	0.93 (0.2)	0.82 (0.2)	1.06 (0.3)
Facial composites	**Complete Image**		**Split image**	
	Accurate cues	Inaccurate cues	Accurate cues	Inaccurate cues
Reaction time (ms)	1673 *(524)*	1790 *(634)*	1851 *(615)*	1822 *(611)*
Accuracy (%)	57.8 *(24.2)*	79.1 *(17.1)*	52.4 *(21.1)*	69.8 *(19.3)*
Inverse efficiency score (Reaction time / proportion correct)	3.49 (2.7)	2.37 (1.1)	4.00 (2.1)	2.98 (1.7)

Table 2.1. Accuracy rates and average median reaction times for correct responses are provided for target photographs and the facial composite images (standard deviations are shown in parenthesis). Inverse efficiency scores provide a global measure of performance: these are calculated by dividing reaction time (in seconds) by the proportion of correct responses; lower scores are indicative of better performance.

Accuracy Data

Analysis of correct responses for photographs showed that there were no significant main effects of presentation format, p = .9, or name cue, p = .7, and no significant interaction between these factors, p = .8. Accuracy for the photographs was unaffected by the veracity of the name cues or split image presentation.

For correct responses to the facial composites, the main effect of presentation format was not significant, $F(1,31) = 3.2$, $p = .08$, $\eta_p^2 = .10$, although it did account for 10% of the variance in accurate scores. There was however, a significant and very large main effect of name cue, $F(1,31) = 26.3$, $p < .001$, $\eta_p^2 =.46$, which accounted for 46% of the variance. There was no significant interaction between presentation format and name cue, $p = .6$. Complete composites tended to be better matched than the split composites, but accuracy was much better for inconsistent name cues than for correct name cues, indicating that it was easier to conclude that a composite image did not represent the person named, than to consider the composite image to be a good likeness.

Reaction Time Data

Repeated measures analysis of median response times for correct responses to the photographs showed that there were significant main effects of presentation format, $F(1,30) = 6.0$, $p = .02$, $\eta_p^2 =.17$, and name cue, $F(1,30) = 64.0$, $p < .001$, $\eta_p^2 =.68$, which were qualified by a significant interaction, $F(1,30) = 9.6$, $p=.004$, $\eta_p^2 =.24$. Paired sample *t*-tests found that responses were faster for accurate name cues than for inaccurate name cues for the complete images, $t(30) = 3.3$, $p = .003$, $d = 0.5$, although the effect was the same but was considerably larger for the split and misaligned photographs, $t(30) = 8.2$, $p < .001$, $d = 1.3$. When the name cues were accurate, responses to split and complete photographs did not differ $p =.8$, but when the name cues were inaccurate, responses were faster for complete images, $t(30) = 3.7$, $p = .001$, $d = 0.6$; therefore, it took longer for participants to establish that a photograph did not match the name when holistic analysis was not possible.

Analysis of median reaction times for correct responses to facial composites revealed a significant main effect of presentation format, $F(1,31) = 5.3$, $p =.03$, η_p^2

= .15, but no significant main effect of name cue, p = .6, and no significant interaction, p = .3. Responses were considerably slower for the split facial composite images.

Inverse Efficiency Scores

For the photographic images there were no effects of condition on accuracy, but reaction time effects were observed, and for the facial composites there was an effect of name cue on accuracy but not response times; therefore, accuracy and reaction time data were combined to provide inverse efficiency scores (response time in seconds divided by the proportion correct) as global measures of performance for each of the experimental conditions (Townsend & Ashby, 1983).

Repeated measures analysis of inversion efficiency scores for photographic images found no significant main effect of presentation format, p = .1, but a significant main effect of name cue, $F(1,30) = 23.3$, $p < .001$, $\eta_p^2 = .44$, and a significant interaction between presentation format and name cue, $F(1,30) = 6.3$, p = .02, $\eta_p^2 = .17$. Performance was better in response to accurate name cues, than to inaccurate name cues for complete photographs t(30) = 1.9, p = .07, d = 0.4, and this effect was substantially larger for split images, $t(30) = 5.3$, $p < .001$, d = 0.9. When the name cues were accurate, performance for split and complete photographs did not differ, p = .6; but when the name cues were inconsistent, performance was better for the complete images, $t(30) = 2.8$, p = .01, d = 0.5. The participants were better able to determine that the name cues were wrong when they were shown complete face images that enabled holistic analysis.

Repeated measures analysis of variance of inverse efficiency scores for the facial composites, revealed that while a main effect of presentation format accounted for 9% of the variance in scores, it failed to reach statistical significance, $F(1,31) = 3.1$, $p = .09$, $\eta_p^2 = .09$. There was a significant main effect of name cue, $F(1,31) = 35.5$, $p = .01$, $\eta_p^2 = .19$, with no significant interaction between these factors, $p = .9$. When accuracy and reaction time were combined, performance in the cued matching task tended to be better for complete composite images, but was significantly poorer for accurate name cues, suggesting that a tendency to reject the composite images as a likeness accounted for 19% of the variance in the inverse efficiency scores.

Discussion

Photographs present accurate facial information and it was predicted that reaction times would be faster for holistic interpretation of complete photographs than for non-holistic interpretation of the split photographs. However, a reaction time decrement for split images was only found when they were inconsistent with the preceding name cue, meaning that even without holistic analysis a name cue could be matched to a consistent facial representation very quickly. Where the inappropriate cue was successfully rejected, participants took longer to respond to the split images, indicating that without the ability to employ holistic analysis it took longer to scan a non-matching image to ensure that the facial information and the name did not correspond (see Schwaninger et al., 2009).

Facial composites portray inaccurate configurations meaning that holistic interpretation of the component information should be more difficult and that split facial composites which reduce configural perception, would be expected to produce faster responses. However, the opposite effect was found and matching of name cues to facial composites was faster with the complete composite images. Complete composite images also tended to produce more accurate responses but there was an indication that participants may have been inclined to reject the composite images more often than they accepted them.

The prediction that reaction times would be faster for split composite images was not supported; although split images enhance recognition of accurate composite elements, the effect may be mediated by prolonged structural analysis. However, splitting the configurally enhanced composites in Experiment 2 impaired recognition; therefore, increased time to process split images cannot compensate for disruption of accurate configural information. The evidence suggests that it is the nature, and not just the amount of processing that is altered when the composites are split and misaligned. The split facial composite advantage may thus be a function of slower, non-holistic interpretation of facial composites.

It is proposed that splitting facial composites disrupts holistic face processing and enables perception of the composite features, but it is not known whether all of the composite features are important for recognition. Face perception studies have noted that facial features have differing salience (e.g. Haig, 1986; Hosie et al., 1988), with the eyes and upper features being particularly important for recognition (O'Donnell & Bruce, 2001). Young et al. (1987) observed that for the split and misaligned images, identification of the upper face half was significantly faster than identification of the bottom half. Experiment 5 evaluates the effectiveness of the upper facial features for identification of facial composite images.

Experiment 5

The importance of upper features

Experiments 1 to 3 established that facial composite recognition can be facilitated if the images are split and the upper and lower portions are misaligned. O'Donnell and Bruce (2001) reported that upper features predominate in recognition of familiar faces; therefore, the aim of Experiment 5 was to determine if this is also the case in identification of facial composite images. If non-holistic analysis of composite images facilitates recognition, and if this effect is driven primarily by the upper portion of the facial composites, we would expect non-holistic interpretation of the upper part of the facial composites to be more effective than presentation of complete composite images. If however, all of the composite information is required to achieve recognition, identification of the upper portions is unlikely to be significantly better.

Participants

Twenty-four participants (12 male) were recruited from Stirling University by opportunity sampling; ages ranged from 17 to 44 years (mean of 20.4 , *s.d.* 5.4). All participants had normal or corrected to normal vision, and none received payment.

Materials

The 32 original composites were used to generate upper composite materials: Adobe Photoshop 5.0 was used to split the composite images horizontally at a point just below the eyes, and to remove the lower portion. With original composites materials these were randomly divided into two booklets, such that each booklet contained 16 original composites and 16 upper composites, with each target represented once. The booklet containing the target photographs was employed to ensure familiarity with the celebrities. Composites measured 8 cm in height and were shown in the centre of white A4 sheets in landscape orientation. An example of the stimuli is shown in figure 2.8.

Figure 2.8. Examples of (a) an original facial composite and (b) an 'upper' facial composite, of the musician Noel Gallagher. The features in the upper portion of the face have been shown to be most important for familiar face recognition. Experiment 5 explored whether the split composite advantage is mediated by non holistic recognition of the upper features, or whether all of the composite information should be presented in a non holistic format for best levels of identification.

Procedure

Participants were tested individually, and were informed that the images they would be shown were facial composites like the ones used by the police and shown on television programmes like "Crimewatch UK". They were informed that the composites were intended to portray famous men and that as they were made from memory, some aspects would be more accurate than others. They were then presented each composite in turn and asked if they could identify the person. As before, explicit biographical information was accepted as a correct response. To provide baseline levels of familiarity, participants were also asked to identify the target photographs. Presentation of the composite stimuli was randomized for each participant.

Results

Familiarity with the target celebrities was good, 88.7% (*s.e.* 2.2%). Complete facial composites generated an identification rate of 14% (*s.e.* 2.5%), while the images showing only the upper portion of the composites produced an identification rate of 10.4% (*s.e.* 1.3%). A paired sample *t*-test *(two-tailed)* found no significant difference in these identification rates, $p = .2$, and no significant correlation, $p = 1$. Across the participant sample identification of complete composites and images that showed just half of the information did not differ. However, across the set of composite items the difference did approached significance, $t(31) = 2.0$, $p = .058$, $d = 0.2$, and identification of complete and upper composites was significantly correlated, $r = .8$, $p < .001$. Whilst the effect size was small, the complete composites that were identified were also more likely to be identified when only the upper portion was shown.

Discussion

The facial composites produced an identification rate of 14% which across participants was not significantly better than identification of the upper part of the facial composites presented in isolation (10.4%). Whilst this may seem like the lower portion of a facial composite has little additional value, identification of the split composite images in Experiments 1 to 3 was substantially higher than original images, strongly suggesting that this is not the case. It is possible that viewing the upper half of composites was partially effective for the same reason that the split composites are effective, the absence of a full face image enabled identification of the composite elements. However, across the set of composite items identification of each image in both formats was directly contrasted, and the difference between full composite and upper composite identification just failed to reach significance. This suggest that the lower part of facial composites provide important identifying information. Disruption of the holistic analysis of composite images can facilitate recognition, but all of the composite information is required to achieve the best levels of performance. The final experiment of this series presents all of the facial composite information, but evaluates whether it is necessary to misalign the images or whether merely breaking the composite image into parts is sufficient to impair holistic analysis and produce a composite identification advantage. In Experiment 6 composite images were split and separated, but not misaligned.

Experiment 6

Presentation of non-holistic facial composite images

The previous experiments have established that facial composite recognition is enhanced if the image presentation precludes holistic interpretation, and that the effect may be mediated by prolonged component analysis. While identification could be achieved from the upper portion of composite images, all of the composite must be presented in order to optimise identification. A final point has yet to be considered, if all of the composite information must be presented, and if the full face image must be disrupted, is any form of holistic disruption suitable, and can the images be presented without the face portions being misaligned? If a complete face image is necessary to promote holistic facial analysis, splitting the image may be sufficient to enable identification of the facial composite elements. Experiment 6 presented composite images that were horizontally split and separated, but the composite parts were not misaligned. Example stimuli can be viewed in figure 2.9.

Participants

Twenty-four participants (8 male) were recruited from Stirling University by opportunity sampling; ages ranged from 17 to 45 years (mean of 23.3, *s.d.* 8.2). All participants had normal or corrected to normal vision, and none received payment.

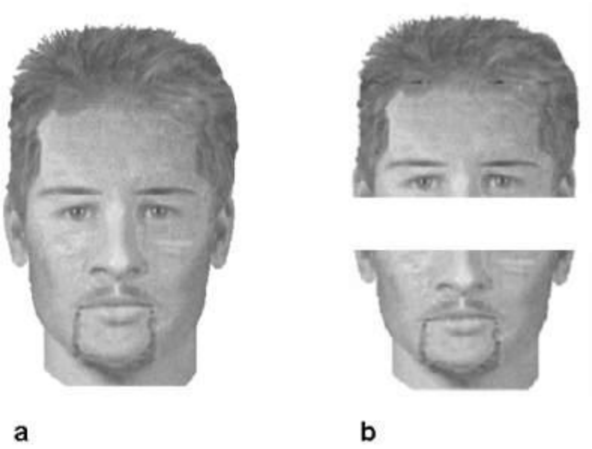


Figure 2.9. Examples of (a) an original facial composite and (b) a split facial composite, of the actor Brad Pitt. Presentation of all of the composite information in a manner that precludes holistic analysis has been shown to significantly enhance facial composite identification. Experiment 6 explores whether disrupting the composite configuration by splitting the image is still effective if the images are not misaligned.

Materials and design

Adobe Photoshop 5.0 was used to generate a set of split facial composites. Each composite was split horizontally below the eyes and the upper and lower portions were separate by a gap of 1 cm. Composites were allocated to two booklets; each contained 16 original composites and 16 split composites, with each target represented once. Baseline familiarity was assessed using the booklet of target photographs. Facial composites measured 8 cm in height, and vertically split composites 9 cm in height. The procedure was identical to Experiment 5 and images were presented in randomised order in the centre of white A4 sheets in landscape orientation.

Results

The mean target familiarity was 87% (*s.e.* 2.1%). The original facial composites generated an identification rate of 17.5% (*s.e.* 2.1%), while split composites produced an identification rate of 19.5% (*s.e.* 3.1%). A paired sample *t*-test *(two-tailed)* found these were not significantly different, $p = 0.6$, and there was no correlation, $p = 0.7$. There was also no significant effect across composite items, $p = .2$, although the correlation between identification of the original composites and the vertically split facial composites was significant, $r = .7$, $p < .001$. The original composites that were identified well were also identified better in the split format.

Discussion

Facial composite recognition is enhanced when the images are presented in a manner that precludes holistic face processing, and all of the composite information must be shown in order to optimise identification. Experiment 6 explored a split presentation format that presented all of the original information but did not misalign the images. The facial composites were split below the eyes and separated, and the composite parts were presented one above the other with a gap between them. Overall identification was fairly good (18.3%) but there was no advantage for splitting the composite images in this way, indicating that when the parts are not misaligned holistic perception is not disrupted. To facilitate identification of composite elements it is necessary to both split and misalign the facial composite information.

While sensitivity to configural information in faces will not reach adult levels until the second decade of life (Freire & Lee, 2001; Mondloch, Dobson, Parsons, & Maurer, 2004; Mondloch, Le Grand, & Maurer, 2002; *although see* Crookes & McKone, 2009 *for a review and discussion of general development theory*), holistic face processing is observed in very young children (3 – 6 years) (Brace et al., 2001; Carey & Diamond, 1986; de Heering, Houthuys, & Rossion, 2007; Mondloch, Pathman, Maurer, Le Grand, & de Schonen, 2007; Sangrigoli & Schonen, 2004). This suggests that holistic face processing is so highly developed it will dominate face perception and will compensate for image inconsistencies; in this case, quite literally filling in the gaps. Hole, George, Eaves, and Rasek (2002) have shown that a configural arrangement will be preserved under extreme distortion, such as stretching the whole face image to twice its height. As maintenance and perception of a configuration is remarkably robust, the ability to perceive and identify a facial composite will be dependent upon the quality of the configural arrangement, and in order to reduce the effects of inaccurate configurations, it will be necessary to present the composite images in a manner that completely precludes holistic interpretation. While splitting and misaligning the upper and lower composite face parts was effective, simply splitting the image in two was not.

Summary and conclusions

When comparison of a facial composite and a target photograph suggests a good likeness, identification rates are nonetheless frequently low. Familiar face perception is sensitive to configural variation (Haig, 1984) and if the configuration of a face image is altered, recognition of the facial features can be impaired (Tanaka & Sengco, 1997; Young et al., 1987); as such, it is possible that configural inaccuracy in facial composites could interfere with identification of accurate composite features. The aim of this chapter was to examine the extent to which inaccurate configural information might impair facial composite identification and to explore techniques designed to inhibit this effect. It was predicted that composites with accurate configurations would be identified better than original composite images, while preventing holistic interpretation of inaccurate configurations would enable recognition of original composite information.

In Experiment 1 facial composites that were enhanced to portray more accurate configurations were identified significantly better than the original composites, confirming that facial composites do not portray accurate configurations and that configural accuracy is important for composite recognition. The second experiment investigated whether the effects of inaccurate composite configurations could be reduced and identification of accurate facial composite information enhanced. The technique devised by Young et al. (1987) was used to disrupt holistic processing of the composites without altering any of the actual composite information. The split and misaligned composite presentation was highly effective for identification of original facial composites but impaired recognition of configurally enhanced composites. This method therefore, successfully impaired perception of configural information, and was effective in reducing the influence of configural inaccuracy in facial composite images.

It is important to note that it is not proposed that the 'split composite effect' exclusively promotes identification of features. When a face image is disrupted by splitting and misalignment, relational and featural information is preserved within each of the face parts, i.e. the distance between the eyes, as well as the eyes. Isolated spatial relations are explicitly represented in memory, and are important for face recognition (Leder & Bruce, 2000); therefore, enhanced composite identification may have been a function of improved featural recognition and improved perception of isolated spatial relationships.

Expert face processing involves the capacity to process faces efficiently as complete units of information. In Experiment 4 it was predicted that in comparison with non-holistic processing of split and misaligned images, reaction times would be faster for holistic processing of accurate configural information (photographs) but slower for holistic processing of inaccurate configural information (facial composites). Reaction times for the facial composites showed the opposite effect, suggesting that the split composite advantage observed in Experiments 1 – 3 may have been caused by analysis of the facial composites in a slower non-holistic manner. It should be noted however, that slower non-holistic analysis only improved recognition when the configural information was inaccurate. This suggests that the logic behind the reaction time evaluation was flawed: Schwaninger et al. (2009) reported significantly shorter reaction times for holistic analysis of faces, than for featural analysis of the same facial information. If one begins with the premise that holistic face processing is a reflexive parallel process, and that to facilitate perception of parts it is necessary to induce a serial featural process, there is no logical reason why split images should ever be processed more quickly than complete face images, even those that are inaccurate.

Schwaninger and colleagues also found that it took longer to reject distracters than to recognise targets, concluding that identification of a match requires detection of a single familiar feature, but rejection of a distracter requires each feature to be scanned to ensure that none are familiar. A similar effect was observed in matching cue names with photographs; without the ability to employ holistic analysis it took longer to scan a non-matching image to ensure that the facial information and the name did not correspond.

Some features are more important for face recognition than others, and studies have shown that the upper features are more useful than the lower features (Hosie et al., 1988; O'Donnell & Bruce, 2001). Identification of full facial composites and the upper portions of facial composites did not significantly differ; this could be interpreted to show that the upper face portions are responsible for recognition of facial composite images. However, in experiments 2 and 3 identification of split and misaligned facial composites produced much higher identification rates suggesting an alternative explanation: splitting face images or showing part of a face image enables the viewer to perceive component facial information in the absence of holistic processes. While this might enable perception of upper facial features to produce comparable results to recognition of complete facial composites, the lower portion also contains identifiable information that will be recognised if the composite is shown without the presence of the full face configuration.

Within this series of experiments the facial composites were more identifiable when the composite information was presented in a manner that did not allow holistic perception of inaccurate composite configurations. The final experiment demonstrated that expert holistic face processing is remarkably robust and will compensate for distortion of images; consequently, it is necessary to misalign the upper and lower face portions to prevent holistic reconstruction and enable interpretation of the facial components. The necessary conditions for a split facial composite advantage are presence of an inaccurate facial composite configuration, that is presented in full, in a format that prevents perceptual reconstruction and holistic analysis.

Theoretical considerations

Facial composite identification requires activation of the stored memory of a familiar person but while familiar face perception can accommodate substantial image variation (Bruce, 1982; Burton et al., 1999), inaccuracies in facial composites make them hard to match with a stored memory representation. Within a facial composite the configuration is likely to be inaccurate, and identification will require recognition of accurate facial features (Collishaw & Hole, 2000; Schwaninger et al., 2009). However, facial composites will be processed holistically, and inaccurate configurations within the composite face image will make it difficult to extract or recognise the component information (Hosie et al., 1988; Tanaka & Sengco, 1997; Tanaka & Farah, 1993; Young et al., 1987).

Within the concept of the Bruce and Young (1986) model, inaccurate facial composites will be perceived holistically as novel faces that won't correspond to any stored face representation, and accurate information within the facial composite will not be recognised. By adopting the split-image technique (Young et al., 1987), holistic face processing will be impaired, which should enable perception and activation of any matching component information. If sufficient composite elements are accurate, and the target is known, a face recognition unit (FRU) will reach the threshold for activation and composite recognition will be enabled. Recognition of configurally enhanced composite images was significantly improved because the accurate configurations enabled the composites to be matched successfully to stored representations; when the images were split activation of FRU's was impaired by loss of the accurate configural information. In a similar vein, identification of the upper composites could only be mediated by the available information and as such was significantly better when all of the split composite information was available.

Applied considerations

Facial composites are a valuable resource that enables witnesses and victims to communicate the appearance of a perpetrator. As a witness is typically unfamiliar with the perpetrator, the resulting composite will not be completely accurate, and identification will depend on the ability of a familiar person to recognise the parts of the image that portray a true likeness. In keeping with face processing expertise, facial composites will be processed holistically as complete units of information, and as the configuration will be flawed, identification of the features will be impaired. The results of this study provide evidence that facial composites can be presented in a manner that reduces the effects of configural inaccuracy and will enhance recognition. The practical implication is that police forces should consider releasing split composites to the public to boost the likelihood of identification. Splitting facial composites requires a simple image editor and is therefore free, yet it has the potential to make a significant impact on the successful detection of crime.

It is possible that a split composite technique might also be useful in the composite construction process. Facial composites require witnesses to select features, or judge facial likenesses within the context of whole faces (e.g. Davies & Christie, 1982; Tanaka & Sengco, 1997); it would be useful to determine if holistic face processing during construction is beneficial, or whether the facial features should be selected in isolation before generating the overall composite likeness. Police artists have been found to produce better facial likenesses than featural composite systems (Frowd, Carson, Ness, Quiston-Surrett, et al., 2005), and it is the police artists' protocol to have the witness identify featural components from a manual, before generating a sketch that combines drawings of these items. Notably, the witness does not view the composite sketch until this process has been completed (Gibson, 2008).

Conclusions

Identification of facial composites can be very poor (Brace et al., 2006; Bruce et al., 2002; Davies et al., 2000; Frowd et al., 2004; Pike et al., 2005) because limitations of unfamiliar memory and difficulty of communicating facial information mean that to some extent, all composite likenesses will be inaccurate or incomplete. Familiar face perception is sensitive to configural properties (Haig, 1984) and if configurations are inaccurate, it can impair identification of facial features (Tanaka & Sengco, 1997). This research demonstrates that holistic interpretation of flawed composite images makes it difficult to identify accurate composite features, but if the images are shown without a full face configuration, holistic analysis will be precluded and recognition can be significantly enhanced. The identification enhancement was mediated by slower, non-holistic interpretation of all of the composite information.

By enabling featural perception and activation of any matching representations in memory composite recognition can be facilitated, but as the match between the facial composite image and the memory representation will be less than optimal, identification judgements will also be susceptible to decisional bias. The smiling face recognition bias is only apparent when identification is difficult, consequently the affective quality of composite images may be very important. Chapter 3 explores the issue of expression in facial composite recognition and confirms that affective quality does exert powerful effects that can be reduced with imaging techniques.

3

Emotional bias in facial composite construction and recognition

The research described in this chapter explores the influence of affective information and facial expression on identification of facial composites. Successful facial composite identification depends upon correspondence between the information that an unfamiliar witness can provide, and the information that is needed to achieve familiar face recognition. As facial composites are produced from memory, they will be unavoidably flawed and difficult to match with any stored face representation: where identification is not easy, recognition judgments are influenced by affective information, such that positive expressions enhance perceived familiarity, while negative expressions produce the opposite effect. The influence of emotional affect on facial composite identification was investigated in a series of experiments to explore whether positive affect could improve facial composite recognition, or whether incorrect identifications might also increase. A barely perceptible level of manipulation was employed to establish whether a positive recognition bias could be incorporated in forensic application.

Unfamiliar witnesses and contextual memory

Exposure to a crime is in most cases, experienced and remembered as an extremely unpleasant event. To convey the appearance of a perpetrator, an unfamiliar witness must create a facial composite from a memory that is structurally limited and dependent on the context at the time of encoding. This means that, in addition to the physical structure of the face, emotional information about the target, the event, and the mood state of the witness will all be encoded and associated within the face memory trace (Eich, 1995; Eich, Macaulay, & Ryan, 1994; Kenealy, 1997; Smith & Vela, 2001; Tulving, 1982, 1984). In line with the encoding specificity principle (Thomson & Tulving, 1970; Tulving & Thomson, 1971, 1973), remembering will also be most effective when the witnesses' mood states at retrieval match their moods at the time of memory encoding; this is termed mood state dependent recall (see Ucros, 1989 for a review).

Context is recognised as important for retrieving information from memory and is accommodated within a context reinstatement phase of the cognitive interview; during this process the witness is guided through a process of recollection and imagery in order to recreate their mood and physical state at the time of the event. When context reinstatement is effective it will enhance recall for the facts or faces of an event; but negative emotions associated with these memories will also be recreated and reinforced. If the witness then goes on to produce a facial composite their memory of the face and the resulting likeness will reflect these negative qualities, and indeed real witnesses are concerned that the facial composites they produce do reflect the negative qualities and expressions they associate with the individual and the event *(Ann Parry Metropolitan Police & Janet Richardson (retired), Forensic Artists, personal communication).*

The smiling face bias and facial composite recognition

Face perception studies concerned with the influence of emotional expression have consistently found that smiling faces are judged to be more familiar than neutral faces (Baudouin et al., 2000; Davies & Milne, 1982; Endo et al., 1992; Gallegos & Tranel, 2005; Kottoor, 1989), while negative expressions produce the opposite effect (Lander & Metcalfe, 2007). For genuinely familiar faces, this is shown in familiarity ratings, and in speeded recognition responses; but where faces are unknown, expression will influence decisions about whether or not a face has been encountered before. The conditions entailed by the smiling face bias are particularly relevant for the production and identification of facial composites.

Facial composites are typically produced by unfamiliar witnesses following an unpleasant event and as far as possible they will portray the remembered unpleasant qualities. Essentially this means that a facial composite image is both unavoidably inaccurate, and intentionally negative. In order to achieve a positive identification a person who is familiar with the perpetrator must be able to recognise the accurate facial elements from within the composite image, which as we have shown is somewhat difficult. Because facial composites are difficult to identify and because the smiling face bias is evident where identification is not easy, the affective quality of the composite images will be very important: familiarity is associated with positive affect but the composite images may be generated to show negative affect; therefore, even if a composite image elicits a sense of familiarity, the negative qualities may influence judgement away from forming any recognition response. In this way, there is an important mismatch between the information provided by the unfamiliar witness, and the information that is necessary to provide a familiar identification.

Aim of study

This series of experiments explored whether the smiling face bias, or indeed the absence of any positive expression, would influence identification of facial composite images, and examined the extent to which affective information in facial composites could be manipulated to enhance identification. Studies of smiling face bias in face recognition have employed images that showed individuals in neutral or smiling poses; however, the applied nature of facial composites mean that it would be inappropriate to portray suspects of serious crime overtly smiling and for this reason, a level of expression that was detectable but did not produce explicit smiles was employed in Experiments 7 and 8. Experiment 7 explored the efficacy of a subtle positive expression transformation and examined the importance of positive affect for recognition of famous facial composites. As there is some evidence that affective qualities may not influence identification of family members (Bruce, 1982; Endo et al., 1992) and therefore faces that are personally familiar, Experiment 8 evaluated the effectiveness of the positive expression transformation on identification of facial composites intended to portray known individuals.

To determine the limits of an effective yet forensically acceptable manipulation, it was important to establish that composite identification could be improved by enhancing expression, and a level of transformation that would produce the best results. Experiment 9 assessed the ability to detect different levels of smile transformation, while Experiment 10 explored the effect of a weaker expression, and of an explicit expression on composite identification. It was predicted that making facial composites to show positive affect would enhance judgements of familiarity; but might also produce higher levels of incorrect identification.

Enhanced expressions in facial composites

To manipulate the facial expressions of composite images, PsychoMorph software (Tiddeman et al., 2001) was used to compute the physiognomy of an average neutral face and the physiognomy of an average smiling face. An average neutral expression was estimated by mapping 179 data points on the facial landmarks of 100 photographs of neutral male faces: these were morphed together and the resulting template used to generate an average neutral expression. An average smiling face was generated in an identical manner using photographs of the same individuals smiling. A PsychoMorph template and the neutral and smiling face averages are shown in figure 3.1.

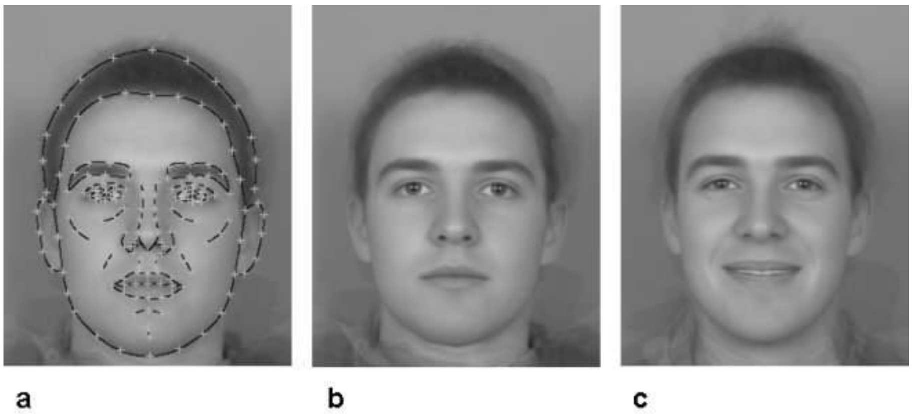


Figure 3.1. Examples of (a) a Psychomorph template (b) the average neutral expression, and (c) the average smiling expression. Average expressions were created by morphing together the PsychoMorph templates of 100 neutral male images, and templates of the same 100 males smiling. The quantified difference between the morphed templates describes how, on average, a face transforms when a person smiles.

From the average expressions it is possible to calculate how, on average, a face will transform when a person smiles. The quantified smile transformation can then be applied to any facial image by creating a PsychoMorph template and using it to manipulate the facial characteristics along the computational difference between the average neutral and smiling expressions. The expression transformation will shift the facial information in precisely the way that, on average, a face will change when it smiles, and this will have the effect of producing a smile expression without altering or averaging any of the actual composite information that was selected by the witness. To avoid portraying suspects of serious crime with explicit smiles, a manipulation of 30% of the smile transformation was deemed to be just detectable and was employed in Experiments 7 and 8. Examples of the composite stimuli are shown in figure 3.2.

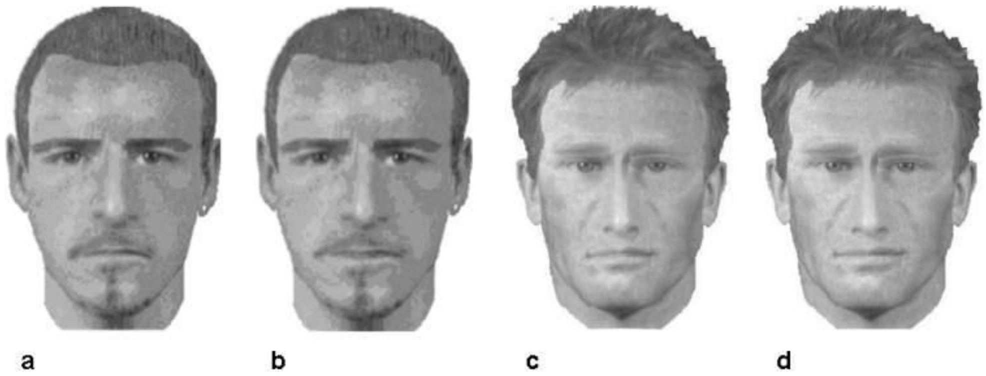


Figure 3.2. Examples of original and transformed facial composites: (a) original E-Fit of David Beckham (footballer), (b) E-Fit of Beckham with 30% of the smile expression transformation applied, (c) original ProFIT of a University of Stirling lecturer, (d) ProFIT image with 30% of the smile transformation applied.

Experiment 7

Enhancing positive affect in famous facial composites

Experiment 7 investigated the effect of positive expression on facial composite recognition. The composites and Psychomorph templates employed in Chapter 2 were used to create 32 expression-enhanced composite images. Each composite was morphed to transform the affective information away from the average neutral expression towards the average smile. The original facial composite information was not altered but was manipulated by the expression transformation to change only in the way that, on average, a face will change when the person smiles *(see figure 3.2).*

Participants

Twenty-eight participants (7 male) aged between 18 and 48 years (mean 22.2 years, s.d. 6.1 years) volunteered in return for course credit. All were students from the University of Stirling and all had normal or corrected to normal vision.

Materials and Design

Original facial composites were randomly assigned to one of two booklets and supplemented by expression-enhanced composites so that each booklet contained 16 original and 16 expression-enhanced composites with each target represented once. The target images were used to establish levels of familiarity. Composites and target images measured 8 cm in height, subtending a viewing angle of 4.6º at a distance of 1 metre, they were shown individually in the centre of white

A4 paper in landscape orientation and presentation was randomised for each participant, the composite identification procedure was identical to Experiment 6.

Results

Target familiarity was 76.9% (*s.e.* 9.2%) and identification rates were calculated for each participant as the number of correct identifications divided by the number of targets known. The original facial composites generated an identification rate of 15.5% (*s.e.* 1.5%), while expression-enhanced composites produced an identification rate of 21.8% (*s.e.* 2.5%). The data were analysed with a 2 x 2 mixed factor repeated measures analysis of variance with composite type (original composites; expression-enhanced composites) as the within participant factor, and test set as the between participant factor. There was a significant main effect of composite type, $F(1,26) = 4.6$, $p = .04$, $\eta_p^2 = .15$, but no significant main effect of test set, $p=.5$, and no significant interaction between image type and test set, $p = .8$. The expression enhanced facial composites were identified significantly better than the original composite images, with the expression transformation accounting for 15% of the variance in accuracy.

Across the set of facial composites, analysis by items with a paired sample *t*-test found no advantage for the expression-enhanced images, $p = .1$, although the correlation between identification of the original composites and the expression-enhanced composites was highly significant, $r = .75$, $p < .001$, indicating that the original composites that were identified well were also identified better when the facial expression was more positive.

Discussion

The results support previous findings that expression influences familiarity and identity decisions (Baudouin et al., 2000; Davies & Milne, 1982; Endo et al., 1992; Gallegos & Tranel, 2005; Kottoor, 1989; Lander & Metcalfe, 2007), and extends them to include facial composite perception. Original facial composites elicited a naming rate of 15.5%, which is in line with previous research (Frowd, Carson, Ness, Quiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005), while the naming rate of the expression-enhanced facial composites was significantly improved (21.8%). Across the sample of participants identification rates were substantially better when the composite expressions were transformed to be more positive, and the facial composites that were most identifiable in the original format remained the most identifiable when the affective information was enhanced. Failure of the effect to reach significance across the full set of composite items could suggest that identification will only be enhanced with certain images or if a facial composite portrays identifiable information; but it could also indicate that the effect is inconsistent and warrants further evaluation with a different set of facial composite images.

Facial composite research often employs celebrity targets to obtain identification data from a wide sample of participants; but as the celebrities are not personally familiar it is possible that their faces may be processed or stored differently, or that they may be more strongly associated with smiling expressions than people known through personal experience. This means that the observed benefit of expression-enhancement in Experiment 7 could reflect a recognition advantage for stored representations of smiling famous faces, rather than an overall smiling face improvement for facial composite images. Experiment 8 addresses this issue by using facial composites of faces that are personally familiar to participants asked to identify them.

Experiment 8

Enhancing positive affect in personally familiar facial composites

The alternative to celebrity targets is to create facial composites of personally familiar people; this is typically achieved by sampling targets and participants from an occupational setting like a university department. While the aim is to explore familiar face recognition, the limited target pool can encourage participants to guess the composite identities by a process of elimination rather than by face recognition (i.e. this person has blonde curly hair so it must be Jane). To evaluate the true influence of affective information on non celebrity facial composite identification, Experiment 8 employed personally familiar targets but included facial composite distracters and the caution that some of the targets were unfamiliar to ensure engagement of genuine face recognition processes.

Familiar targets and distracters

To examine familiar recognition of facial composites a target set was created comprising eight familiar lecturers from the Psychology Department at the University of Stirling (4 female): familiar targets were identifiable to staff and final year students; to ensure that identification would not be confounded by forced choice identification they were matched to eight unfamiliar lecturers from Napier University on the basis of gender, age, weight and hairstyle. No target had distinctive characteristics, and on the basis of superficial features alone, each composite could be mistaken for another staff member.

Target materials

To mimic real life conditions, unfamiliar witness participants must be briefly exposed to a target: each lecturer was photographed in a number of poses displaying a number of facial expressions and all of the images provided contextual background cues. Four images were selected of each target to be used as stimuli: one showed a full face pose, one a ¾ profile, one a full profile and one portrayed the target looking up and away from the camera. Two of the images displayed neutral expressions, one a smiling expression, and one an angry expression. The resulting familiarisation materials therefore were not image-, or expression-specific, and provided a richer representation than would be possible from one full-face image. The photographs were captured at a distance of approximately 3 metres with a Sony Cyber Shot digital camera (5 mega pixels resolution), which was mounted on a tripod. Each image was sized to a width of 10cm and subtended a viewing angle of 7º at a distance of 80 cm; all four were displayed simultaneously in landscape orientation within a word document on a Dell Inspiron 6400 laptop. An example of the target stimuli is shown in figure 3.3.

Witness participants

Thirty two witness participants (6 male) aged between 18 and 46 years (s.d. 6.5 years) were recruited from staff and students at Stirling University. All had normal or corrected to normal vision, were unfamiliar with the targets, and were paid £5 for their participation.

Figure 3.3. Example of target exposure stimuli. Targets were shown in four positions: full-face pose; ¾ profile; full profile; looking up and away from the camera. Two images displayed neutral expressions, one a smiling expression and one an angry expression. The series of images provided contextual background cues and a richer memorial representation than would be possible from one basic face image.

Composite construction procedure

Witness participants attended the lab individually on two occasions. On the first visit they were randomly allocated a target and allowed to view the target stimuli for one minute, the only instruction given was to view the person shown in the photographs; an appointment was then made to return in two days. Facial composite construction at the second appointment was designed to mimic current police practice: each witness participant received a full cognitive interview, including imagery and context (Fisher, Geiselman, Raymond, Jurkevich, et al., 1987; Geiselman et al., 1985). Following the cognitive interview each participant worked with the experimenter to produce a facial composite of the target using ProFIT facial composite software. When the participant was satisfied with the likeness, they were debriefed and paid for their time. Two witness participants created a facial composite for each target, generating a set of 32 composite images. Examples of a target, his composite likeness and a matched distracter composite are shown in figure 3.4.

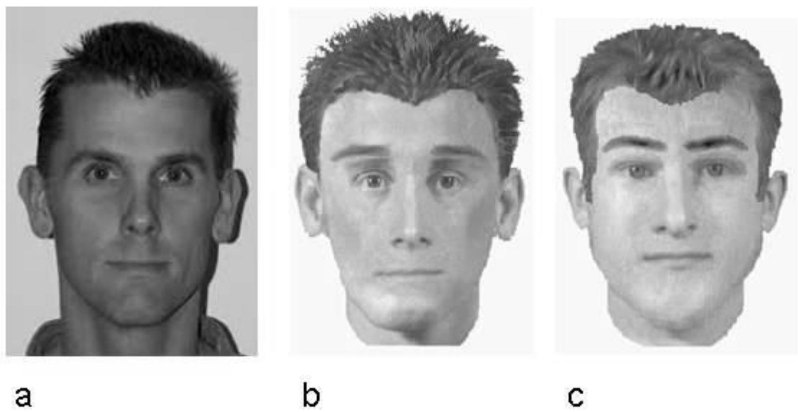


Figure 3.4. (a) A University of Stirling lecturer, (b) the ProFIT facial composite intended to represent the lecturer, (c) the ProFIT facial composite intended to portray the Napier University lecturer matched for age, gender, and physical attributes.

Enhancing positive affect in the facial composites

Experiment 8 investigated the importance of affective information for non celebrity facial composite recognition. As female faces tend to be more expressive (Dimberg & Lundquist, 1990; Fischer, 1993; Hess & Bourgeois, 2010) and may move in a different way than male faces when they smile, average neutral and smiling female images were created using PsychoMorph and the resulting templates formed the references for enhancing the facial expressions of female facial composites. Individual templates were created for each facial composite image and were used to manipulate the composite facial characteristics along the computational difference between the average female or male, neutral and smiling expressions by 30%. Examples of the original and enhanced-expression female composite stimuli are shown in figure 3.5.

Figure 3.5. (a) A ProFIT facial composite intended to represent a lecturer, (c) ProFIT composite with 30% of the female smile transformation applied. The expression transform shifts the composite image in the way that on average a female face changes when the person smiles. None of the component composite information is altered.

Recognition: Materials and Design

The original and expression-enhanced Stirling University composites and Napier University distracter composites were assigned to two test sets of power point slides such that each set contained both composite attempts at each identity: one the original composite and one the expression-enhanced composite. Allocation of participants to the test sets was counterbalanced and the items within each test set was randomised within two half books; each contained one composite attempt at each identity and the presentation order was reversed for half of the participants.

Participants

Forty-seven participants (8 male) were recruited from staff and final year psychology students at Stirling University. They ranged in age from 19 to 63 years (mean 27.2 years, s.d. 10.7 years). All had normal or corrected to normal vision and all were familiar with the Stirling targets, they were paid £2 for their time.

Procedure

Participants were tested individually or in groups of up to eight. They were informed that they would be shown a series of facial composites, that some of them were intended to portray members of staff from the Psychology Department and some portrayed people from other universities. They were informed that they should attempt to identify each one and record their answer on the form provided. Where they could not provide a name, but could offer information that would identify the person, they were asked to clearly note this down. The facial composites were presented sequentially in randomised order and participants responded in their own time.

Results

All of the targets were familiar to the participants. The original facial composites produced an identification rate of 12.5% (*s.e.* 1.5%), which was much lower than the identification rate of the expression-enhanced composite images, 22.3% (*s.e.* 1.7%). The data were analysed with a 2 x 2 mixed factor repeated measures analysis of variance with composite type (original composites; expression-enhanced composites) as the within participant factor, and test set as the between participant factor. There was a significant main effect of composite type, F(1,45) = 39.3, $p < .001$, $\eta_p^2 = .47$, but no significant main effect of test set, $p = .8$, and no significant interaction between image type and test set, $p = .1$. The expression enhanced facial composites were identified significantly more often than the original composite images, accounting for 47% of the variance in the scores.

Across the set of facial composite items the expression-enhanced facial composite images also generated a significantly higher identification rate than the original composites, $t(15) = 2.8$, $p = .01$, and there was a strong correlation between identification of the original and expression-enhanced composites, r .9, $p < .001$. When original composites contained identifiable information, enhancing the affective quality of the image increased recognition.

Whilst it is promising that expression-enhancement can improve identification of facial composites, positive expressions have also been shown to erroneously elicit judgements of familiarity for unknown faces (Baudouin et al., 2000; Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2004) and it would be of some concern, if the familiarity bias for positive affect was also associated with higher rates of incorrect facial composite identification. Comparison of incorrect naming levels revealed that the original composites were wrongly identified 8.6% (*s.e.* 0.9%) of the time in comparison with 6.8% (*s.e.* 0.9%) for expression-enhanced composites, indicating that positive affect did not increase false identification of facial composite images, $p = .1$.

Signal detection measures of sensitivity *d'* and response criterion *c* (Green & Swets, 1974) were also calculated from the hits and false identifications of the original and expression-enhanced composite images. Response criterion was more conservative for the original composites $c = 1.58$, than for expression-enhanced images, $c = 1.34$, $t(46) = 2.2$, $p = .03$, while sensitivity was significantly better for expression enhanced images $d' = 1.02$, then for the original images $d' = 0.11$, $t(46)$, 4.6, $p < .001$. These results suggest that while incorrect identifications did not increase, greater sensitivity to the expression-enhanced composite images is associated with greater willingness to offer up a name.

Discussion

The results of this experiment replicate those found with the celebrity composites in Experiment 7, and extends the finding that facial affect influences identification judgements to identification of facial composite images. The original non-celebrity composites elicited a naming rate of 12.5%, which increased to 22.3% when facial expressions were transformed to be more positive. The proposal that identification of celebrity composites might have improved because participants were more familiar with smiling celebrities than with neutral celebrities, is rejected because the recognition advantage was considerably larger for non-celebrity facial composites (η_p^2 .47) than for famous composites (η_p^2 .15). Affective information was more important for composites of people with whom we have personal interactive memories, rather than people that are frequently shown smiling in the media.

There is no obvious reason why enhancing composite images to show more positive expressions should enable greater levels of sensitivity. One possibility is that enhanced positive affect increased the level or quality of attention that was given to the facial composites. Endo et al. (1992) found that participants took longer to inspect unfamiliar face images when they were smiling, suggesting that positive expression motivated greater engagement before an identity decision was reached. Ellis and Young (1990) propose that in normal familiar face perception, recognition of the structural face stimulus is accompanied by a positive affective response. In this way it is possible that a positive affective response will cue familiarity causing the face image to be studied more carefully in a bid to match it to a stored familiar face representation; where matching of the face to a stored representation is possible, but in doubt, a positive affective input could be sufficient to promote the identification judgement.

Previous work has shown that unfamiliar faces are more likely to be judged familiar if they are smiling, suggesting that positive expression might reduce the criterion for recognition judgements, (e.g. Baudouin et al., 2000; Garcia-Marques et al., 2004). These studies employed familiarisation procedures and subsequently recorded old/new identification judgements; they therefore assessed the influence of affect on memory for face images with no corresponding personal or semantic information. Criterion for a facial composite recognition response was reduced in this experiment but it is notable that levels of false identification did not increase. As sensitivity to the face images also improved, positive affect may have signalled familiarity, leading images to be studied more carefully and increasing the probability of a positive recognition response. While this could also have reinforced activation of incorrect representations, the combination of positive affective signal and increased sensitivity or attention to the composite images did not produce this outcome.

In applied use facial composites are produced from memory following an unpleasant event and while identifiable information may be present it will be difficult to distinguish amidst other inaccurate information. These results indicate that under such conditions affective information and responses will be important, and where a member of the public has some doubt over recognition, the absence of any positive emotional signal will support rejection. Indeed given a lack of positive affect, the facial composite may fail to engage sufficient attention for an identification to be achieved.

Experiment 9

Perception of affective transforms

In designing a study that explored the utility of affective information to enhance composite identification, it was desirable to employ a technique that would be acceptable in forensic application. Facial composites portray perpetrators of crime, and witnesses who construct them would be offended if they were shown with the presence of an explicit smile. The transformation level of 30% was initially selected because the modification was extremely subtle but could be perceived; given the success of the expression-enhanced composites it was important to establish whether this level of manipulation was explicitly perceived, or whether the recognition enhancement might operate at a sub-conscious level. Experiment 9 assessed detection of both an explicit 60% expression-enhancement and the 30% expression transform.

Materials and Design

To generate a fully counterbalanced within participant design, fifteen of the University of Stirling facial composites were used to generate expression-enhanced composites manipulated to portray 60% of the smiling expression transform. Three presentation sets were created to display five pairs of identical original composites, five original composite and 30% expression-enhanced composite pairs, and 5 original composite and 60% expression-enhanced composite pairs. Image pairs were allocated such that each composite was represented once within a test set and the order was randomised for each participant. The images were 5 cm in height and pairs were presented side by side in the centre of an A4 page in landscape orientation. Examples of the composite stimuli are shown in figure 3.6.

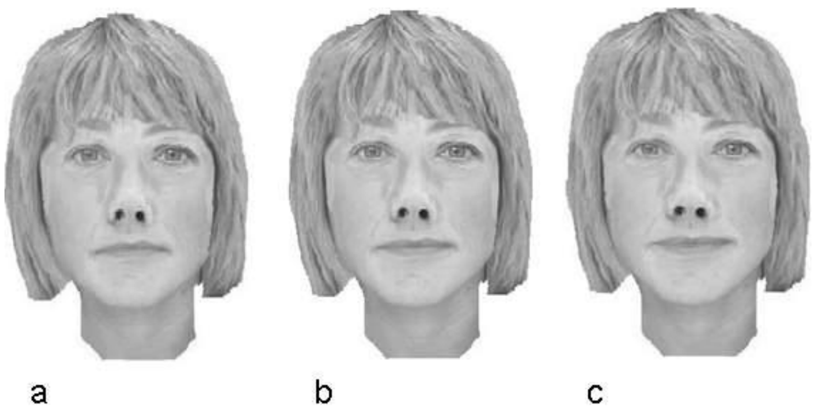


Figure 3.6. (a) An original ProFIT facial composite of a lecturer; (b) the ProFIT facial composite with 30% of the enhanced-expression transformation applied; (c) the ProFIT composite with 60% of the enhanced-expression transformation applied. The expression transformation shifts the facial landmarks of composite image in the way that on average a female face would smile.

Participants

Twelve participants (2 male) were recruited from staff and students at Stirling University. They ranged in age from 21 to 55 years (mean 36, *s.d.* 11.8). All participants were familiar with the University of Stirling targets, and all had normal or corrected to normal vision; none received payment.

Procedure

Four participants were randomly allocated to each presentation set and were tested individually; they were shown one pair of images at a time and asked to judge whether the images were identical or differed in some way, no other information was supplied and no feedback was given.

Results and Discussion

Participants correctly judged the original face pairs to be identical 78.3% of the time (*s.e.* 7.6%), and accuracy was significantly better than chance (50%), $t(11) = 3.7$, $p = .003$, $d = 1.5$. Discrimination of original facial composites and composites enhanced by 60% of a smile was also good at 80% (*s.e.* 5.5%), and again significantly better than chance, $t(11) = 5.5$, $p < .001$, $d = 2.2$. However, ability to distinguish the original composites and composites enhanced by 30% of the smile was much poorer at 53.3% (*s.e.* 9.6%) and did not differ from chance, $p = .7$. These effects are replicated across the set of composite items: correct identification of original composite pairs was significantly better than chance, $t(14) = 6.1$, $p < .001$, $d = 2.2$; discrimination of original composites and composites transformed by 60% was also significantly better than chance, $t(14) = 4.3$, $p = .001$, $d = 1.6$; while discrimination of original composites and composites transformed by

30% was at chance levels, p = .6. The results indicate that while expression-enhancement by 30% of the smile transformation can significantly improve facial composite identification, the alteration to the composite image is too subtle to be reliably detected in an explicit discrimination task. It is however, on the threshold for detection, and may thus be an effective and acceptable forensic manipulation. Experiment 10 explored the limits of this effect for composite identification with both weaker and stronger levels of expression-enhancement.

Experiment 10

Identification with different levels of transformation

The previous experiments showed that affective information is important for identification of facial composites and provided evidence that facial composite identification can be significantly improved if the images are manipulated to display more positive facial expressions. By quantifying how an average a face will change when a person smiles, it was possible to transform each facial composite by a 30% proportion of this transformation. This manipulation was highly effective for composite identification yet was not reliably perceived in an explicit discrimination task. While this indicates that affective influence of the expression-enhanced composites was not explicit or consciously perceived, it raised the possibility that a weaker transform would also be effective and prompts the question of whether an explicit expression transform would produce better improvements. Experiment 10 examined the influence on facial composite identification of a weaker 25% expression transform and of a stronger 50% expression transform.

Materials and design

The University of Stirling and Napier University composites were manipulated with the Psychomorph software and templates to generate a set of expression-enhanced composite images showing 25% of the smile transformation, and a set comprising 50% of the transformation. A 2 x 2 mixed factor design was employed with composite type (original composites; expression-enhanced composites) as the within participant factor, and test set (25% transformation; 50% transformation) as the between participant factor. Original composites and 25% expression-enhanced composites were allocated to two sets of power point slides; each contained one exemplar of each facial composite, half were expression-enhanced and this factor was counterbalanced across the test sets. Equivalent materials were generated for the 50% expression-enhanced composites. Examples of composite stimuli are shown in figure 3.7.

Figure 3.7. (a) An original ProFIT facial composite of a University of Stirling lecturer; (b) the ProFIT facial composite with 25% of the enhanced-expression transformation applied; (c) the ProFIT composite with 50% of the enhanced-expression transformation applied. The male expression transform shifts the facial landmarks of composite image in the way that on average a male face would move when it shows a smile.

Participants

Forty-eight participants (9 male) were recruited from staff and final year psychology students at Stirling University. They ranged in age from 20 to 54 years (mean 28.2, *s.d.* 10.7). All had normal or corrected to normal vision and all were familiar with the Stirling targets, they were paid £2 for their time.

Procedure

Participants were randomly allocated to one of the four test sets and were tested individually or in groups of up to six. The procedure replicates the composite recognition method employed in Experiment 8.

Results

In the 25% transformation group original facial composites had an identification rate of 15.3% (*s.e.* 3.5%), which was lower than identification of the expression-enhanced composites, 23.6% (*s.e.* 3.4%), while in the 50% transformation group original composite identification was 25.7% (*s.e.* 3.5%) rising to 29.9% (*s.e.*3.4%) when the images were enhanced. The data were analysed with a 2 x 2 x 2 mixed factor repeated measures analysis of variance with composite type (original composites; expression-enhanced composites) as the within participant factor, and level of expression transformation and test set as the between participant factors. There was no significant between subject effect of test group, $p = .6$. The main effect of composite type marginally failed to reach significance, $F(1,44) = 3.8$, $p = .06$, $\eta_p^2 = .08$, but did account for 8% of the variance in accuracy scores. A significant main effect of transformation level, $F(1,44) = 4.9$, $p = .03$, $\eta_p^2 = .10$ accounted for 10% of the variance. There was no significant interaction between test group and composite type, $p = 1$, between test group and transformation level, $p = .2$, between transformation level and composite type, $p = .5$, and no significant three way interaction between test group, transformation level, and composite type, $p = .7$.

Expression enhanced facial composites were identified better than original composite images. However, composite identification was uncommonly high in the 50% transformation group, and original facial composites achieved an identification rate of 25.7%, which might have masked effects of the expression transformation. Comparison of original and expression-enhanced composite images by group found no significant improvement for the expression-enhanced images in the 50% transformation group, $p = .4$, but within the 25% transformation group the expression-enhanced images did achieve significantly better rates of identification, accounting for 19% of the variance, $F(1,23) = 5.3$, $p = .03$, $\eta_p^2 = .19$. However, these effects were not replicated across the set of composite items: a 2 x 2 analysis of variance with composite type and level of transformation as factors found no significant main effect of composite enhancement, $p = .2$, no effect of transformation level, $p = .2$, and no significant interaction, $p = .6$. Examination of false identification also found no significant main effects of composite type, $p = .8$, level of transformation, $p = .1$, and no significant interaction, $p = 1$. The incorrect identification rate for the 25% transformation group was 7.8% (*s.e.* 1.5%), and for the 50% transformation group was 11.3% (*s.e.* 1.5%).

Identification of facial composites was both more successful and elicited higher levels of false identification in the 50% transformation group; the exceptionally high identification rates within this group were provided by two sets of participants using the two sets of counterbalanced test materials and the results obtained from each subset were not significantly different, $p > .5$. These groups did not differ from the 25% transformation groups in terms of age ($p > .3$), gender ($p > .7$), or experimental conditions, which would suggest that the nature of the test materials rather than any extraneous factor caused the participants in the 50% group to adopt a more lenient response criterion and be more willing to offer identification responses.

Signal detection measures of *d'* sensitivity and response criterion *c* were calculated from the hits and false positive rates of both sets of participants. Paired sample *t*-tests found that sensitivity was marginally improved when 25% of the smile transform was applied (*d'*=.34 *v's* *d'*=.95), $t(23) = 1.8$, $p = .08$; but in the 50% transformation group sensitivity for original items was not significantly altered when the expressions were enhanced (*d'*=.68 *v's* *d'*=.87), $p = .7$. Examination of response criterion found that in the 25% transformation group response criterion for original composite images was significantly lowered when the expressions were enhanced ($c = 1.8$ *v's* $c = 1.4$), $t(23) = 2.2$, $p = .04$, but in the 50% transformation group a more lenient response criterion was shown for both original composite images and expression-enhanced composite images which did not significantly differ ($c = 1.3$ *v's* $c = 1.2$), $p = .6$. These results suggest that when the expression-enhancement is difficult to perceive, sensitivity to the composite images will improve as will the willingness to offer a name; when expression-enhancement is explicit and some of the composite images show explicit positive expressions, sensitivity is comparable for all of composite items and is associated with a stronger overall tendency to offer an identification response.

Discussion

An expression transformation that manipulated facial composites to show 30% of a smile was difficult to see but was extremely effective for identification (η_p^2 = .47). This suggested that the affective influence was not conscious and raised the possibility that an even weaker transformation could also be effective. When composite images were enhanced by 25% of the smile transform the effects were replicated but the effect size was somewhat smaller (η_p^2 = .19), and while sensitivity also improved, the effect did not achieve significance at the 0.05 level. However, the response threshold was significantly lower for expression enhanced images, indicating that although the weaker 25% transformation was less effective in terms of sensitivity or identification, it nonetheless influenced participant decisions. These findings suggest that affective information that is too subtle to be explicitly perceived will influence identification judgements, but that accuracy may require some perception of affect to enhance sensitivity or level of attention.

Expression-enhancement of facial composites by 50% of the smile transform enabled assessment of the smiling face bias when positive facial composite expressions were easier to perceive. The identification rates for these participants were uncommonly high in facial composite research and failed to replicate the findings of the 30% and 25% transformation groups. For the 50% expression-enhanced composites there was no identification advantage (η_p^2 = .03), sensitivity was not improved, and response bias was not reduced relative to the original composites. This experiment sought to establish whether a stronger explicit expression transformation could produce even better levels of identification than expressions that were difficult to perceive: there is no evidence that 50% expression-enhanced images were identified better than the original composites viewed at the same time, but overall sensitivity was better, the response threshold was lower and identification of all of the composite images was exceptionally high in comparison with the 25% and 30% expression transformation groups.

Given such unusually high identification levels for both enhanced and original composite images, it is possible that the presence of noticeably positive composite images may have influenced participants to attend more carefully to all of the composite items, and with an increased level of attention and analysis they were also more willing to offer up a name. In this way, explicit positive affect may have influenced the level of attention and judgement for both the enhanced and the original composite images. Endo et al. (1992) found that participants took longer to inspect unfamiliar faces when they were smiling, which would suggest that enhanced attention should be confined to the expression-enhanced composites images. However, a smiling face bias in facial composite identification may not replicate Endo et al. (1992) because, in contrast to unsmiling unfamiliar faces, the facial composites are all novel face images that are nonetheless intended to portray familiar people.

The results provide more evidence that facial affect influences identification judgements for facial composites, but are problematic with regard to the extent and conditions of the effect. Perhaps the smiling face bias operates at more than one level: when expressions are very subtle only face images that embody positive affect will non-consciously engage more attention, but when expressions are explicit any face that signals some level of familiarity will consciously receive more attention. These conclusions are speculative and in most visual discrimination or identification tasks it would be possible to record reaction times as an index of attention, or masking procedures to study covert effects; but the time taken to achieve facial composite recognition is highly variable, can be lengthy, and is subject to a wide range of individual factors (i.e. level of personal contact or familiarity). In Experiment 4 facial composite reaction time data was obtained in a cued identification task but as names cue familiarity, this paradigm would confound positive expression familiarity effects. It could, however, be used to study negative expressions; Lander and Metcalfe (2007) found that negative expressions reduced perceptions of familiarity. Therefore, one would predict that composites showing negative expressions would fail to engage as much attention as original images. In order to establish reliability of the 50% transformation group results, it would be advisable to replicate this experiment with a different set of composite images, and to repeat the evaluation with a blocked design: if mixed presentation with the explicit expression-enhanced composites also caused original composite images to be examined and identified better, one would predict that presenting enhanced and original composites in blocks would cause original items to be processed less carefully and identification rates would be lower.

Summary and conclusions

Studies have consistently shown that smiling faces are judged to be more familiar than neutral faces or faces showing negative expressions, particularly when identification is not easy (Baudouin et al., 2000; Davies & Milne, 1982; Endo et al., 1992a; Gallegos & Tranel, 2005; Kottoor, 1989; Lander & Metcalfe, 2007). Facial composites are produced from memory by unfamiliar witnesses who will generally have experienced an unpleasant event, and as far as possible, the composites will portray those negative qualities (Parr & Richardson, personal communication). As composite identification requires a match between the information that a witness can provide and the information needed to achieve familiar face recognition, these affective qualities may be very important. The aim of this chapter was to establish whether the smiling face bias is an important consideration for facial composite identification, and to explore the effectiveness of an image manipulation designed to enhance facial expressions in composite images. Manipulating composites to show more positive expressions was expected to enhance composite identification, but might also increase false identification.

Sets of smiling and neutral face pairs were used to calculate how, on average, a face moves when a person smiles; the resulting transformations from neutral to smiling expression could then be applied to any face image causing it to show the transformed positive expression. Experiment 7 altered celebrity facial composites to show 30% of the average smile transform and confirmed that manipulating facial composites to show more positive expressions enhanced identification. Experiment 8 replicated these findings with non-celebrity facial composites. When the composites represented personally familiar people, the smiling face bias was much stronger, indicating that the recognition advantage has less to do with stored representations of smiling faces, than with association between positive affect and familiarity. The combination of increased sensitivity with reduced response criterion suggests that positive expression may have induced participants to attend more closely to the enhanced composite images, which consequently made them more willing to offer a name.

A goal of this research was to develop a forensic technique that might enable more accurate facial composite identification in applied settings. While this work attempts to harness the smiling face bias it would be unacceptable to display ‘live' facial composites with explicit smiling expressions. The 30% transform was adopted in Experiments 7 and 8 because it produced alterations that, to the experimenter, could be perceived but did not display explicit smiles. Experiment 9 established that while a 30% transform significantly improved identification, alterations at this level were not reliably discerned by people familiar with the targets. This suggests that the smiling face bias in facial composite identification can operate at a non-conscious level.

The final experiment explored different levels of expression enhancement. A 25% transform replicated Experiment 8 but produced weaker effects. Thus, although a smiling face bias does not require explicit perception, the manipulation must be strong enough to produce a robust effect. A stronger 50% transform enabled assessment of the smiling face bias when composites showed explicit expressions. Identification rates were exceptionally good, but did not show improved identification of expression-enhanced composites. These results suggest that positive expressions can improve facial composite identification, but that performance and behaviour will be different for explicit smiles.

Theoretical considerations

Facial composite identification requires a match between the information reproduced by an unfamiliar witness, and a stored representation of a familiar person. Because of inaccuracies in facial composites, forming a correspondence will require effort on the part of the viewer and where identification is difficult, the presence of positive affect will be important to signal familiarity (Baudouin et al., 2000; Davies & Milne, 1982; Ellis & Young, 1990; Endo et al., 1992; Gallegos &

Tranel, 2005; Kottoor, 1989): where the image portrays negative affect, it is more likely that it will be rejected as unfamiliar (Lander & Metcalfe, 2007).

The nature of facial composites and the type of events that make them a necessity entail that in many cases composite images will portray negative qualities; but the evidence indicates that this will have important consequences for identification: as members of the public fail to immediately identify accurate composite features, negative valence will signal that the person portrayed is unfamiliar and the composite may fail to engage attention for any useful length of time. Even where a composite has elicited some level of familiarity and the possibility of identification, a lack of positive signal will promote a more conservative criterion and positive identification will be less likely.

The Bruce and Young (1986) model proposes that expression and identity are processed in parallel pathways *(see figure 1)*, and while this is supported by neuropsychological studies and neurophysiological observations (e.g. Caharel et al., 2005; Hornak et al., 1996), it has been shown that identity and emotional expression processes interact when identification is difficult (e.g. Endo et al., 1992; Ganel & Goshen-Gottstein, 2004). Recognition of a composite requires correspondence with a stored representation and will take longer than recognition of an accurate face image (Bruce, 1982); this will allow affective information to contribute to cognitive appraisal of the face image. At the most basic level, the presence of positive affect will signal the likelihood of familiarity, while negative affect will indicate an absence of familiarity. This may be sufficient to determine whether structural analysis of the face image will continue and is consistent with the findings that sensitivity, and potentially attention, was increased for the expression enhanced composite images. Where elements of the composite image have been recognised but activation of a PIN is not sufficient to promote name generation, affective information from the parallel expression analysis route may feed back information from the cognitive system to the PINs, and the combined input from the FRU and the cognitive system may then reach the activation threshold for name generation. This proposition is consistent with an increase for facial composite identification but no elevation of false identification rate; for the smiling face bias to enhance facial composite identification a specific person representation must be activated and the affective information is therefore supplementary.

Applied considerations

The process of constructing a facial composite is designed to enable the witness to show what a perpetrator looked like and the interview procedure employs context reinstatement and imagery to recreate the event, in order to help the witness remember as much detail as possible. Memories of unfamiliar faces are context specific, and while this means the structural representation of the face will be defined by the environmental conditions, the character and affective quality will depend on the emotional conditions at the time of encoding. As most criminal acts are perceived by victims and bystanders as extremely unpleasant, these are the affective qualities that will be encoded as part of the face memory and they will be reproduced in any subsequent composite likeness.

The smiling face bias is shown to be important for facial composite identification. Therefore, real life facial composites that portray negative facial affect will significantly reduce the potential for identification by a member of the public. These results demonstrate that composite images can be modified to reduce negative qualities, and it is shown that very subtle smiling expressions are sufficient to substantially improve facial composite recognition. As a finished facial composite must be signed by the witness and cannot thereafter be altered, police practitioners should consider employing expression modification within the composite construction process. Within the evolutionary composite systems (e.g. EvoFIT) the face images are already defined by templates and can be modified at the click of a mouse, what is more by incorporating the full neutral to smiling transform the facial composite image could be adjusted by way of a slider to establish the best level of transformation that the witness will accept.

Conclusions

This chapter explored the influence of affect on facial composite identification and found that like real face recognition, identification judgements are facilitated by positive facial expressions. The effect was found with celebrity composites but was stronger for composites of personally familiar people, indicating that positive affect can mediate successful activation of specific face memories. All of the evaluations employed composites produced in the lab and as such they do not portray the type of negativity one would expect the victim of a serious crime to reproduce. Because real composites will reproduce highly negative emotional affect, it is likely that the effects of the smiling face bias, or rather a negative expression bias, will be even more pronounced within an applied setting. It is therefore intuitive that this will necessitate stronger levels of expression-transformation and also that an optimal level of transformation will be specific to each individual composite image.

The results of this work are also of consequence to any research where facial expression is of primary concern. Studies of interaction and behaviour generally incorporate stimuli designed to portray unambiguous positive, neutral, or negative affect (e.g. Pictures of Facial Affect, Ekman & Friesen, 1976; Karolinska Directed Emotional Faces, Lundqvist, Flykt, & Öhman, 1998) but this work demonstrates that barely perceptible expression can significantly alter behaviour; therefore the reported effects of 'gross' expression may mask more subtle but important effects. Finally, these results show that humans are extremely sensitive to subtle affective facial cues and that such cues will influence decision making; the use of cosmetic procedures that immobilise facial muscles have become common place and it would be useful to investigate the consequence of such procedures for both visual and efferent face to face communication.

4

Familiarity bias in face matching

Discrimination of unfamiliar faces is difficult, and where facial identification requires effort, affective information has been shown to influence judgements of familiarity (e.g. Baudouin et al., 2000). Positive expressions enhance perceived familiarity, while negative expressions will produce the opposite effect (Lander & Metcalfe, 2007). In Chapter 3 the smiling face bias was found in identification of facial composites. The effect was associated with increased sensitivity for composites that were manipulated to show a more positive expression; consistent with Endo et al. (1992), it was proposed that positive affect induced participants to attend better to the images that had been enhanced. This chapter comprises a series of experiments designed to evaluate whether positive affect can also induce attention, and consequently improve performance, in face discrimination tasks. However, as a smiling face bias can increase false familiarity judgments, this work also explores whether positive affect will elicit more false positive face matching responses.

Discrimination of unfamiliar faces

Face perception ability varies considerably from person to person (e.g. Bruce et al., 1999; 2001; Megreya & Burton, 2007) and it varies consistently and substantially depending on the familiarity of the face. When faces are familiar, recognition is effortless and even in difficult viewing conditions, such as with poor quality CCTV, identification of familiar people will be almost perfect; in these conditions recognition of unfamiliar faces will be near to chance levels (Burton et al., 1999). Perceptual discrimination of faces from CCTV footage mirrors these effects; ability to match photographs to CCTV images was significantly poorer for unfamiliar participants who showed reduced sensitivity to the faces in comparison with people who were familiar with the targets (Bruce et al., 2001).

Confirming identity from a person to photographic identification is an example of unfamiliar face matching and while it is common practice and vital for border control, evidence shows that we are consistently bad at making these judgements. Kemp et al. (1997) demonstrated that if a person bears some resemblance to the image on a fraudulent photo-card, the acceptance rate will be around 64% and even if the resemblance is poor, the ID will be accepted about one third of the time. Within the legal system juries may also be asked to match a defendant to images or security footage presented as evidence in court; Davis and Valentine (2009) have shown that matching a live suspect to a person in a video produced a false identification rate of 17%, yet participants failed to identify one fifth of the correct matches. The images that are shown on photographic identification and on CCTV footage are often unclear, but while image quality is an important factor, Bruce et al. (1999) found that even with good quality images that were captured on the same day, accuracy did not exceed 70% for unfamiliar face matching.

In spite of substantial evidence that unfamiliar face matching is prone to high rates of error and although the cost of such errors can be high, identification from face images is less invasive and cheaper than biometric alternatives, and will become increasingly common. With a growing reliance on visual media and applications, image quality in identity verification is likely to improve; but to significantly improve accuracy, it will also be important to identify methods of making unfamiliar faces more distinguishable.

The smiling face bias and facial discrimination

When face recognition is difficult, positive expression will enhance perceptions of familiarity and negative expressions will produce the opposite effect. In Chapter 3 manipulating facial composites to show subtle positive expressions significantly improved identification and was marked by greater sensitivity and a reduced criterion for offering an identification response. This was the first demonstration of a smiling face bias for identification of facial likenesses. The effect has typically been shown in speeded identification of familiar faces, and in erroneous recognition judgements for unfamiliar faces. The influence of positive affect on false recognition might suggest that a smiling face bias would be detrimental to any unfamiliar facial identification procedure; but facial image comparison does not involve memory judgements, and as smiling faces enhance sensitivity and engage attention for longer (Endo et al., 1992), positive affect could facilitate attention and differentiation of unfamiliar faces.

Aim of study

The effect of expression on unfamiliar face matching was previously evaluated by Bruce et al. (1999) in a simultaneous face matching task where the smiling targets were matched to arrays of neutral images. It was shown that if the facial expression of the target image and the potential matches is different, correspondence of matching items will be significantly impaired. The research presented here was not designed to consider face matching across facial expression but to determine whether positive facial affect could be used to enhance attention and facial discrimination.

Experiment 11 evaluated whether discrimination of pairs of face images would be enhanced when the images showed more positive expressions. If positive affect engages more attention, it was predicted that sensitivity to the unfamiliar face images would be enhanced. Positive affect is also associated with a more liberal response criterion; therefore it was anticipated that false identification might also increase. In applied settings a target image may be compared with an array of images. Positive affect enhanced sensitivity to individual facial composites, but this effect might not be found when a number of faces are shown simultaneously: Experiment 12 therefore explored the influence of positive affect on face matching with multiple item arrays. It was predicted that with simultaneous comparison of a number of images, positive affect might fail to enhance sensitivity, but that reduced criterion could elevate identification judgements and increase the rate of false positive identification. To determine whether positive affect could enhance discrimination of individual faces, yet accomplish multiple comparisons, Experiment 13 employed a sequential matching procedure.

When the target and the array items show different expressions correspondence of the unfamiliar faces will be reduced and matching will be impaired (Bruce et al., 1999). However, in applied settings face image comparison often requires that individual persons be matched to photographs or footage in which expressions differ. Live face matching does not involve simultaneous comparison of images as the face and the image are regarded one at a time. To explore the effects of differing expressions in this type of practice, Experiment 14 employed a delayed face matching procedure in which only one face was manipulated to show positive affect. If differing expression reduced correspondence the rate of selections would reduce; but if positive affect enhances familiarity, false identification would increase. Positive affect may, however, enhance sensitivity, in which case face matching could improve.

Within this series of experiments it was possible to explore the influence of positive affect in perceptual face matching tasks, and thus to determine whether memory is a necessary component of affective bias in face recognition judgements.

.

Experiment 11

Enhancing attention with positive affect

Experiment 11 evaluated whether affective information could be manipulated to improve perceptual discrimination of pairs of face images. If positive affect engages more attention, it was predicted that sensitivity to the images would be enhanced and discrimination would improve; but as positive affect is also associated with a more liberal response criterion, false identification might also increase.

Participants

Forty-three participants (7 male) aged between 18 and 48 years (mean 22.2, *s.d.* 6.1) volunteered in return for course credit. All were students from the University of Stirling and all had normal or corrected to normal vision.

Materials and Design

This series of experiments employed the face matching materials compiled by Bruce et al. (1999). Original matching face pairs comprised a photograph and a video still of the same individual, both showing full face neutral poses. Non-matching original face pairs comprised the photograph of the target with the video still of a second individual that was paired with the target most often in a similarity matrix (Bruce et al., 1999). To create face items with enhanced positive affect the Psychomorph software package (Tiddeman, Burt, & Perrett, 2001) was used to create a template of each image and the smile transformation that was developed for Chapter 3 was then applied.

Face pairs were created for 80 targets and each target was viewed by each participant once. A 2 x 2 repeated measures design was employed: half of the face pairs showed the target image with the correct match (target present) and half with the foil image (target absent). In half of the target present and target absent trials, both of the images were transformed to show positive affect. The left / right position of photographs and video stills was equally sampled and trials were presented in fully randomised order. All conditions were counterbalanced across participants. Examples of the face matching pairs are shown in Figure 4.1.

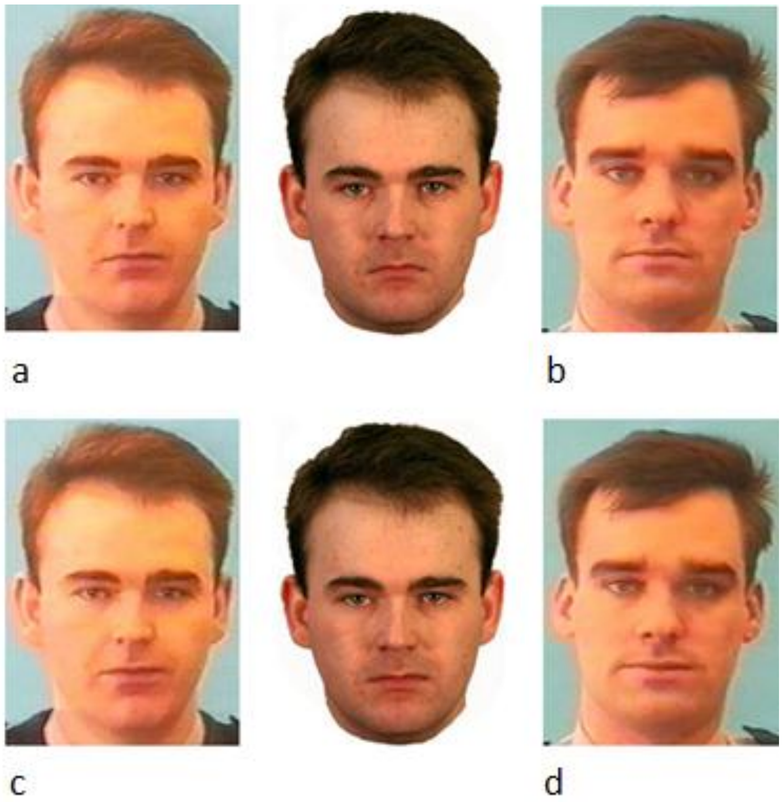


Figure 4.1. Examples of the original and expression enhanced face matching pairs. Expression enhanced images were generated using Psychomorph imaging softeware (Tiddeman et al., 2001) to apply 30% of the smiling expression transformation developed for the experiments in Chapter 3. The top row shows (centre) the original target; (a) the original correct match; (b) the original foil. The bottom row shows (centre) the expression enhanced target; (c) the expression enhanced match; (d) the expression enhanced foil. On half of the trials the target was paired with the correct match and on half with the foil image.

Procedure

Participants were tested individually. The experiment was conducted using E-Prime software with a 17 inch LCD monitor at 1024 x 768 pixels resolution and responses were recorded via the keyboard. Each trial consisted of a photograph and a video still shown side by side on a white background at the vertical mid-point of the screen. The head images gave a viewing angle of approximately 2.5° at 70 cm viewing range and were separated by a distance of 125 mm. The instructions were provided both verbally and with on screen commands. Participants were told that they would be shown pairs of face images and should decide whether the pictures were of the same person, or of different people. If they thought the images were of the same person they were instructed to press ‘m’ on the keyboard; if they thought that the images were of different people, they were asked to press the ‘z’ key (this instruction was always visible and keys were reversed for half of the participants).

Results

Contrary to predictions, positive affect reduced the rate of matching responses, as shown in both fewer correct identifications and lower levels of false identification. The means and standard deviations for accurate responses are shown in table 4.1.

	Original images		Expression enhanced images	
	M	*S.D.*	M	*S.D.*
Matched pairs	88.5	*13.0*	82.5	*16.9*
Foil pairs	73.9	*17.5*	77.4	*18.3*

Table 4.1. Evaluation of face matching for original and expression enhanced face pairs. The table shows the percentage of correct response, and associated standard deviations, in each condition. Correct matching was impaired when the images were enhanced to show positive affect but accurate rejection of the foils improved.

Accurate response rates were analysed with a 2 x 2 repeated measures analysis of variance with target presence and affective transformation (original images, expression enhanced images) as factors. There was a significant main effect of target presence, $F(1,42) = 6.9$, $p = .01$, $\eta_p^2 = .14$, no significant main effect of expression enhancement, $p = .3$, and a significant interaction between these factors, $F(1,42) = 14.3$, $p < .001$, $\eta_p^2 = .25$. Paired sample *t*-tests showed that when the face pairs did not match they tended to be rejected better when the smile transform was applied, $t(42) = 1.9$, $p = .06$, $d = 0.2$, but when the images did match, positive affect significantly impaired face matching performance, $t(42) = 3.4$, $p = .002$, $d = 0.4$.

Correct identification and false positive rates were combined to determine signal detection measures of sensitivity (*d′* prime) and response bias (criterion *c*) for the original and expression enhanced face pairs. In line with the effects on facial composite recognition, it was predicted that positive affect would enhance sensitivity but would produce a more liberal response criterion. However, sensitivity was better for the original images, $d' = 2.5$, than for the expression enhanced images, $d' = 2.1$. A paired sample *t*-test found that in contrast to ognition judgements, for perceptual discrimination the presence of increased positive affect significantly reduced sensitivity, $t(42) = 2.8$, $p = .007$, $d = 0.4$. Response criterion was also higher for the expression enhanced images, $c = -0.5$, than for the original images, $c = -0.2$, $t(42) = 4.5$, $p < .001$, $d = 0.4$: in a perceptual discrimination task positive affect elicited a more conservative response bias, as shown in fewer matching decisions.

Discussion

A smiling face bias is consistently found to increase perception of familiarity. In Chapter 3 it was shown that this effect might be mediated to some extent by enhanced sensitivity and attention, together with a reduced response criterion and greater willingness to make an identification judgement. It was predicted that enhanced sensitivity with reduced response criterion would also be observed in a face discrimination task, but a contrasting effect was obtained: when the face pairs were manipulated to show positive affect, discrimination declined and a more conservative response criterion was adopted.

The results support findings that expression influences identification judgements (Baudouin, Gilibert, Sansone & Tiberghien, 2000; Gallegos & Tranel, 2005; Kottoor, 1989; Davies & Milne, 1982; Endo et al. 1994; Lander & Metcalf, 2007), but indicate that while positive affect will enhance recognition, perceptual match judgments will decline. This might suggest that in face matching positive affect was distracting, but transforming faces to show a smile may have exaggerated perceptual differences produced by different image properties, making correspondence of matching images less likely. If the matching detriment is caused by image properties, the image transformation should reduce accuracy in any face matching procedure, but if positive affect is distracting, one might expect this effect to be weaker when simultaneously comparing multiple images.

Experiment 12

Applying positive affect to parallel arrays

In Experiment 11 sensitivity was reduced and fewer matching decisions were obtained for expression enhanced face pairs. If the image manipulation exaggerated differences between the images, a similar effect would be observed in matching with multiple images; but if affective quality altered behaviour, the effect might be weaker for multiple image comparisons. Experiment 12 explored whether comparison of a number of images would reduce the effect of the expression transform, or whether consistent imaging effects would be observed.

Participants

Nineteen participants (8 male) were recruited by opportunity sampling at Stirling University. They ranged in age from 19 to 63 years (mean 27.2, *s.d.* 10.7). All had normal or corrected to normal vision and were paid £2.

Materials and Design

Experiment 12 employed the 80 face matching arrays from Bruce et al (1999). Trials consisted of a video still target image shown above the 10 photographs paired with the target most often in a similarity matrix (target absent condition), or above the 9 most similar images with a photograph of the target (target present condition). The items were arranged in two rows and numbered 1-10. Target position within the arrays was randomly sampled with the constraint that each position was equally sampled for original and expression enhanced arrays. Within the expression enhanced trials both the target and array items were transformed. Target images were cropped to show head and shoulders (50 mm x 80 mm); array images were cropped to show only the head. All of the head images gave a viewing angle of approximately 2.5° at 70 cm viewing range. The complete array with target image measured 260 mm x 270 mm. An example of an original array is shown in Figure 4.2; an example of an expression enhanced array is shown in Figure 4.3.

A repeated measures design employed target presence and image type (original images; expression enhanced images) as factors. Half of the arrays showed original images and half the expression enhanced images; the target was present in half of each set. Presentation of the trials was fully randomised and all conditions were counterbalanced across participants.

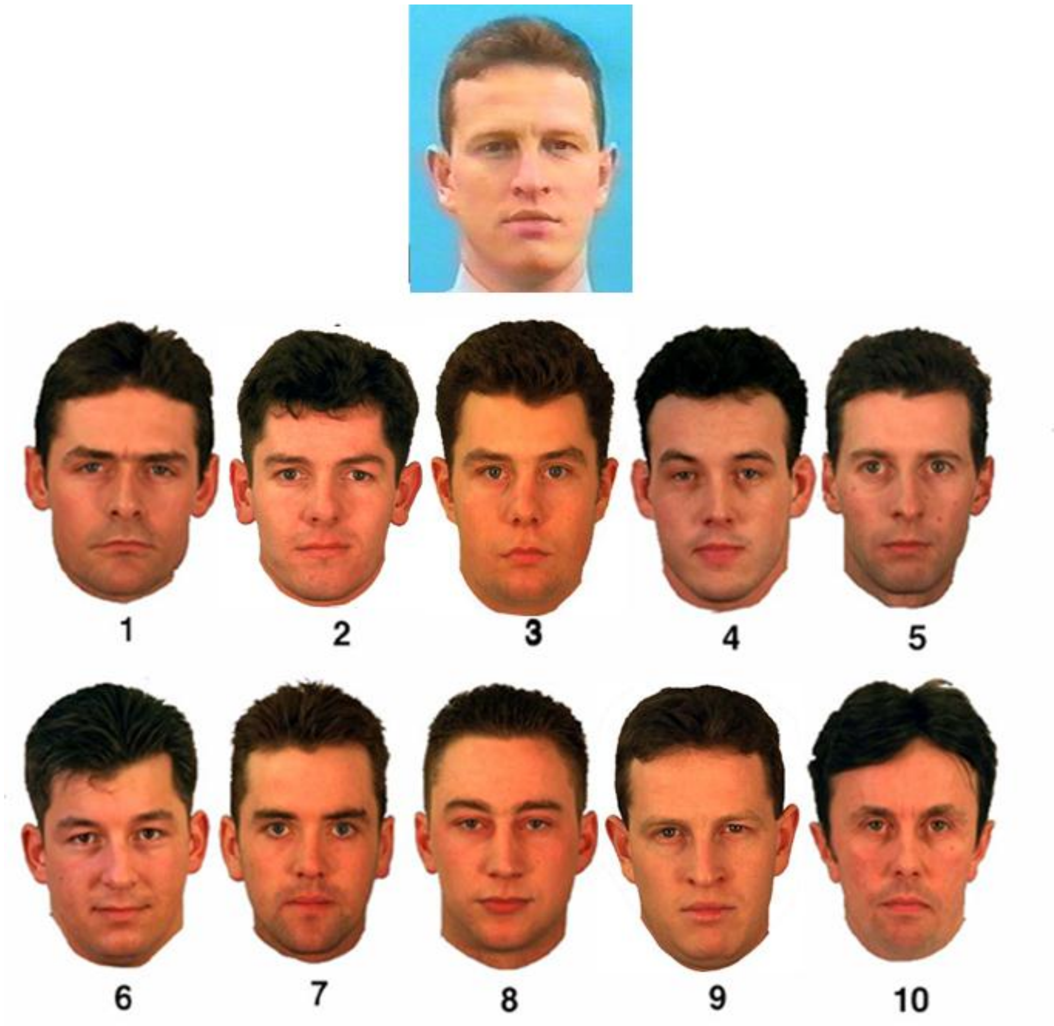


Figure 4.2. Example of an original simultaneous face matching array. The target, or probe image shows a video still. The array shows photographs of 10 males with neutral expressions. In target present arrays, position of the correct match was equally sampled.

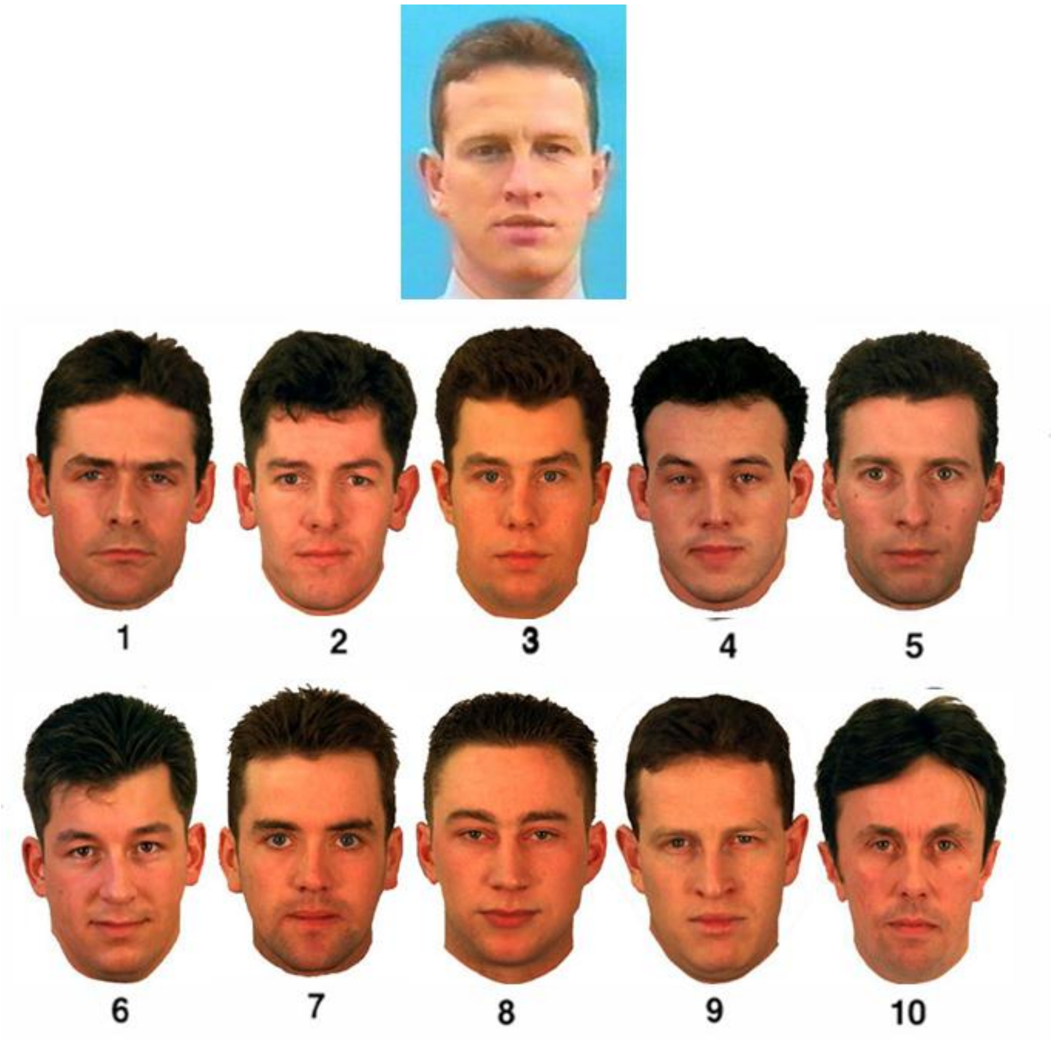


Figure 4.3. Example of a simultaneous face matching array comprising expression enhanced faces. Both target and array items were manipulated using Psychomorph software to enhance expressions by 30% of the smiling expression transform developed for Chapter 3.

Procedure

The experiment was conducted using E-Prime software with a 17 inch LCD monitor at 1024 x 768 pixels resolution; responses were recorded via the keyboard. The participants were tested individually and the instructions were provided both verbally and with on screen commands. They were asked to complete a face matching task and informed that the target shown at the top of the screen may, or may not, be pictured in the accompanying line-up array. To identify an array image as a match they were to type the corresponding number, but if no match was identified they were instructed to press the space bar.

Results

Within multiple item simultaneous arrays positive affect again impaired accurate face matching performance, but within this procedure correct rejection of the target absent arrays also declined. The means and standard deviations for different types of response are shown in table 4.2.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	56.1	*37.7*	19.3	*23.1*	24.6	*32.6*	60.5	*32.5*
Expression enhanced	48.3	*36.8*	24.6	*25.1*	27.2	*24.3*	46.5	*29.7*

Table 4.2. Evaluation of face matching for original and expression enhanced faces within ten item simultaneous arrays. The results are broken down by the percentage of hits, misses and false identification for target present arrays, and the percentage of correct rejections for target absent arrays. Correct matching was impaired when the images were enhanced to show positive affect.

Accurate responses were analysed with a repeated measures analysis of variance with target presence and image type as factors. There was a significant main effect of image transform, $F(1,18) = 9.0$, $p = .008$, $\eta_p^2 = .14$, but no significant main effect of target presence, $p = .8$, and no significant interaction, $p = .3$. Overall face matching performance was poor but performance was significantly reduced when the expression enhancement was applied, with the smiling expression transformation accounting for approximately 25% of the variance in the face matching scores. For the incorrect responses to target present arrays, paired sample comparison indicated no significant increase in either misses, $p = .3$, or false identification, $p = .5$. Sensitivity for the original images was $d' = 0.6$ and for the expression enhanced images was $d' = -0.6$; a paired sample *t*-test showed that discrimination was significantly impaired by the expression transformation, $t(18) = 3.2$, $p = .005$, $d = 0.3$. There was no effect of expression enhancement on measures of response criterion, $p = .2$ (original $c = 0.2$; enhanced $c = 0.01$).

Discussion

Within the simultaneous face matching procedure, sensitivity was reduced for images that were manipulated to show positive affect, indicating that the expression transformation increases disparity between matching faces. When faces were matched one at a time response bias also became more conservative; but in the simultaneous procedure this did not occur, suggesting that the influence of positive affect may have been weakened in comparison of multiple faces. Within the simultaneous face matching procedure the expression transformation significantly impaired accurate face matching performance.

Experiment 13

Positive affect within sequential face matching arrays

The final face image comparison experiment presented multiple item arrays in a sequential format. It was initially predicted that sensitivity would be improved by positive affect and that sequential presentation would elicit a conservative response criterion and reduce false positive identification. However, given the results of Experiments 11 and 12, it was expected that the expression transformation would impair discrimination, but that the sequential format with the expression manipulation would elicit a conservative response criterion and false positive identification would not increase.

Participants

Twenty participants (6 male) were recruited from staff and students at Stirling University. They ranged in age from 16 to 49 years (mean 28.6, *s.d.* 12). All had normal or corrected to normal vision and none received payment.

Materials and Design

Sequential face matching arrays were generated from the materials employed in Experiment 12. Within the sequential procedure each trial consisted of the target image shown to the left of each array item in turn and the order of array items was randomised with the constraint that target position was equally sampled. A repeated measures design was employed and presentation of all 80 arrays was randomised for each participant; half contained original faces and half contained expression enhanced faces, with the target present in half of each set. All conditions were counterbalanced across participants.

Procedure

The trials were presented with E-Prime software *(apparatus as before)*. Each pair of images was presented at the vertical mid-point of the screen and was separated by 125 mm with each head subtending a viewing angle of 2.5 ° at 70 cm viewing distance. Participants were tested individually and instructions were provided verbally and with on screen prompts. The participants were told that they would be shown pairs of faces and should decide whether the images were of the same person, or of different people. They were advised that a number of faces could be shown for each target and that there may or may not be a matching face within each set. The participants were asked to respond in their own time; if they judged the faces to be of different people they were to type 'n', and if they thought the images were of the same person they were to type 'y'. In accordance with Lindsay and Wells (1985) each item was viewed once, the number of array items was not disclosed, an array was shown until a selection was made or until all of the items had been rejected, items could not be reviewed, and decisions could not be changed.

Results

Within the sequential arrays positive affect did not influence face matching accuracy but a sizeable increase was observed in false identification within target present arrays. The means and standard deviations for different types of response are shown in table 4.3.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	35.9	*36.3*	46.7	*41.7*	17.5	*27.8*	57.5	*34.4*
Expression	35.0	*33.7*	37.5	*36.2*	27.5	*28.2*	58.3	*37.3*

Table 4.3. Evaluation of face matching for original and expression enhanced face pairs within ten item sequential arrays. The results are broken down by the percentage of hits, misses and false identifications for target present arrays, and the percentage of correct rejections for target absent arrays.

Accurate response rates were analysed using a repeated measures analysis of variance with target presence and image transformation as factors. There was a significant main effect of target presence, $F(1,19) = 7.6$, $p = .01$, $\eta_p^2 = .29$, but no significant main effect of expression transformation, $p > 1$, and no significant interaction between these factors, $p = .8$. Face matching performance was generally poor but was significantly better when the target was not in the array, suggesting that overall selection of images was reduced when the arrays were presented in a sequential format. The only evidence of affective influence on face matching was in the type of errors for target present arrays: when the images were manipulated to show more positive expressions there were significantly more false positive identifications, $t(19) = 2.4$, $p = .03$, $d = 0.4$, with a trend for fewer misses, $t(19) = 1.9$, $p = .07$, $d = .02$.

On the basis of Experiments 11 and 12 it was expected that sensitivity would be impaired and a more conservative response criterion would be adopted for the expression enhanced images, but paired sample comparisons showed that with sequential face matching arrays expression transformation did not affect *d*-prime measures of sensitivity, $p > 1$ *($d = -0.5$)*, or response criterion, $p = .8$ *($c = 0.7$).*

Discussion

Enhancing positive affect reduced sensitivity in discrimination of face pairs and in simultaneous comparison of multiple images; a similar effect was thus expected when multiple item arrays were presented sequentially. Criterion for reaching a face matching decision was not affected by the expression transformation when an array of faces was compared but was heightened when two images were compared; hence it was proposed that sequential presentation of a face matching array would also produce a more conservative response bias. Neither prediction was supported, and the overall levels of accuracy were not affected by the image transformation.

From the sequential face matching procedure two effects were notable: the rate of accurate face matching was low and as response criterion was more conservative than observed with the simultaneous arrays *(c = 0.7 v's c = 0.1),* this might suggest that a sequential procedure engenders a reluctance to make an image selection irrespective of facial affect. However, correct rejection of target absent arrays was comparable with the simultaneous format, indicating that it was more difficult to form accurate face matching decisions when images were evaluated in pairs, than when simultaneous presentation allowed multiple image comparisons. This difficulty was compounded when the expression transformation was applied; although the level of accurate face matching did not change, false identification from target present arrays increased providing evidence that where judgement is difficult, positive affect can influence decisions in a face matching task if one cannot view the images simultaneously. Within the sequential procedure the expression transformation did not enhance face matching performance but increased levels of inaccurate image selection.

Experiment 14

Positive affect in a delayed matching task

Facial expression has been shown to influence recognition judgements but the results of the perceptual discrimination experiments suggest that the effect does not operate at a purely perceptual level and that memory will also be involved. It is possible that positive expressions engage more attention in general face perception but that the expression transformation exaggerates image characteristics in a way that reduces correspondence between two face images of the same person. Bruce et al. (1999) demonstrated that if expression differed between images, face matching performance would be reduced and for this reason the expression manipulation was applied to both the target and the array items in these evaluations. Conflicting facial expressions are, however, common in live identity matching where a person will be matched against a photograph or to video footage. What is more, in this process the face and the face image are not simultaneously contrasted, but the practice will take the form of looking at the person and then at the image, and vice versa. Experiment 14 attempted to approximate this effect, and matching of inconsistent facial expressions was evaluated within a delayed face matching task. If different facial expressions reduce correspondence of the facial images, the number of face matching selections would be reduced, causing poorer rates of correct identification but increasing correct rejection of non-matching images. However, if iconic memory enables positive affect to enhance familiarity, the expression enhanced images may be selected more often.

Participants

Forty-six undergraduate psychology students at the University of Stirling participated in return for course credit. Their ages ranged from 20 to 54 years with a mean of 28.2 years (*s.d.* 10.7). All had normal or corrected to normal vision.

Materials and design

The delayed face matching task presented each target image followed by a two item array. The video stills were employed as target items and face matching arrays were generated using the photographs. Arrays were prepared for each target and displayed two items: in target present trials the photograph of the target was shown with the photograph of the most similar foil; in target absent trials the two most similar foils were shown. A repeated measures design employed target presence and expression transformation as factors. On each trial one of the array items would be shown with the enhanced positive expression, for half of the target present trials the photograph of the target was transformed, while for the remainder it was the foil image that was transformed and the target position was counterbalanced. The trials were presented in fully randomised order, and conditions were counterbalanced across participants.

Procedure

The experiment was conducted using E-Prime software *(apparatus as before).* Each pair of images was shown at the vertical mid-point of the screen, they were separated by a distance of 125 mm and the head in each image subtended a viewing angle of approximately 2.5º at a distance of 70 cm. On each trial a fixation cross was shown on the centre of a white screen for 3000 ms followed by presentation of the target image for 750 ms. There followed an inter stimulus interval of 1000 ms and then the two item array was shown until a response was provided. A diagram of the trial procedure is shown in Figure 4.4.

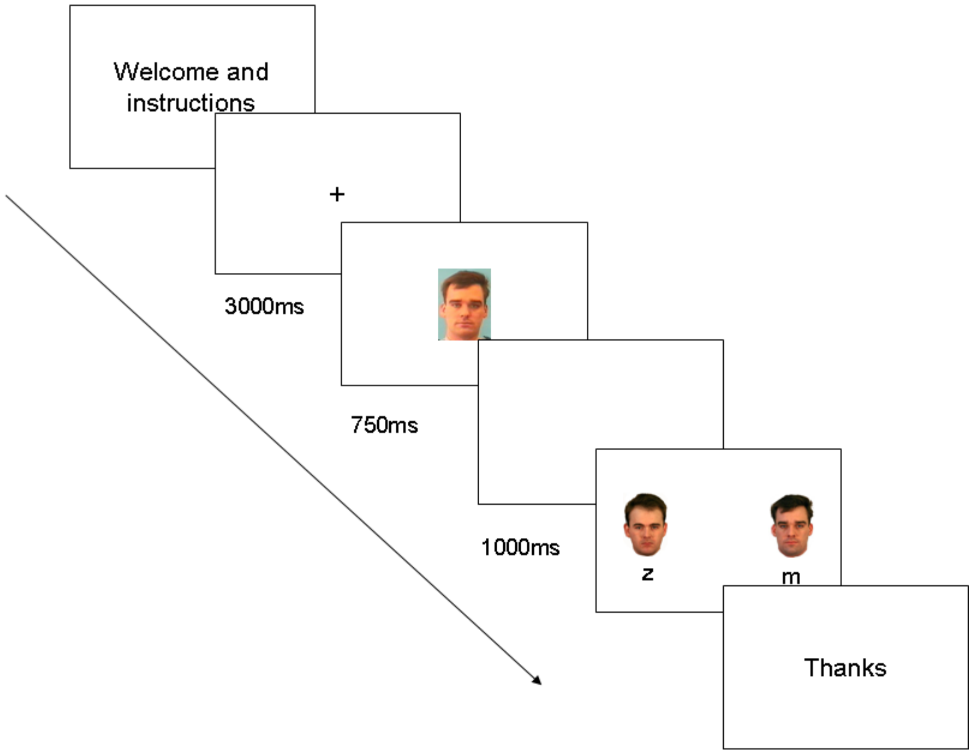


Figure 4.4. A diagram of the delayed face matching task employed in Experiment 14. The trials were activated manually by key press when the participant was ready to begin. For each trial a fixation cross was shown in the centre of a white screen for 3000ms, followed by a photograph of the target for 750ms, there followed and inter stimulus interval of 1000ms before the two item array was shown until a response was provided. To identify the image on the left as a match the participant typed 'z', to identify the image on the right the letter 'm' was entered. If no image was identified as a match the participants pressed the space bar and the next trial commenced.

Each participant was tested individually. They were instructed that they would complete a face matching task and that on each trial they would briefly be shown a target image, followed by two potential matches. The participants were cautioned that the target may not be represented by either image; if they believed that one of the images showed the target person they should press the appropriate key, but if they thought that neither image showed the target person they should press the space bar. Instructions were provided verbally and prompts to the appropriate key press were always in view on the computer screen.

Results

In the delayed face matching task the target absent pairs obtained a correct rejection rate of 59.7% (*s.e.* 4%) and when the target was present, the identification rate was 74.5% (*s.e.* 1.9%). For the target present trials there was no observable effect of the expression transformation, but when the target was not present, there were fewer false positives for the expression enhanced images. The means and standard deviations for different responses are shown in table 4.4.

	Target Present						Target Absent	
	Hits		Miss		False ID		False Positive	
	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	36.7	*7.9*	6.9	*6.3*	5.4	*6.3*	21.7	*15.5*
Expression enhanced	37.8	*7.0*	6.8	*6.0*	6.4	*6.2*	18.6	*13.3*
Total	74.5		13.7		11.8		40.3	

Table 4.4. Evaluation of unfamiliar face discrimination for original and expression enhanced faces within a delayed face matching task. The results are broken down by the percentage of hits, misses and false identifications for target present arrays, and the percentage of false positive identifications for the target absent arrays. Responses for target present trials show no effect of the expression transformation. In target absent arrays fewer false positives were observed for images that were transformed to show more positive affect.

A series of paired sample t-tests confirmed that when the target was represented the presence of positive affect did not influence the rates of correct identification, $p = .3$, false identification, $p = .2$, or misses, $p = .9$. The transformation of one of the images should have reduced correspondence between images of the same person, while the delayed procedure might have allowed positive affect to influence judgements from memory; but there was no detrimental effect to accurate face matching when the target was represented. When the target was not represented there was a marginally non-significant effect of expression on correct rejection of the target absent arrays, $t(45) = 1.9$, $p = .06$, $d = 0.2$. Contrary to the effects of the smiling face bias, expression enhanced images were less likely to be incorrectly identified than the original images. Signal detection measures of sensitivity and response criterion were also calculated and *t*-tests established that in the delayed matching task signal detection measures of sensitivity and response criterion were equivalent for the original and expression enhanced images($d' = 0.7$; $c = 0.7$).

.

Discussion

Discrimination of unfamiliar faces is difficult and any difference between the images can impair correspondence of matching items. In a simultaneous face matching task Bruce et al. (1999) demonstrated that face matching would be poorer if the target and array faces showed different expressions. However, positive expression can provoke false identification (Baudouin, Gilibert, Sansone & Tiberghien, 2000; Gallegos & Tranel, 2005; Kottoor, 1989; Davies & Milne, 1982; Endo et al. 1994; Lander & Metcalf, 2007). Conflicting facial expression is common in identity verification; thus it would be useful to better understand the way expression might induce facial identification errors.

A delayed face matching procedure was employed in which a target image was shown followed by two potential matches, one of which was enhanced to display a more positive expression. It was predicted that if different facial expressions caused matching images to appear less alike, fewer selections would be generated which would reduce accurate identification but would improve rejection of target absent arrays. If, however, affective information influenced identity verification judgements, a positive familiarity bias would cause expression enhanced faces to be matched to the target more often, with the consequence that false positive identification would increase.

There was no evidence of a positive familiarity bias in responses for the delayed face matching task. When the target was present, responses for original images and expression enhanced images were comparable, and when the target was not present, faces enhanced to show positive affect were less likely to be selected. The results indicate that the expression enhancement may have facilitated discrimination of non matching items but it did not produce enough disparity to impair correspondence of matching images. In the delayed matching procedure, expression transformation did not enhance face matching performance, but it did improve correct rejection when the target was not present in the array.

Summary and conclusions

A familiarity bias for positive facial expressions is consistently found in studies of face recognition (Baudouin, Gilibert, Sansone & Tiberghien, 2000; Gallegos & Tranel, 2005; Kottoor, 1989; Davies & Milne, 1982; Endo et al. 1994; Lander & Metcalf, 2007) and in Chapter 3 enhanced facial composite identification was associated with increased sensitivity for images that were manipulated to show positive affect. As there is no apparent reason why discrimination should be improved by facial expression, it was concluded that positive valence induced more attention with subsequent gains in recognition (see also Endo et al., 1992). Perceptual discrimination of unfamiliar faces is difficult and although there are no memory judgements, the aim of this chapter was to determine if positive affect could enhance attention and improve face matching performance in a range of face discrimination tasks, or indeed whether the positive familiarity bias would elicit more false positive identifications.

In Experiment 11 face pairs comprised a photograph and a video still, in either the original format or with the expressions enhanced. In contrast to the effects obtained for face recognition, positive affect reduced sensitivity and elicited a more conservative response bias. This was observed in reduced matching judgements such that accurate face matching declined while rejection of the foil images improved. Face matching with ten item simultaneous arrays was evaluated in Experiment 12. As before, sensitivity was reduced for expression enhanced images, but with comparison of multiple images, response bias did not change; consequently, accurate face matching and correct rejection of target absent arrays both reduced. Multiple item arrays were evaluated with sequential presentation in Experiment 13. Sensitivity, response criterion, and accuracy were not influenced by positive affect, but accuracy was poor in comparison with the simultaneous format: therefore, face matching was easier when simultaneous presentation enabled multiple image comparisons. In the target present arrays the types of error were influenced by expression. False identification significantly increased, while incorrect rejection of the arrays declined, but it is not clear why this effect was not also observed with the target absent arrays. Positive affect was shown to enhance sensitivity in identification judgements for facial composites and it was anticipated that this effect might also be obtained in facial discrimination tasks; but sensitivity and face matching was not enhanced, and when multiple images were compared the levels of inaccurate selection increased.

The final experiment employed a delayed face matching procedure in which presentation of the target preceded a two item array in which one of the faces showed enhanced positive expression. In half of the target present trials the correct match was transformed, for the remainder it was the foil image that was transformed. Within this procedure sensitivity and response criterion were not affected and there was no evidence of affective bias. While positive expression did influence false positives from target absent arrays, the effect was contrary to predictions: the images with enhanced expressions were less likely to be selected, suggesting that expression enhancement may have facilitated discrimination of non-matching items.

These results provide evidence that although positive expression can enhance perception of familiarity in face recognition, in perceptual discrimination tasks, sensitivity will not be enhanced and accurate face matching judgements can be significantly impaired. While rejection of target absent arrays was improved in the delayed matching procedure, the most likely explanation is that the expression enhancement increased perceptual disparity of the target and foil images.

Theoretical considerations

In face recognition it has been shown that positive expression can enhance perceptions of familiarity (Baudouin, Gilibert, Sansone & Tiberghien, 2000; Gallegos & Tranel, 2005; Kottoor, 1989; Davies & Milne, 1982; Endo et al. 1994; Ellis & Young, 1990). In Chapter 3 it was found that when facial composites were manipulated to show more positive expressions, enhanced identification was associated with increased *d*-prime measures of sensitivity. This finding and evidence that smiling faces engage attention for longer (Endo et al., 1992) prompted speculation that positive affect might enhance attention and that inducing positive affect could be an effective way to facilitate differentiation of unfamiliar faces. The findings presented in this chapter provide evidence that while positive affect may induce attention in an attempt to remember a face, the effect is not found in perceptual discrimination tasks. The results indicate that the smiling face advantage in face recognition is implicated at the point of memory retrieval, and is a function of enhanced identification of a face representation from memory, or from the erroneous impression that this has been achieved, rather than just ensuring that the image is regarded for longer.

Applied considerations

Facial image comparison and identity verification from face images is common practice in security settings, yet there has been little formal evaluation of human face matching abilities, or factors that will impede, or enhance performance. This research explored the possibility that positive affect could be manipulated to induce attention and enhance face discrimination, and while the manipulation was ineffective, the results have important consequences for applied face image comparison. Prior to this work it was known that facial expression can significantly impair face matching performance if the expressions on the target face and potential matches are different (Bruce et al., 1999). Using the same face matching materials this knowledge may now be extended to show that these effects can be ameliorated if a delayed, or consecutive matching procedure is employed. In Experiment 14 it was found that if the target and the array items are not regarded simultaneously subtle positive expression will not impair accurate face matching and might enhance successful discrimination of different identities. This finding is particularly encouraging, as in identity verification from documents, such as driving licences or passports, facial expression will often be different; as the face and the image cannot be scanned simultaneously this result suggests that expression disparity may be less detrimental than previously believed. This experiment should now be replicated with different image sets to evaluate different expressions and stronger affect.

Experiments 11 – 13 evaluated face matching when both the target and the array items were manipulated to show subtle positive affect. While it might be expected that image correspondence should not be impaired if the target and array expressions match, each evaluation indicated that discrimination was superior when the images were shown with the original neutral expressions. The expression transformation may have exaggerated image properties and increased disparity between matching items, but in Experiments 12 and 13 incorrect face matching selections increased indicating that this is unlikely to completely explain the effect. The alternative is that the positive expressions may have detracted from structural comparison of the face images with the consequence that discrimination was impaired. These results suggest that for security purposes, facial image comparison and formal identification should be determined from neutral facial expressions.

Conclusions

Smiling facial expressions enhance identification judgements when recognition is difficult (Baudouin, Gilibert, Sansone & Tiberghien, 2000; Gallegos & Tranel, 2005; Kottoor, 1989; Davies & Milne, 1982; Endo et al. 1994; Ellis & Young, 1990). This chapter to explored whether positive affect could also be manipulated to induce attention and enhance face discrimination. The results provide evidence that the positive identification bias is not evident at a purely perceptual level and that judgements regarding retrieval of a face memory must also be involved. Matching of face images that were enhanced to portray positive affect showed reduced performance in comparison with original items, indicating that positive facial expression may detract from structural face image comparison and that formal identification protocols should, where possible, incorporate expression free images. The materials employed here were developed by Bruce et al. (1999); the authors assessed the impact of expression change within simultaneous arrays, demonstrating that face matching will be impaired if the facial expressions of a target and array items are different. This result was extended to show that if images with differing expressions are matched in a delayed or consecutive format, accuracy may not be impaired and discrimination of non-matching items might be enhanced. Facial image comparison is a common security practice yet relatively little is known about factors that will influence performance. Future study should attempt to replicate the effects reported here with different image sets and with different expressions of varying strength.

5

Increasing distinctiveness in unfamiliar face discrimination

Matching faces may seem like a trivial task but when the faces are unfamiliar, even high quality images produce poor results. This chapter continues to evaluate unfamiliar face discrimination, and explores an imaging method intended to increase facial distinctiveness. Valentine and Bruce (1986a; 1986b) found a consistent recognition advantage for faces that were distinctive. These faces also took longer to classify as faces; therefore, they were less like other faces, but they were also less like a prototypical face. Facial caricatures exaggerate facial features that are useful for identification, thereby making the faces less average and enhancing their distinctiveness: caricatures have also been shown to confer a recognition advantage (e.g. Tanaka & Simon, 1996). In a face matching task exaggerating the differences between faces should also make it easier to tell them apart; this chapter examines the use of caricature as a means to increase facial distinctiveness and enhance unfamiliar face discrimination.

Distinctiveness and caricature

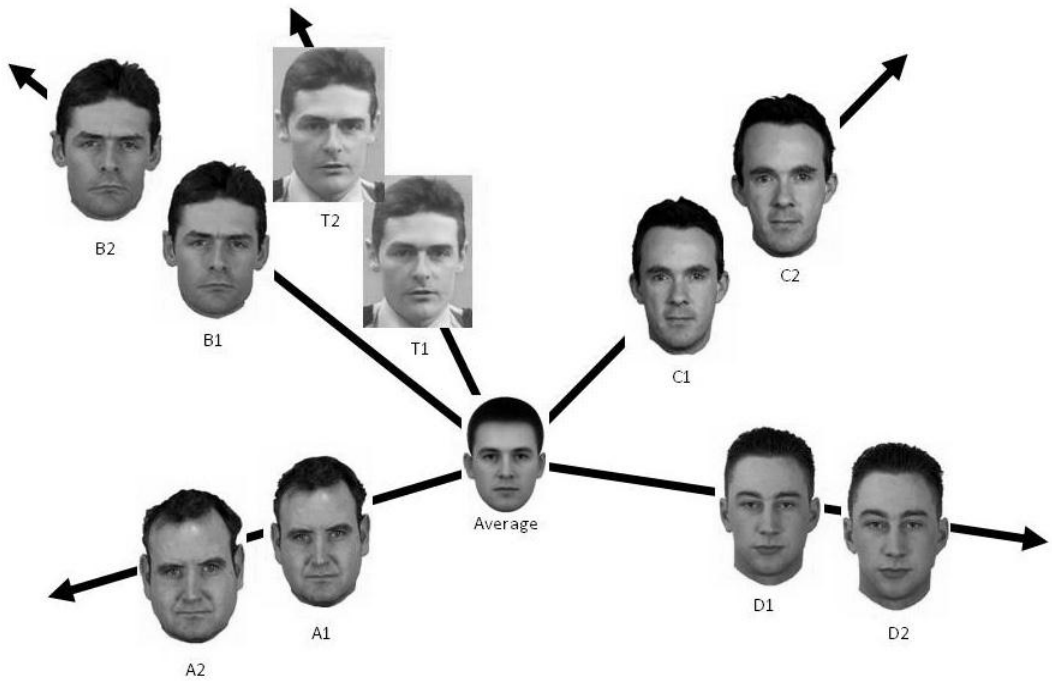
The art of caricature can be traced to ancient Greece: depictions are generally cartoon-like but despite gross exaggeration, they are highly identifiable because they emphasize distinctive characteristics. Agreement is, generally, good about what features should be parodied (Goldman & Hagen, 1978) and famous caricatures can become iconic; for example, a woman with wild curly hair and a huge smile is Julia Roberts, or a man with small eyes and a prominent cleft chin is John Travolta. For familiar faces, artists' caricatures can be identified faster than accurate line-drawings (Stevenage, 1995), producing results that are similar to those found with distinctive faces (Bartlett et al., 1984; Goldstein & Chance, 1980; Light et al., 1979; Shepherd et al., 1991; Valentine & Bruce, 1986; Winograd, 1981). This suggests that caricature might systematically be employed to increase the distinctiveness for unfamiliar faces.

Caricatures can be produced with imaging software by tagging the landmarks of a face image with data points, which are then warped to exaggerate each point relative to a norm face. This will make distinctive features more salient in comparison with typical characteristics, and computer generated caricatures show a reliable recognition advantage (e.g. Benson & Perrett, 1994; Brennan, 1985). For line-based images, fairly pronounced levels of manipulation are estimated to produce the best likeness (e.g. 50% - Benson & Perrett, 1994; 16% - Rhodes et al., 1987); while photographs are judged to be a better likeness if small levels of caricature are employed (e.g. 4% - Benson & Perrett, 1991; 6% - Ellis, 1990). These findings indicate that computerised caricature transformations are easily perceived and can reliably influence facial identification judgements.

Caricature and Multi-Dimensional Face Space

Caricature effects are most commonly explained with reference to multi-dimensional face space models (Valentine, 1991). Within a MDFS each face is encoded along an unspecified number of dimensions, and each contributes to the face memory structure. Face similarity is determined by distance and density: typical faces are perceptually similar, and are encoded in a cluster around the origin of the space, while faces that are more distinctive occupy sparser positions further away from the central tendency. Within a MDFS any characteristic that causes a face to deviate from the norm is considered to be distinctive, and because the face will be positioned further away from the origin and from typical exemplars, recognition and discrimination will be promoted (Valentine & Bruce, 1986).

In terms of MDFS, a computer-generated caricature exaggerates the difference between a face image and an average face, which increases the distance between the face image and the centre of the space. This separates the manipulated face from surrounding faces, making it easier to discriminate and identify (Valentine, 1991; Valentine & Endo, 1992). The MDFS model would predict that if each face in a matching array is caricatured relative to a norm face, the trajectory of each from the centre of the face space would be maintained, but the distance between each face, and from the origin, would become greater, making the faces appear more dissimilar. An illustration of the effect of caricature on face matching within a MDFS is shown in figure 5.1.



5.1. A 2-dimensional representation of MDFS and how caricaturing might increase distinctiveness to enhance perceptual face matching. An average face image is shown at the origin of the face space[2]; around the origin, the original array items A1, B1, C1, and D1 are perceptually similar to the target face T1. Each face differs from the average face in different ways, and caricaturing the target and array items relative to the average face exaggerates these differences by pushing them all further from the origin, and further from each other (A2, B2, C2, D2 and T2). Within MDFS increasing the distance between each face should reduce perceptual similarity and improve discrimination, thereby reducing false positive matches. For correct face matching the caricatured target (T2) and the caricatured correct match (B2) are further apart than the original target (T1) and the original correct match (B1), but the distance between the caricatured target (T2) and the other caricatured array items has also increased and may thus enhance discrimination of the correct matching face. N.B. The directions shown are arbitrary, A and D are not in any sense near opposites within the MDFS.

---

[2] Within a norm based MDFS model faces are encoded with reference to a norm face, while within an exemplar based MDFS model faces are encoded relative only to other face exemplars. Figure 5.1 is intended to describe the effect of caricature within any MDFS and is not intended to comment on these models or to selectively represent a norm based account.

Aim of study

Sensitivity to unfamiliar faces is generally poor (Kemp et al, 1990; O'Donnell & Bruce, 2001) and face images that are not distinctive will be difficult to discriminate (e.g. Bruce et al. 1999). A caricature technique that systematically increases distinctiveness could enhance discrimination, and it might be possible to generate standardised imaging techniques for safer facial identification procedures. The aim of this chapter was to evaluate whether a systematic caricature transform could increase distinctiveness and improve accuracy in unfamiliar face matching tasks.

Increasing distinctiveness with caricature should make faces more dissimilar and reduce false positive identification, but the effects on accurate identification are more difficult to predict. Unfamiliar faces have no stored memory representation; therefore, each unfamiliar face must be interpreted on the basis of structural codes, in relation to existing face representations. In unfamiliar face matching each image is unique; therefore, the target will occupy a different face space position to the matching image, and their proximity will depend considerably on image properties (e.g. Adini et al., 1994). Caricaturing the images will shift the target face and the array faces away from the centre of the face space, and from each other. If this causes the distance between the caricatured target and the caricatured match to be smaller than the distance between the caricatured target and caricatured foils, identification will improve; but if caricature increases the distance between matching faces too much, perceptual similarity, and performance, will decline. Three levels of caricature were assessed with simultaneous matching arrays in Experiments 15 to 17 and in a sequential matching task in Experiment 18.

Creating distinctiveness with caricature

This series of experiments employed the arrays developed by Bruce et al. (1999). To create caricatured arrays, the templates of each image created with Psychomorph *(for details see Chapter 4)* (Tiddeman, Burt, & Perrett, 2001) were used to compute the average location of each data point for all 120 photographs. The resulting average template was then used as the reference, or norm face, for all caricature transformations. Caricatured images were created by using the Psychomorph templates to warp the shape of each individual image 30%, 50%, and 70%, away from the shape of the norm face. Examples of the Psychomorph template, the average face image, and original and caricatured images are shown in Figure 5.2.

Experiment 15

Face matching with 30% caricatures

Experiment 15 evaluated whether a subtle 30% caricature manipulation could increase distinctiveness and discrimination within simultaneous face matching arrays. It was predicted that caricaturing would reduce false positive identification; the effects on correct identification were exploratory. If the distance between the caricatured target and the caricatured match was smaller than the distance from the caricatured foils, identification might improve; but if the distance between matching faces was too great, accurate matches will decline.

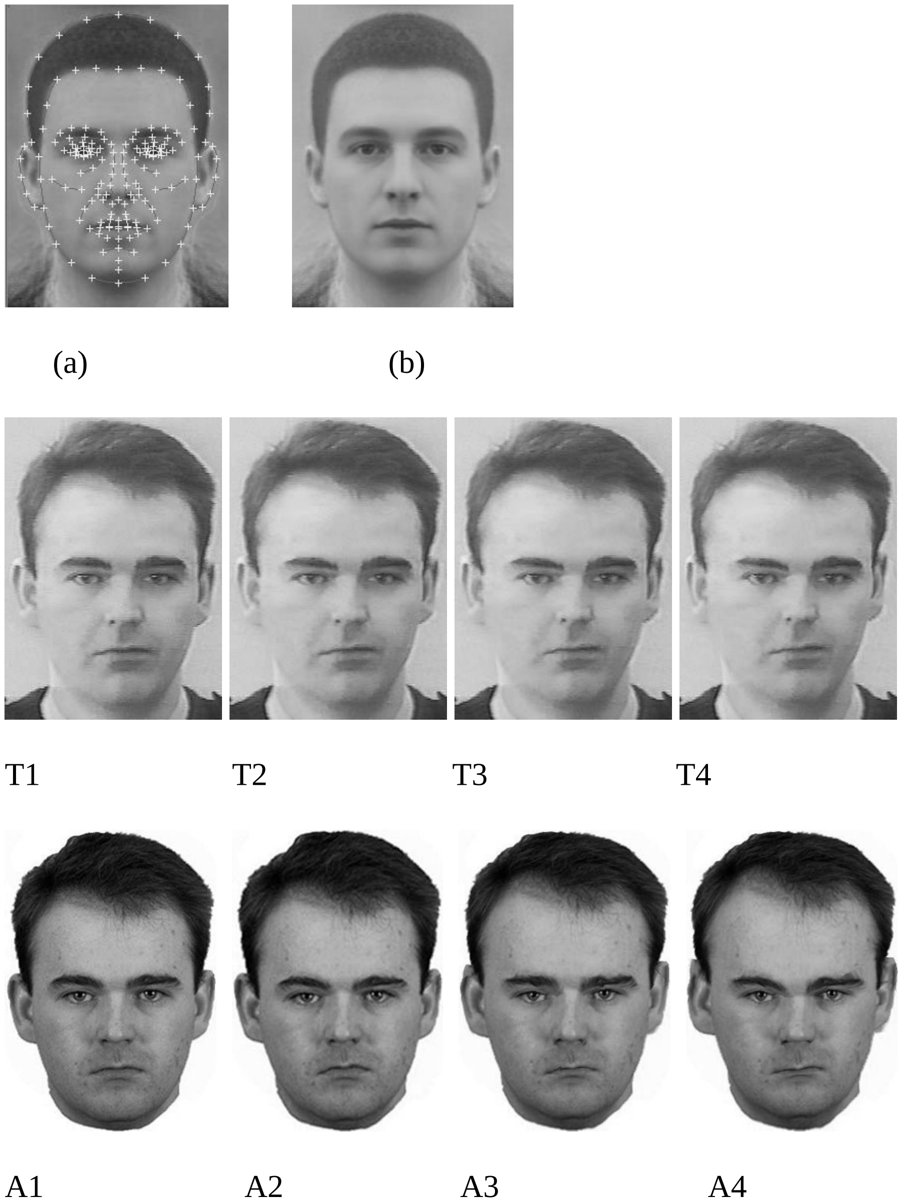


Figure 5.2. Caricatured images were generated using Psychomorph imaging softeware (Tiddeman et al., 2001). Examples of (a) the Psychomorph template and (b) the average face image are shown above. The norm face is the average of 120 individual templates. Items T1-T4 show a target image as the shape is caricatured from (T1) veridical, through levels of (T2) 30%, (T3) 50%, (T4) 70%. Items A1-A4 show the matching array item at each level: (A1) original face, (A2) 30% caricature, (A3) 50% caricature, (A4) 70% caricature.

Participants

Twenty-four students from the University of Stirling participated in return for course credit. Seven participants were male, and ages ranged from 19 to 54 years with a mean of 26.5 (*s.d.* 10.7). All had normal or corrected to normal vision.

Materials

This experiment employed a subset of the 32 most difficult arrays from Bruce et al. (1999). Trials consisted of a video still target image above the 10 photographs paired most often in a similarity matrix (target absent condition), or above the 9 most similar images and a photograph of the target (target present condition). Arrays were arranged in two rows and numbered 1-10. Target position was randomly sampled with the constraint that each position was employed at least three times and none more than four. The target images were cropped to show head and shoulders, and measured 50mm x 80mm; the array images were cropped to remove clothing and each head gave a viewing angle of approximately 2.5° at 70 cm viewing range. The array with target image measured 260 mm x 270 mm. Within caricatured arrays the target and arrays were caricatured by 30% from the norm face. Examples of original and caricatured arrays are shown in Figure 5.3 and Figure 5.4, respectively.

Design

A 2 x 2 repeated measure design was employed with target presence, and image type (original; caricatured) as factors. All arrays were employed for each participant; half portrayed caricatured images, half the original images, and the target was present in half of each set. All conditions were fully counterbalanced across participants, and the trial presentation order was randomised for each participant.

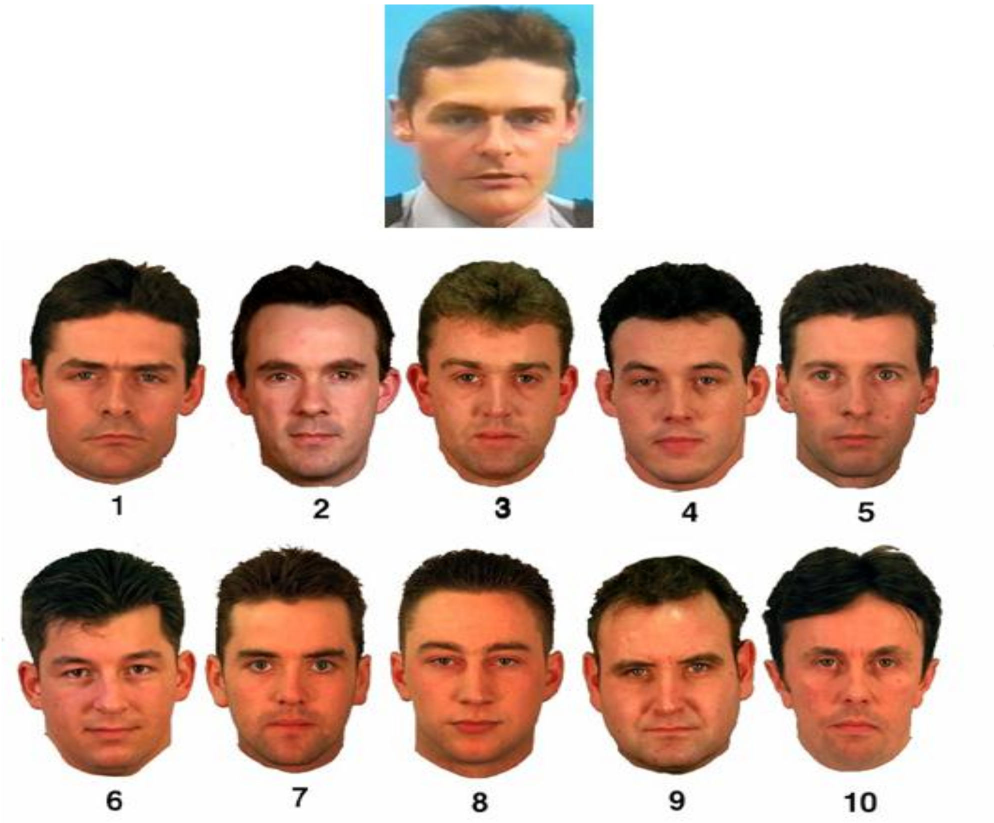


Figure 5.3. An example of a simultaneous face matching array comprising original faces. The target, or probe image shows a video still. The array shows photographs of 10 Caucasian males with neutral expressions. In target present arrays array position of the correct match was equally sampled.

Procedure

Each participant was tested individually and the experiment was conducted using E-Prime software with a 17 inch LCD monitor at 1024 x 768 pixels resolution. The responses were recorded via the keyboard. Instructions were provided both verbally and with on screen commands. The participants were asked to complete a face matching task and informed that the target may or may not be present in each line-up array. To identify an array item as a match they were to type the corresponding number, if no match was identified they were instructed to press the space bar.

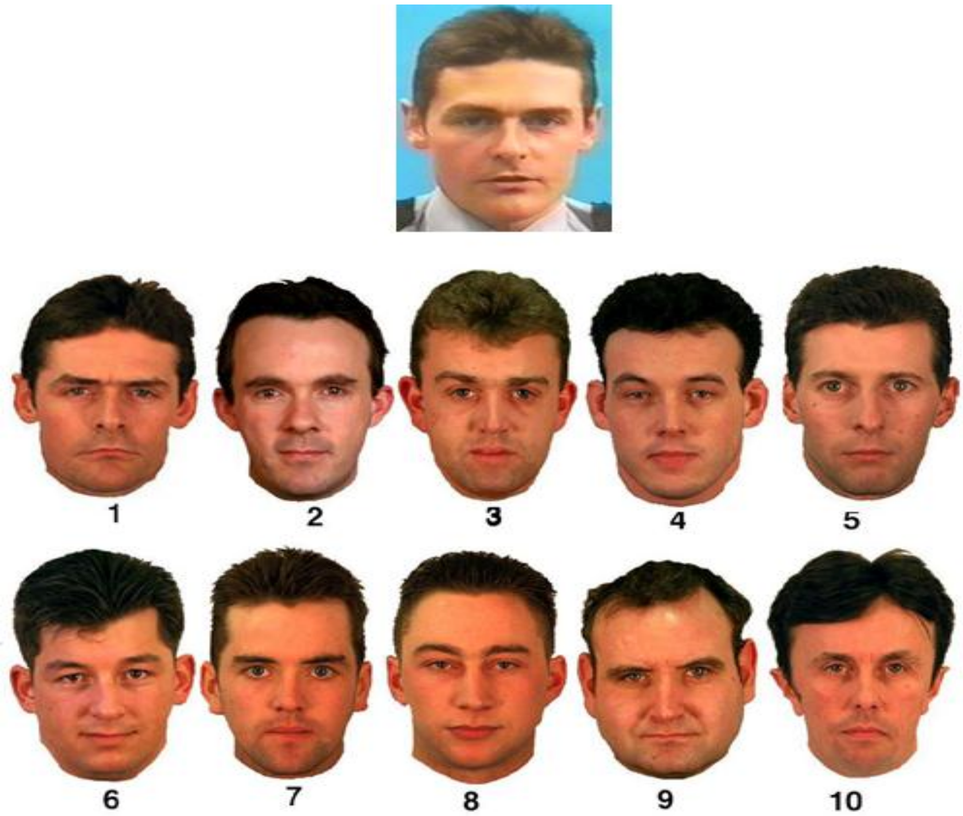


Figure 5.4. An example of a simultaneous face matching array comprising 30% caricatured faces. The target, or probe image shows a video still; the array shows 10 Caucasian males with neutral expressions. Both target and array items were caricatured by 30% from the norm face.

Results

Caricature applied at the 30% level did not influence target present accuracy, but when the target was not present false positives were substantially reduced. The means and standard deviations are shown in Table 5.1.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	64.1	*23*	19.8	*23*	16.1	*15.8*	56.8	*21.8*
30% Caricature	64.6	*25*	20.3	*19.1*	15.1	*15.6*	73.4	*21.6*

Table 5.1. Evaluation of face matching for original and 30% caricatured faces within simultaneous arrays. Results are broken down by the percentage of hits, misses and false identification for target present arrays, and the percentage of correct rejections for target absent arrays.

Correct face matching scores were analysed using a repeated measures analysis of variance, with factors of target presence (target present; target absent), and image type (original; 30% caricature). This showed no significant main effect of target presence, $p = .9$, but there was a significant main effect of caricature, $F(1,23) = 5.1$, $p = .03$, $\eta_p^2 = .18$, and there was a significant interaction between caricature and target presence, $F(1,23) = 5.5$, $p = .03$, $\eta_p^2 = .19$. Planned comparisons confirmed that when the target was not present caricature produced significantly more correct rejections, $t(23) = 3.0$, $p = .01$, $d = 0.8$, but there was no effect on correct identification from the target present arrays, $p = .9$.

To assess whether this effect was consistent across the full set of targets and arrays, an analysis of variance by items was also performed. There was no significant main effect of target presence, $p = .9$, but there was a significant main effect of caricature, $F(1,31) = 12.0$, $p = .002$, $\eta_p^2 = .28$, and a significant interaction between target presence and caricature, $F(1,31) = 4.6$, $p = .04$, $\eta_p^2 = .13$. Planned comparisons confirmed no effect of caricature on correct face matching, $p = .9$, but when the target was not present, 30% caricature reduced false positive identification across the full set of arrays, $t(31) = 3.8$, $p = .003$, $d = 0.7$.

It was anticipated that caricature would enhance discrimination; therefore, signal detection analysis of hits and false positives was conducted to obtain *d'* prime measures of sensitivity, and response bias criterion *c*. Sensitivity was better for caricatured images (*d'* = 1.43 *v's d'* = 0.67), *t*(23) = 2.2, *p* = . 04, *d* = 0.5, and response bias was also more conservative (*c* = 0.17 *v's c* = -0.14), *t*(23) = 2.8, *p* = .01, *d* = 0.6, indicating that 30% caricature enhanced discrimination but made the participants more cautious.

Discussion

When the target was not present caricature enhanced discrimination, but when the target was present, correct identification remained at the same level. Within the face space metaphor, 30% caricature shifted foil images far enough away from each other to improve discrimination, but the distance between the target and the correct match, remained small enough to avoid detriment to accurate matching.

In the facial caricature literature, a transformation of 30% is perceived as too large for photographs of familiar faces (e.g. Benson & Perrett, 1991); in contrast, the visible effect on unfamiliar faces was barely discernible. This is probably because we are extremely sensitive to the facial properties of familiar faces (e.g. Haig, 1984); however, Lee and Perrett (1997) found that if familiar images were presented for very short durations (33ms), a caricature advantage emerged at much stronger levels (50%). This suggests that the caricature advantage may depend on some level of doubt or difficulty, not generally found with familiar face recognition. Given that discrimination of unfamiliar faces is difficult, stronger levels of manipulation might be even more effective. Accordingly, Experiment 16 evaluated unfamiliar face matching with a more pronounced caricature manipulation.

Experiment 16

Face matching with 50% caricatures

Experiment 16 replicated the design of the previous experiment but explored the effects of a stronger 50% caricature manipulation. It was also important to assess whether a caricature advantage would be found with a more variable set of face matching materials, and to this end the full set of 80 arrays was employed (Bruce et al., 1999). It was predicted that caricature would reduce false positive identification but might also impair correct identification of matching faces.

Participants

Forty students from the University of Stirling participated in return for course credit. Twenty-four were male and ages ranged from 18 to 46 years with a mean age of 20 years (s.d. 6.2). All had normal or corrected to normal vision.

Materials

Caricatured images were produced from the complete set of 80 arrays (Bruce et al., 1999) by warping the shape information of each image 50% away from the norm face. In target present arrays target position was randomly sampled with the constraint that each position was employed four times for the original arrays and four times for the caricatured arrays.

Design and Procedure

All details were as per Experiment 15 with the exception that when half of the trials had been completed, a rest break was offered and the experiment was resumed by key press at the participant's instigation.

Results

The means and standard deviations are shown in Table 5.2 and are broken down by type of response. When caricature was applied at the 50% level correct rejection of the target absent arrays was again enhanced.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	70.5	*19*	19.4	*15.7*	10.1	*8.7*	72.8	*18.9*
50% caricature	71.8	*17*	20.1	*13.9*	8.1	*8*	78.5	*17.5*

Table 5.2. Evaluation of face matching for original and 50% caricatured faces within simultaneous arrays. Results are broken down by the percentage of hits, misses and false identification for target present arrays, and the percentage of correct rejections for target absent arrays.

A 2 x 2 repeated measures analysis of variance of correct responses revealed no significant main effect of target presence, p = .2. There was a significant main effect of image type $F(1,39) = 5.6$, $p = .02$, $\eta_p^2 = .13$, but no significant interaction between target presence and image type, p = .2. Across participants performance was significantly better when the arrays were caricatured by 50%.

The analysis of variance by target items found a significant main effect of target presence, $F(1,79) = 4.3$, $p = .04$, $\eta_p^2 = .05$, and a marginally non-significant main effect of image type, $F(1,79) = 3.6$, $p = .06$, $\eta_p^2 = .04$, but no significant interaction between target presence and image type, p = .2. Across target items accuracy was better for target absent arrays than for target present arrays, and caricatured arrays produced more accurate responses.

The prediction that caricaturing would enhance discrimination was again investigated using signal detection analysis of hits and false positives. Sensitivity was a little better for caricatured images (d' = 1.70) than for original images (d'=1.45), $t(39)$ = 1.8, p = .04, d = 0.2 *(one-tailed),* while response criterion also tended to be more conservative for the caricatures, (c=.19) than original images (c = .06), $t(39)$ = 1.7, p = .05, d = 0.3 *(one-tailed)*.

Discussion

The 50% caricature transformation increased overall face matching performance and the advantage for caricatured arrays held across the set of target items but failed to reach statistical significance (p = .06). Results for the original materials are remarkably similar to those obtained by Bruce et al. (1999) with the same arrays, and whilst accuracy is high in comparison with Experiment 15, it should be noted that the first experiment employed only a subset of the most difficult arrays. Sensitivity and response criterion were altered by 50% caricature but the effects were smaller than found with 30% caricature.

Within the face space metaphor, application of 30% or 50% caricature moved the array faces further away from the centre of face space and away from each other, improving discrimination and correct rejection of target absent arrays. With regard to correct identification, at 30% caricature it was concluded that increased distance between the target and correct match was small enough to avoid any face matching detriment; however, with a stronger 50% caricature, and greater distance between the target and correct match, a caricature advantage was observed. This could indicate that stronger caricature reduced similarity between array items and enabled the match to be discriminated, but it is possible that the effect reflects difficulty across the set of array items. When the faces are pushed further apart it may just become easier to identify the items that are less similar.

Post Hoc analyses

To investigate the effect of difficulty, a repeated measures analysis of variance on accuracy for the 32 most difficult arrays revealed a significant main effect of caricature, $F(1,40) = 4.4$, $p = .04$, $\eta_p^2 = .10$, but no significant main effect of target presence, $p = .3$, and no significant interaction between target presence and caricature, $p = .2$. In line with the full data set, correct identification and correct rejection of target absent arrays was significantly better with arrays that were caricatured by 50%.

Across target items the effects for difficult arrays were more consistent with Experiment 15: there was no significant main effect of target presence, $p = .3$, and no significant main effect of caricature, $p = .3$, but there was a significant interaction between these factors, $F(1,31) = 4.2$, $p = .05$, $\eta_p^2 = .12$. Paired sample t-tests found that when the arrays were more difficult, 50% caricature significantly improved correct rejection of target absent arrays, $t(31) = 2.0$, $p = .05$, $d = 0.4$, but did not enhance correct identification, $p = .7$, suggesting that although caricature can enhance matching of easier items, it will not consistently do so for more difficult ones. Importantly, at both 30% and 50% levels of transformation, caricature has the power to enhance unfamiliar face discrimination with no observed detriment to accurate identification.

While 30% and 50% caricature both improved sensitivity and accuracy, the effects were weaker at 50%, and confined to target absent arrays for difficult items. This might suggest that a weaker caricature is more beneficial, but to determine whether 30% caricature can enhance easier matches, it was necessary to repeat the experiment with the full set of arrays. The prediction that strong caricature will make matching faces too dissimilar, and thus impair identification, would also be tested.

Experiment 17

Evaluation of caricature levels

Experiment 17 employed a split factor design to replicate Experiment 16 with caricature levels of 30% and 70%. This allowed all of the caricature levels to be interpreted within a mixed factor analysis of variance. An interaction was predicted, such that correct rejection of target absent arrays would be enhanced at each level, but correct matches would be impaired at the strongest level of transformation.

Participants

Fifty-four students from the University of Stirling participated in return for course credit. Fourteen participants were male, and ages ranged from 17 to 68 years with a mean of 24.1 (*s.d.* 11.2). All had normal or corrected to normal vision.

Method

A mixed factor design was employed with level of caricature as the between participant factor and within participant factors of target presence and image type (original and caricatured). Mixed factor analyses of variance would explore the effects of discrimination at all three levels of caricature across the full set of arrays. The procedure and all other details are a precise replication of Experiment 16.

Results

Accuracy was improved for target absent and target present arrays when the images were caricatured by 30%. When the arrays were caricatured by 70% correct rejection of target absent arrays was improved but identification of matching images was substantially reduced. The means and standard deviations by caricature level and condition are shown with the results of Experiment 16 in Table 5.3.

	Hits		Target Present Miss		False ID		Target Absent Correct	
30%	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	65.6	*16.3*	20.5	*9.5*	13.9	*14.4*	73.3	*15.4*
Caricature	69.5	*15.3*	22.8	*9.2*	7.7	*12.4*	75.6	*19.0*
50%	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	70.5	*19.0*	19.4	*15.7*	10.1	*8.7*	72.8	*18.9*
Caricature	71.8	*17.0*	20.1	*13.9*	8.1	*8.0*	78.5	*17.5*
70%	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	66.1	*17.0*	11.4	*10.7*	22.5	*16.2*	62.7	*34.1*
Caricature	60.2	*17.3*	19.8	*12.7*	20.0	*16.3*	70.7	*29.5*

Table 5.3. Evaluation of face matching for original faces, and face caricatured by 30%, 50%, or 70%, within simultaneous arrays. The results are broken down by the percentages of Hits, Misses and False identification for target present arrays, and percentages of correct rejections for target absent arrays.

Analysis by participants

A 3 x 2 x 2 mixed factor analysis of variance of correct responses was conducted with level of caricature (30%, 50%, 70%) as the between participant factor, and target presence (target absent, target present) and image type (original faces, caricatured faces) as within participant factors. The results showed a significant main effect of target presence, $F(1,91) = 4.6$, $p = .04$, $\eta_p^2 = .05$, and a significant main effect of image transformation, $F(1,91) = 5.7$, $p = .02$, $\eta_p^2 = .06$, but no significant main effect of caricature level, $p = .1$. There was no significant interaction between target presence and level of caricature, $p = .8$, or between image type and level of caricature, $p = .6$, but there was a significant two way interaction between target presence and image type, $F(1,91) = 7.9$, $p = .006$, $\eta_p^2 = .08$, and there was a significant three way interaction between target presence, image type, and level of transformation, $F(2,91) = 4.4$, $p = .02$, $\eta_p^2 = .09$.

Planned comparisons revealed that correct rejection of target absent arrays and correct identification of matching faces were not enhanced by 30% caricature, $p = .3$ and $p = .1$ respectively. When the arrays were caricatured by 50%, correct rejection of target absent arrays improved, $t(39) = 2.9$, $p = .007$, $d = 0.3$, but identification of matching faces did not, $p = .6$. In line with predictions, caricaturing by 70% improved rejection of target absent arrays, $t(21) = 2.4$, $p = .03$, $d = 0.3$, but impaired identification of matching faces, $t(21) = 2.2$, $p = .04$, $d = 0.3$.

One way analyses of variance found no effect of the level of caricature on correct rejection of caricatured arrays, $p = .4$, but there was a significant effect of caricature level on identification of the caricatured matches, $F(2,93) = 3.6$, $p = .03$, $\eta_p^2 = .07$. Post hoc Tukey HSD tests showed no significant difference between matches caricatured by 30% and by 50%, $p = .8$, or between images caricatured by

30% and 70%, *p* = .1, but identification of images caricatured by 70% was significantly poorer than of images caricatured by 50%, *p* = .03, *d* = 0.7.

In summary, there was no effect of 30% caricature with the larger array set. Both 50% and 70% levels of caricature enhanced rejection of target absent arrays but 70% caricature also significantly reduced levels of correct identification.

Analysis by items

The analysis of variance by items revealed a significant main effects of target presence, $F(1,237) = 6.9$, $p = .009$, $\eta_p^2 = .03$, of image transformation, $F(1,237) = 5.4$, $p = .02$, $\eta_p^2 = .02$, and of caricature level, $F(1,237) = 9.1$, $p < .001$, $\eta_p^2 = .07$. There was no significant interaction between target presence and level of caricature, $p = .8$, or between image type and level of caricature, $p = .4$, but there was a significant two way interaction between target presence and image type, $F(1,237) = 5.4$, $p = .02$, $\eta_p^2 = .02$, and there was a significant three way interaction between target presence, image type, and level of transformation, $F(2,237) = 3.9$, $p = .02$, $\eta_p^2 = .03$.

Planned comparisons confirmed that correct rejection of target absent arrays was not enhanced by 30% caricature, $p = .3$, but identification of matching faces was somewhat improved, $t(79) = 1.7$, $p = .09$, $d = 0.2$. When the arrays were caricatured by 50% rejection of target absent arrays improved, $t(79) = 2.2$, $p = .03$, $d = 0.3$, but identification of matching faces did not, $p = .6$. Caricaturing by 70% also improved rejection of target absent arrays, $t(79) = 2.4$, $p = .02$, $d = 0.3$, but significantly impaired identification of matching faces, $t(21) = 2.1$, $p = .04$, $d = 0.2$.

One way analyses of variance also found a significant effect of caricature level on correct rejection of target absent arrays, $F(2,237) = 3.0$, $p = .05$, $\eta_p^2 = .02$, and on identification of caricatured matches, $F(2,237) = 5.6$, $p = .004$, $\eta_p^2 = .04$. Post hoc Tukey HSD tests showed that for correct rejection the 30% group did not differ from either the 50% group, $p = .5$, or the 70% group, $p = .4$, but performance for 70% caricature was much poorer than for the 50% caricature, $p = .04$, $d = 0.5$. Correct identification of caricatured images also did not differ between the 30% and 50% groups, $p = .9$, but identification of images transformed by 70% was significantly poorer than identification of 30% caricatures, $p = .02$, $d = 0.2$, and of 50% caricatures, $p = .007$, $d = 0.3$.

Across the full set of items correct rejection of target absent arrays was not influenced by 30% caricature but was enhanced by 50% and 70% caricature, with the 50% caricature producing significantly better results. For correct matches, no advantage was observed for 50% caricature, but there was a non-significant trend for 30% caricatures to be identified better. As predicted, 70% caricature significantly impaired accurate face matching.

Signal detection analyses

Discrimination of the faces was again investigated with signal detection measures of sensitivity and bias. A mixed factor analysis of variance on d' scores with image type (original and caricature) and level of caricature (30%, 50%, 70%) as factors, revealed a significant main effect of image type, $F(2, 91) = 4.8$, $p = .03$, $\eta_p^2 = .05$, but no significant main effect of caricature level, $p = .2$, and no significant interaction between these factors, $p = 1$. Sensitivity was better for the caricatured arrays than for the original arrays ($d' = 1.44$ *v's* $d' = 1.21$). Analysis of response bias also found a significant main effect of image type, $F(1, 91) = 8.7$, $p = .004$, $\eta_p^2 = .09$, but no significant main effect of caricature level, $p = .9$, and no significant interaction, $p > .1$; therefore, response criterion was consistently more conservative for caricatures than for the original images ($c = 0.23$ *v's* $c = 0.09$).

In summary, although sensitivity improved at all caricature levels, the 30% caricature transformation produced no significant effect on accuracy. Consistent with MDFS predictions, the 70% caricature transformation significantly impaired identification of matching images, and while levels of correct rejection were enhanced, performance was much poorer than those obtained with the 50% caricature. It is notable that the 30% caricature produced no effect when evaluated with the larger more variable array set but produced a significant advantage with the subset of difficult items. To explore the effect of difficulty at all caricature levels a final set of comparisons were conducted on data for the difficult arrays.

Post Hoc comparison for difficult arrays

A mixed factor analysis of variance of accuracy for difficult arrays found a significant main effect of image type, $F(1,91) = 7.6$, $p = .007$, $\eta_p^2 = .08$, but no significant main effects of caricature level, $p = .09$, or of target presence, $p = .1$. There was no significant interaction between target presence and caricature level, $p = .1$, or between image type and caricature level, $p = .3$, and there was no significant three way interaction between target presence, image type, and level of caricature, $p = .5$, but there was a marginally non-significant interaction between target presence and image type, $F(1,91) = 3.1$, $p = .08$, $\eta_p^2 = .03$.

Investigation of the interaction found no effect of caricature on correct identification, $p = .5$, but that caricature significantly improved rejection of the target absent arrays, $t(93) = 3.5$, $p = .001$, $d = 0.3$. Post hoc Tukey HSD tests revealed that performance for the 30% arrays did not differ from 50% arrays, p =.6, or from 70% arrays, p = .4, but there was a trend for accuracy to be better with 50% arrays than for the 70% arrays, p = .07.

A mixed factor analysis of variance of d'-prime scores found a significant main effect of image type, $F(1,91) = 6.1$, $p = .02$, $\eta_p^2 = .06$, but no significant main effect of caricature level, $p = .09$, and no significant interaction between these factors, $p = .7$. Sensitivity was greater for caricatured images ($d' = 1.9$ *v's* $d' = 1.4$), but did not differ significantly by caricature level. The mixed factor analysis of variance for response criterion also found a significant main effect of image type, $F(1,91) = 4.3$, $p = .04$, $\eta_p^2 = .05$, with no significant main effect of caricature level, $p = .9$, and no significant interaction, p =.5. Response criterion was more conservative for all of the caricatured arrays ($c = 0.26$ *v's* $c = 0.03$).

As observed throughout, sensitivity was better and response bias was more conservative for caricatured images. Consistent with the results of Experiment 16, there was no effect of caricature on accurate identification of matching images, but correct rejection of target absent arrays was significantly better for caricatured images. Performance with the 30% caricatures did not differ significantly from the 50% and 70% arrays, but the 50% caricatures were more effective than the 70% caricatures, confirming the effect that was found in both participant and items analyses for the full set of arrays.

Discussion

When the full set of arrays was employed, caricaturing images by 30% did not enhance face matching performance, although enhanced rejection of target absent arrays for the difficult subset of items was replicated. The 50% caricature transformation produced the most consistent results: false identification rates were reduced within the full set of arrays and within the subset of difficult arrays. MDFS predictions were supported by the results for images caricatured by 70%: the strongest level of caricature enhanced rejection of items that did not match, but impaired identification of images that did, providing support for the premise that if the caricature is too strong it will make images of the same person appear too different. In summary, the most consistent effect of caricature for unfamiliar face matching is enhanced rejection of target absent arrays. The 70% level of caricature produced the poorest levels of performance overall, and the 30% level of transformation was only effective with more difficult arrays. At the 50% level, the caricature transformation consistently improved rejection of target absent arrays with no detriment to accurate face matching decisions.

Perceptual sensitivity was significantly enhanced by caricature, while response criterion was more conservative. These effects did not differ across different levels of transformation. Sequential line-up procedures are also associated with a conservative response bias, which generally produces better levels of correct rejection but reduced levels of identification (Lindsay & Wells, 1985; Lindsay & Bellinger, 1999). As the caricature transformation consistently improves discrimination, it is possible that in a sequential matching task greater sensitivity to caricatured images might combine with higher criterion to facilitate accurate face matching decisions.

Experiment 18

Sequential face matching with caricature

The effectiveness of a sequential face matching procedure was explored with original and caricatured faces. In the sequential procedure, array items were presented individually and each remained on screen until a matching decision was made. It was predicted that in the sequential procedure enhanced sensitivity to** caricatured images would reduce false positive identification and might also increase accurate identification.

Participants

Eighty-eight students from the University of Stirling participated in return for course credit. Twenty-nine of the participants were male, and ages ranged from 17 to 53 years with a mean of 23.1 (*s.d.* 7.4). All had normal or corrected to normal vision.

Materials and design

The 70% level of caricature was shown to impair accurate face matching and was not included in this evaluation. A 2 x 2 x 2 mixed factor design was employed with level of caricature (30%, 50%) as the between participant factor, and target presence and image type (original, caricatured) as within participant factors.

Sequential face matching arrays were generated from the full set of original arrays, and with the 30% and 50% caricature materials. Each trial consisted of the target image shown to the left of each array item in turn and the order of array items was randomised with the constraint that in target present trials the sequential position of the matching image was equally sampled for both original and caricatured arrays.

The trials were presented using E-Prime software on a 17 inch LCD monitor at 1024 x 768 pixels resolution; each pair of images was presented at the mid point of the screen and separated by 125 mm with each head subtending a viewing angle of 2.5 ° at 70 cm viewing distance. Presentation of all 80 arrays was randomised for each participant; half contained original faces and half contained caricatured faces, with the target was present in half of each set. All conditions were counterbalanced across participants.

Procedure

Testing was conducted individually and instructions were provided both verbally and with on screen prompts. The participants were informed that they would be shown pairs of faces and should decide whether the images were of the same person or of different people. They were advised that a number of faces could be shown for each target and that there may or may not be a matching face within each set. They were asked to respond in their own time; if they judged the faces to be of different people they were to type 'n', and if they thought the images were of the same person they were to type 'y'. In accordance with Lindsay and Wells (1985) each item was viewed once, the number of array items was not disclosed, an array was shown until a selection was made or until all of the items had been rejected, and decisions could not be changed.

Results

Within a sequential matching procedure correct identification of matches was improved for both 30% and 50% levels of caricature, but the rates of correct rejection for target absent arrays do not appear to have been enhanced. The means and standard deviations by caricature level and condition are shown in Table 5.4.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
30%	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	44.9	*17.1*	11.7	*11.8*	43.4	*18.8*	61.0	*24.8*
Caricatured	56.4	*18*	14.5	*12.2*	29.0	*17.7*	62.5	*26.2*
50%	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Original	38.4	*14.8*	8.2	*8.6*	53.4	*18.1*	49.7	*27.3*
Caricatured	51.8	*13.9*	8.1	*11.5*	40.2	*17.3*	52.4	*26.3*

Table 5.4. Evaluation of face matching for original faces, and for faces caricatured by 30% or 50%, within sequential arrays. The results are broken down by the percentages of Hits, Misses and False identification for target present arrays, and percentages of correct rejections for target absent arrays.

A mixed factor analysis of variance of correct responses with level of caricature (30%, 50%) as the between participant factor, and target presence (target absent, target present) and image type (original faces, caricatured faces) as within participant factors, found a significant main effect of target presence, $F(1,86) = 18.2$, $p < .001$, $\eta_p^2 = .17$, a significant main effect of image transformation, $F(1,86) = 56.5$, $p < .001$, $\eta_p^2 = .40$, and a marginally non-significant main effect of caricature level, $F(1,86) = 3.7$, $p = .06$, $\eta_p^2 = .04$. There was no significant interaction between target presence and caricature level, $p = .2$, or between image type and caricature level, $p = .4$, but there was a significant two way interaction between target presence and image type, $F(1,86) = 31.1$, $p < .001$, $\eta_p^2 = .27$, with no three way interaction between target presence, image, and caricature level, $p = .8$.

The interaction was explored with paired sample *t*-tests. Caricaturing the images did not enhance rejection of the target absent arrays, $p = .2$, but when the target was present, identification rates were better for caricatured than original images, $t(87) = 11.3$, $p < .001$, $d = 0.7$. Overall performance was better for the 30% caricature group than for the 50% caricature group.

A mixed factor analysis of variance on *d'* scores with image type (original, caricature) and caricature level (30%, 50%) as factors found a significant main effect of image type, $F(1,86) = 29.0$, $p < .001$, $\eta_p^2 = .25$, and a marginally non-significant main effect of caricature level, $F(1,86) = 3.8$, $p = .05$, $\eta_p^2 = .04$, but no significant interaction between these factors, $p = .5$. Sensitivity was better for the caricatured arrays than for the original arrays ($d' = 0.43$ *v's* $d' = 0.03$), and was also better within the 30% caricature group than the 50% caricature group ($d' = 0.51$ *v's* $d' = -0.04$). The mixed factor analysis of variance for response bias also found a significant main effect of image type, $F(1,86) = 12.6$, $p = .001$, $\eta_p^2 = .13$, but no significant main effect of caricature level, $p = .1$, and no significant interaction, $p = .3$. Response criterion for the original faces was more conservative than for the caricatured images ($c = 0.23$ *v's* $c = 0.09$).

Discussion

In keeping with previous observations, sensitivity was enhanced by caricaturing; however, with the sequential format the effects on criterion were reversed and response bias for the caricatured faces was lower than for the original images. A liberal response criterion is generally associated with increased levels of selection. When the target was present correct identification increased but false identification did not; therefore, more selections were made but only for the correct images. When the target was not present, reduced criterion should produce higher levels of false identification but as this was not observed; enhanced discrimination appears to have compensated and accuracy relative to the original images was maintained.

Within the sequential matching procedure caricature significantly enhanced identification of matching items but did not improve rejection of target absent arrays. This is a direct contrast to the effects obtained with simultaneous arrays which showed enhanced rejection of target absent arrays, with no effect on accurate identification. Of more importance, however, the overall levels of accuracy were poorer with the sequential procedure than observed previously with the simultaneous procedure. This indicates that for perceptual discrimination the sequential format is less effective.

Summary and conclusions

Caricatures exaggerate the aspects of a face that make it distinctive, and have been shown to enhance identification of familiar faces (Benson & Perrett, 1991; 1994; Ellis, 1990; Lee & Perrett, 1997; Lee, Byatt, & Rhodes, 2000; Rhodes, Brennan, & Carey, 1987; Rhodes & Tremewan, 1994; Stevenage, 1995; Tanaka & Simon, 1996). The work described in this chapter investigated whether caricaturing unfamiliar faces could also confer an identification advantage by increasing distinctiveness and discrimination in a face matching task.

Theoretical considerations

The effect of caricature was conceptualised within a MDFS framework which entailed certain predictions: if all of the faces in a matching array were caricatured relative to a norm face, the distance between each of the faces, and from the origin, would become greater, resulting in images that were more distinctive. Systematic caricature would therefore make the faces more dissimilar and would reduce false positive identification. Since caricaturing would also increase perceptual differences between the target image and the correct array image, the effects on accurate identification would depend on the degree of image similarity, and the level of caricature transformation: if caricaturing caused the difference between the target and the correct match to be smaller than the difference between the target and the foil images, accurate identification would be enhanced; but if caricature caused the difference between the images to become too great, correct identification would be impaired.

Within the simultaneous face matching procedure MDFS predictions were supported. Caricature enhanced discrimination and reduced false positive identification at all levels, although at 30% this was only observed with a subset of difficult arrays. At the 70% caricature level accurate face matching also declined, indicating that with this amount of transformation images of the same person had become too dissimilar. Accurate identification did not significantly improve, although a trend for enhanced matching of easier arrays at 30%, hinted that a small amount of caricature may sometimes be beneficial. The most consistent effects were obtained with the 50% caricature transformation, which improved rejection of target absent arrays with no detriment to accurate face matching decisions.

In memory tasks sequential line-ups show enhanced rejection of foils in comparison with a simultaneous procedure (Lindsay & Wells, 1985), but in face matching the sequential procedure was less effective, indicating that perceptual discrimination is more successful with comparison of multiple images. The effects of caricature were also reversed such that response criterion for caricatured images was reduced instead of heightened, which meant that more selections would be made. As sensitivity was also enhanced, this resulted in higher levels of correct identification but no observable effect on rejection of the target absent arrays. While it is consistent that caricature would increase distinctiveness and improve sensitivity, it is not clear why the sequential procedure should reduce levels of response bias when the contrasting effect is typically observed. In summary, systematic caricaturing can enhance discrimination of unfamiliar faces, and will do so most effectively within a simultaneous matching procedure. Within this set of images the 50% transformation produced the most consistent results.

The caricature transformation was conceptualised within a MDFS framework and is akin to the effects of distinctiveness as they are described within a norm based MDFS model. It should be noted that this is an artefact of this methodology and this work is not intended to provide support for norm based encoding. Within a MDFS model the discrimination advantage for caricatured faces is caused by increasing the Euclidian distance from origin of the face space, and therefore between the face representations. The act of shifting representations in a face space and the reported caricature effects do not provide any information as to whether the faces are coded relative to a norm face, or relative to other face exemplars.

Unfamiliar face matching relies on being able to form a good correspondence between the structural codes that are defined by different image properties (Bruce, Henderson, Greenwood, et al. 1999). This task can be challenging and depending on image clarity, viewpoint, and lighting, different people will appear more similar than images of the same person (Adini, Moses, & Ullman, 1994). The appearance of facial shape and structure is determined by distance, angle of view, and reflectance from the available lighting; but while differing viewpoints can be interpreted fairly well, lighting differences significantly impair face matching performance (Hill & Bruce, 1996). The images employed in this study were captured on the same day in good conditions that attempted to avoid lighting effects and control for angle of view (Bruce et al., 1999). In spite of this, superficial differences were exaggerated by the caricature transformation: with images captured at different times, in different lighting, or at different distances, the effects of image quality or environment would reduce perceptual correspondence between pictures that show the same person, and performance would be notably poorer. A potential solution to these effects may be found in 3-dimensional face modelling: with the technology to fit 2-dimensional face images to 3d models that can reconstruct and rotate head images, many artefacts and imaging errors could be routinely corrected.

It might now seem logical to carry out a larger study designed to establish an optimal level of caricature for unfamiliar face matching, but there are a number of factors that would make this unproductive. Perceptions of best likenesses produced by caricature are highly variable (Ellis, 1990; Frowd et al., 2007; Benson & Perrett, 1991; Benson & Perrett, 1994; Rhodes et al., 1987), while the effect sizes for the target present condition are also rather small; to evaluate several levels of caricature would require a very large number of participants to produce significant results, which would in any case be linked to a particular set of images. A more fruitful approach would be to establish whether the caricature advantage observed here can be replicated with different image sets of varying task difficulty. It would also be interesting to conduct some formal modelling of the results, to see whether simple face space models do adequately account for the data.

Applied considerations

Identity verification from face images is increasingly commonplace and demand for enhanced security for financial, industrial and law enforcement has fuelled development of biometric face recognition applications. To appreciate how this is important for human face perception, it is necessary to understand how the applications are used: when a person attempts to gain access to a secure environment that is controlled by a face recognition system, their face will be presented along with the best matches from the database and at this point security personnel must decide whether any of the images is a match. The ultimate decision is a human one, and difficulty in unfamiliar face matching means that identification errors will be common and procedures such as the one described here will be beneficial.

Within a biometric system, each face in the dataset is represented by a complex template, or 'mesh' of data points, which means that the functionality to manipulate images automatically is already in place. Thresholds for image matching are determined by system administrators; therefore, it would be possible to set levels of correspondence between the biometric templates that would determine how much of a caricature transform should be applied. Studying caricature within a live biometric dataset would allow caricature effects to be determined across variable face images, and would enable development of a systematic transform that can determine an appropriate level of caricature by degree of biometric correspondence; that is. with poorer biometric association, a weaker caricature transformation would be prescribed.

Conclusions

Within this series of experiments the images were caricatured relative to the average face shape of all of the photographs in the dataset. With computational caricatures the reference image will determine the metrics of the average face and therefore how each individual image deviates from the norm. Different pictures of the same person should differ from the norm in a similar way, while pictures of different people should differ from the norm in different ways. It is therefore important that the same reference image is used to caricature the target and the images that will be shown in a matching array; if different average images are employed, distinctiveness will not be generated along the same dimensions and correspondence between the target and array items will be reduced.

There is, however, an alternative caricature procedure that would not employ average reference images and might also enhance identification of accurate matches: if the items in an array are caricatured not with reference to a norm or average image but with reference to the target face, then items that do not match, or that vary in different ways, should become more dissimilar to the target than images that portray the same person. By this logic, a target specific caricature would greatly increase sensitivity and should be even more effective than a generic caricature transformation. If effective, this transformation would also have the advantage that no global average would need to be computed in order to transform a set of images. A pilot study of target specific caricature effects is currently under way at the University of Stirling.

Future study should also return to the issue of sequential face matching decisions. The results of Experiment 18 provided evidence that one to one correspondence of images produces fewer correct matching decisions than simultaneous face matching arrays. However, the most common method of facial image verification involves inspection of individual images and this typically involves comparison of a person with photographic identification. Kemp et al. (1997) evaluated the effectiveness of photographic credit cards and showed that in a live sequential matching procedure participant checkout operators demonstrated high acceptance levels for both the legitimate *(m = 89.8%)* and the fraudulent cards *(m = 51.2%).* It is clearly not possible to caricature a live person, but as there is a gross discrepancy in the size and clarity of photographic ID and the bearer, it would be useful to evaluate whether applying caricature to the ID image might highlight diagnostic characteristics and enable accurate matching decisions. A pilot study of caricatured identification cards is presently under way at the University of Stirling.

In summary, systematic caricaturing can enhance discrimination of unfamiliar faces, and will do so most effectively within a simultaneous matching procedure. Within this set of images the 50% transformation produced the most consistent effects. These results indicate that biometric applications may have the potential to harness variable caricature transforms to enhance discrimination of unfamiliar faces.

6

Expert face processing and the own race bias

Own race bias in memory is consistently shown in poorer recognition of other race faces. Theories about the cause of this effect typically focus on differences in the way that other race faces may be encoded, or stored in memory; consequently, perceptual discrimination of other race faces has received much less attention. The research in this chapter evaluates face matching and investigates whether difficulty identifying other race images is caused by inexperience with the facial variation of other ethnic groups, or because the faces are processed in a less efficient manner. To explore the importance of perceptual experience a computerised image transform was employed to transform African American and Japanese faces towards an average Caucasian face shape. Identity and information such as skin tone and hair type was maintained but the featural shapes, and configurations were altered to approximate facial dimensions with which Caucasian viewers are familiar.

Perception of other race faces

Anecdotally other race faces appear to have less variability than own race faces, but while there is no evidence that this is true (Goldstein, 1979a, 1979b; Goldstein & Chance, 1978; Valentine & Endo, 1992), numerous studies show that recognition of other race faces is generally much poorer (Meissner & Brigham, 2001). This has been observed in memory tasks (e.g. Rhodes, Locke, Ewing, & Evangelista, 2009), delayed matching tasks (e.g. Tanaka, Kiefer, & Bukach, 2004), discrimination tasks (e.g. Megreya, White, & Burton, 2011), and eyewitness paradigms (e.g. Pezdek & Blandon-Gitlin, 2005). Most studies have employed Caucasian and African American samples, but ORB is reported in Asian (e.g. Hayward, Rhodes, & Schwaninger, 2008), Hispanic (e.g. MacLin, MacLin, & Malpass, 2001), Arabic (e.g. Rattner, Weimann, & Fishman, 1990) and African (e.g. Wright, Boyd, & Tredoux, 2003) populations.

Asymmetry in the occurrence of ORB in different racial groups is common and there is some indication that inter-racial contact may reduce, or eliminate the effect. For example, level of contact was inversely associated with ORB in Africans and Caucasians in South Africa and the UK (Chiroro & Valentine, 1995), but this was not found in Caucasians and Asians in Canada and Singapore (Ng & Lindsay, 1994). However, a study of Caucasian and South Asian teenagers in the UK reported that it was not just exposure, but the need to individuate faces that predicted discrimination ability (Walker & Hewstone, 2006). Meissner and Brigham (2001) suggest that levels of ORB in different racial groups may also reflect "social utility" (Malpass, 1990): racial minorities have to individuate out-group majority faces and will not exhibit ORB, while a majority race with less need to individuate minority faces, will show ORB. Another social theory is conceptualised as cognitive disregard (Rodin, 1987). In this model, out-group faces have less importance than in-group faces and are processed at a superficial level; consequently they will not be effectively encoded in memory. Levin (2000) proposes that racial categorisation determines how faces will be encoded because when race is encoded as a feature, it is at the expense of individuating information. Some support for this account is found in studies showing that instruction to individuate faces can eliminate ORB (Hugenberg, Miller, & Claypool, 2007; Lebrecht, Pierce, Tarr, & Tanaka, 2009; Rhodes et al., 2009). However, instructions intended to enhance depth of processing (i.e. character judgements or preparation for a memory test) do not reduce ORB, which suggests that other race face processing is different in quality rather than just quantity (Chance & Goldstein, 1981; Devine & Malpass, 1985).

Holistic face processing and configural analyses are associated with face processing expertise (Diamond & Carey, 1986). A number of studies have shown that these processes are more efficient for own race faces, indicating that other race faces may be processed in a more featural, or piecemeal way (Rhodes, Hayward, & Winkler, 2006; Rhodes et al. 1989; Michel et al., 2006; Tanaka et al., 2004; Michel, Caldara, & Rossion, 2006). Yet the type of face processing that is employed, i.e. holistic or featural, may not be determined by the facial attributes. When ambiguous faces (i.e. face images that are a hybrid or could belong to one of two races) are cued as own race, or as belonging to another race (MacLin & Malpass, 2001; Pauker & Ambady, 2009; Pauker et al., 2009), face memory is better for the images identified as own race, and Michel, Corneille, and Rossion (2007) have shown that racial categorisation determines whether the ambiguous faces are processed holistically. Conversely, Hugenberg and Corneille (2009) found that categorisation as social out-group (belonging to another university) caused own race faces to be processed in a featural way. The effects of race and social grouping may also interact depending on motivation; Hehman, Mania, & Gaertner (2010) found that when faces were categorised by university there was no within own university evidence of ORB, and other race faces labelled as own university, were remembered better than same race faces that were affiliated with another university. Collectively these studies indicate that when social grouping is a salient factor, categorisation rather than racial experience may determine whether expert face processing strategies are applied. It should be noted, however, that Hehman et al. (2010) employ the same images at study and test; therefore, the effects of social categorisation on other race *picture* recognition may not generalise to recognition of other race *faces*.

In contrast to cognitive allocation of processing strategies that determine how well other race faces are encoded, perceptual expertise theories suggest that discrimination, or lack thereof, is determined by experience. By this account, limited exposure to other race faces means that an appropriate range of variation to distinguish them is never learned (MacLin & Malpass, 2001). Perceptual learning theories are supported by findings that individuation experience enhances recognition (e.g. Walker & Hewstone, 2007; Tanaka & Pierce, 2009), while performance can be improved by training viewers to attend to racially appropriate dimensions (Hills & Lewis, 2006). The theory is also compatible with the concept of multi-dimensional face space (Valentine, 1991): in a MDFS model, other race faces are encountered infrequently and will be distinctive on at least one race specifying dimension. Within the face space other race faces would be encoded in a cluster away from the central tendency of own race faces (Rhodes & McLean, 1990), making them perceptually similar and difficult to discriminate. Consistent with perceptual learning, neonates show no preference for own race faces (Kelly et al., 2005), but there is evidence of perceptual tuning according to the faces that are encountered throughout early childhood (Bar-Haim et al., 2006, Kelly et al., 2007; 2009; Sangrigoli et al, 2005).

There is evidence that perceptual expertise and expert face processing strategies may be absent, or lacking, in discrimination of other race faces. However, social categorisation studies suggest that these effects are evidence of out-group discrimination within which other race face processing can, on occasion, be superior (e.g. Hehman et al., 2010). The interplay of social categorisation, perceptual expertise, and motivation is described in the categorization-individuation model (Hugenberg, Young, Bernstein, & Sacco, 2010). Within this account, face processing begins with spontaneous social categorisation (i.e. race.), directing attention to information that defines the category, rather than information that discriminates among category members. If there is sufficient motivation to individuate a face, attention shifts from category features to unique characteristics. However, it remains unclear whether racial categorisation of a face as out-group elicits less effective face processing strategies, or whether perceptual expertise cannot be effectively applied to other race faces because facial variation is unfamiliar. Face matching employs perception without memory and, as such, is an ideal paradigm to investigate the influence of perceptual expertise. If face processing is tuned to own-race variation and is thus ill equipped to differentiate other race faces, making the other race faces vary along familiar own race dimensions should enhance performance.

Aim of study

The aim of this chapter was to evaluate a novel cross race face transformation and explore the effect of perceptual expertise in facial identification. If a substantial cause of the cross race recognition deficit is inexperience with facial variation, manipulating the shape of other race faces to portray Caucasian dimensions should improve discrimination by Caucasian observers; conversely, manipulating Caucasian faces to other race dimensions would impair own race discrimination. If however, racial categorisation determines how effectively the faces are encoded, the transformation might be ineffective.

A second aim was to test the effect of different face matching procedures. In memory research multiple image comparison in line-ups is associated with observers forming relative judgements about the best possible match, which is thought to increase false identification (Wells, 1984; Lindsay & Wells, 1985; Lindsay & Bellinger, 1999; Smith, Stinson, & Prosser, 2004). Sequential line-ups require absolute judgements for each image consecutively, leading to fewer false positives (Lindsay & Wells, 1985; Lindsay & Bellinger, 1999). As relative judgements and false identification are also more likely in cross race identification (Smith, Lindsay, Pryke, & Dysart, 2001; Jackiw, Arbuthnott, Pfeifer, Marcon, & Meissner, 2008), simultaneous and sequential face matching procedures might produce different effects for the own and other race faces.

Perceptual expertise was explored in a multiple item simultaneous face matching task in Experiment 19. Own race Caucasian, and African American faces were presented with ten item arrays in both their original format, and when they were transformed towards the dimensions of the contrasting race. Experiment 20 evaluated face matching with the same images within sequential arrays, while Experiment 21 employed a novel procedure that included both sequential and simultaneous presentation. It was predicted that discrimination of African American faces would be poorer than of Caucasian faces, but would be enhanced by transformation towards average Caucasian dimensions. Identification of Caucasian items was expected to decline when images were transformed toward the African American dimensions. False identification rates were expected to be better when a sequential procedure was adopted.

Experiments 22-24 were intended to replicate the cross race effects observed in Experiments 19 - 21 with images of a different race. The Caucasian shape transformation was explored with Japanese face images within simultaneous face matching arrays in Experiment 22, sequential face matching arrays in Experiment 23, and within the novel combined procedure in Experiment 24. Within this series of experiments it was possible to explore perceptual expertise in other race face discrimination within simultaneous and sequential face matching procedures.

Defining race and transforming facial characteristics

Race characteristics were defined and race transformed images were produced with Psychomorph software (Tiddeman, Burt, & Perrett, 2001). Within three sets of male faces; Caucasian *(n* = 97*)*, African American[3] *(n* = 116*)*, and Japanese[4] *(n* = 80*)*, a template was generated for each image by tagging the facial features and face outline with 179 data points. Within each set of faces the average location of each data point was then computed, and the resulting template was used as the average shape reference for that race. From these shape references it was possible to quantify how the average Caucasian face shape differed from the

[3] Caucasian and African American images supplied by Meissner; (Meissner, Brigham, & Butz, 2005).
[4] Japanese images supplied by ATR.

average African American face shape (or from the average Japanese face shape), and to define a mathematical transformation that describes the difference between the average characteristics. A multi-dimensional face space representation of the race transformation is shown in figure 6.1.

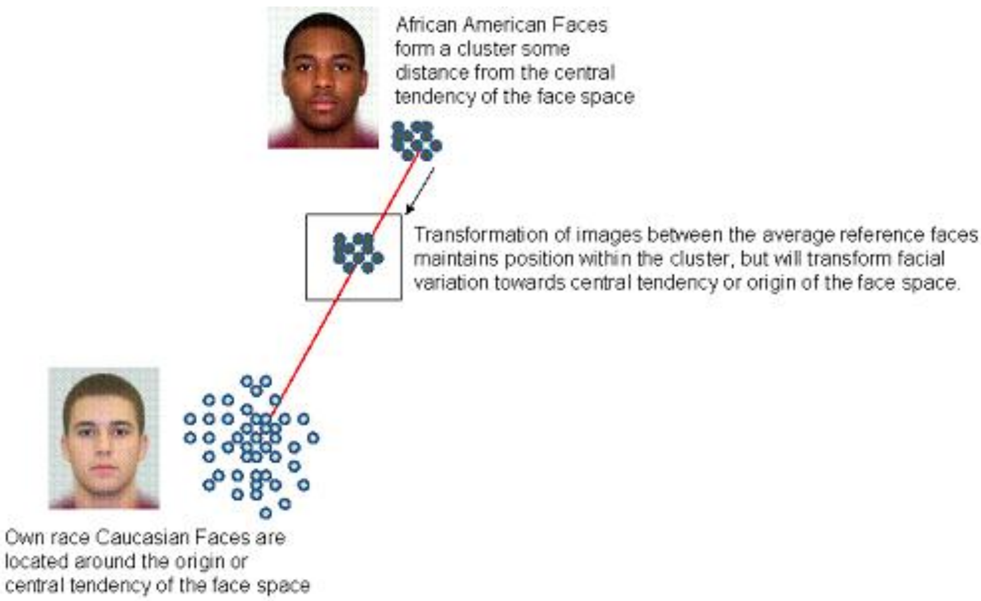


Figure 6.1. A MDFS representation of the average shape race transform. Each circle represents a single face in the face space. Own race images are distributed around the norm or origin and form the central tendency; other race faces will be coded along at least one race specifying dimension and will form a cluster some way from the centre. Transformation shifts the other race faces towards the area of learned variation, and should make the images easier to discriminate. Note that the position of each image is maintained relative to the other images within the cluster.

Individual images could then be transformed by warping the shape information 100% away from the same-race reference face, toward the reference shape of another race. For example, warping the shape information of individual African American faces away from the African American average reference toward the Caucasian average reference would cause the faces to show shapes and variation with which Caucasian observers were familiar, although the face images would continue to differ from each other. Examples of original and race-transformed images are shown in figure 6.2.

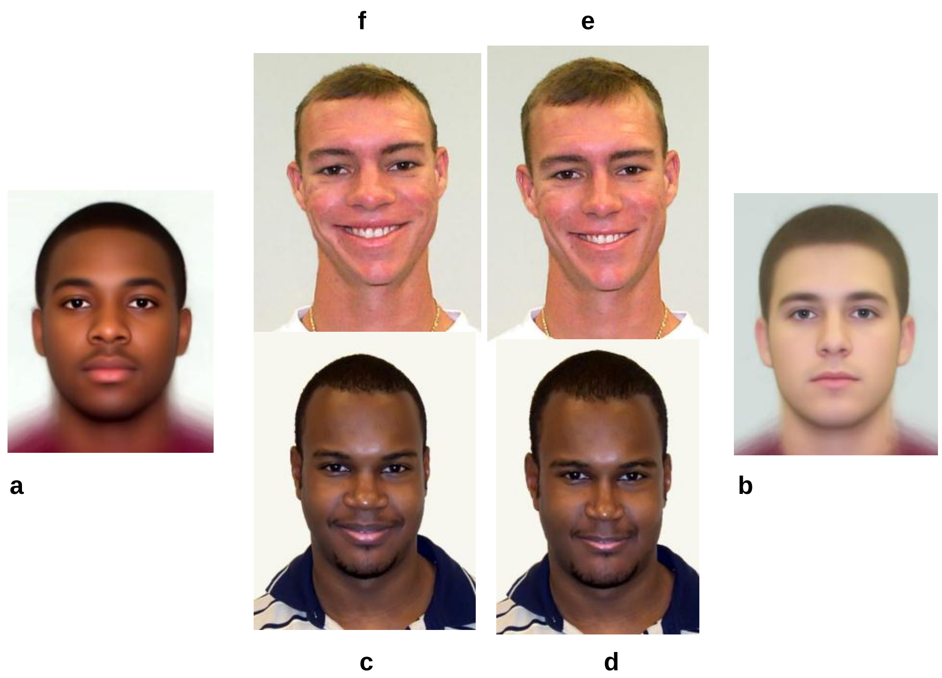


Figure 6.2. The African American average face shape reference (a) and the Caucasian average face shape reference (b). The difference between the average templates could then be applied to any face image: (c) shows an original African American image, (d) shows the African American image when it is transformed 100% away from the average African American shape towards the average Caucasian shape; (e) shows an original Caucasian image, (f) shows the Caucasian image when it is transformed 100% away from the average Caucasian shape towards the average African American shape.

Matching own race Caucasian and African American faces

Experiment 19

Matching original and race transformed faces in simultaneous arrays

Own race bias in perceptual discrimination was evaluated within simultaneous matching arrays of Caucasian and African American faces. To explore the influence of perceptual expertise on other race face processing, half of the arrays presented original images, and half presented images that were transformed toward the facial shape of the other race.

Participants

Seventeen Caucasian students from the University of Stirling participated in return for course credit. Four participants were male, and ages ranged from 18 to 37 years with a mean of 21.3 (*s.d.* 5.8). All had normal or corrected to normal vision.

Materials and design

Face matching arrays were generated for 32 Caucasian and 32 African American male targets. The probe images showed each target wearing street clothes in a full frontal smiling pose. Arrays were constructed using photographs that bore resemblance to the targets[5]: each item showed a male wearing a burgundy sweatshirt in full frontal pose with a neutral expression, all images were in colour. Trials consisted of a target image shown above 8 photographs of foil images (target

[5] Line-ups should be constructed by an investigator of the same race as the target (Brigham & Ready, 1985); these arrays were agreed by Caucasian observers for investigation of Caucasian ORB, they would not be appropriate for study with African American participants.

absent condition), or above 7 foil images and an equivalent photograph of the target (target present condition). Array items were arranged in two rows and numbered 1-8; position of the target was randomly sampled with the constraint that each position was employed four times for each race. The images were cropped to show head and shoulders, and measured 50 mm x 80 mm. The head within each image gave a viewing angle of approximately 2.5° at 70 cm viewing range; the complete array with target image measured 260 mm x 270 mm. Within the race-transformed arrays both the target and the array items were altered using the same transformation and reference images. An example array is shown in Figure 6.3. A fully within participant 2 x 2 x 2 repeated measure design was employed with target presence, race (Caucasian; African American), and image type (original images; race-transformed images) as factors.

Procedure

Participants were tested individually: the experiment was conducted using E-Prime software (apparatus as before), and responses were recorded via the keyboard. Instructions were provided both verbally and with on screen commands. Participants were asked to complete a face matching task and informed that the target may or may not be present in each line-up array. To identify an item as a match, they were to type the corresponding number; if no match was identified, they were to press the space bar. All 64 arrays were employed for each participant; half of each race set portrayed race-transformed images and the target was present in half of each subset. The conditions were fully counterbalanced across participants, and presentation order of trials was randomised for each participant.

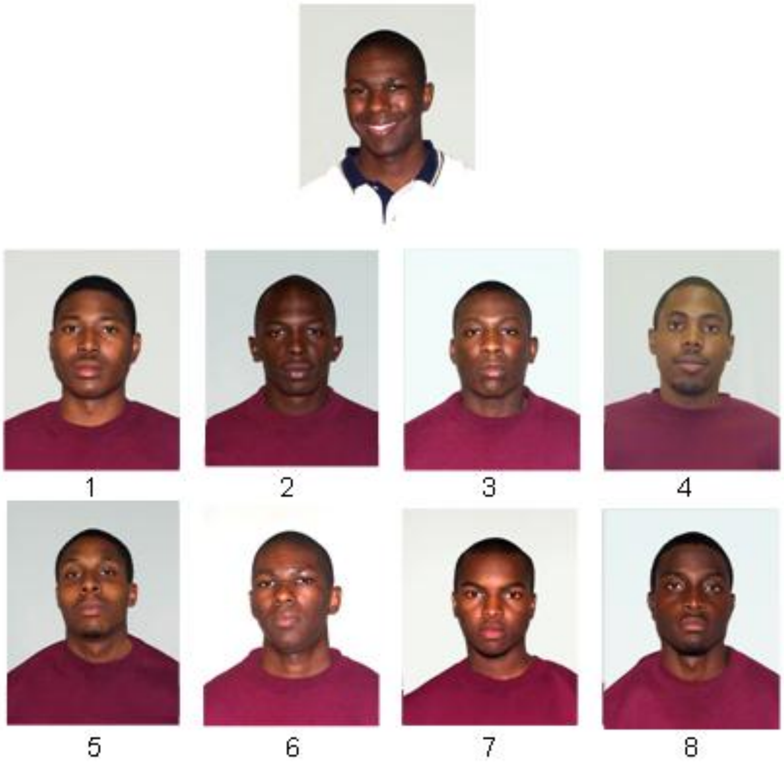


Figure 6.3. An example of a simultaneous face matching array comprising original African American faces. The target, or probe image shows a smiling African American male in street clothes. The array shows eight African American males with neutral expressions and identical clothing. In target present arrays, array position of the correct match was equally sampled, and within race-transformed arrays the target and the array items were altered using the same transformation and reference images.

Results

Accuracy was poorer with African American than with Caucasian arrays but is not evidence of ORB as the relative difficulty of the arrays is unknown. Of more interest, is the effect of the other race shape transform: while transforming own race faces towards the African American average showed the expected decrement, transforming African American images towards the average Caucasian shape also reduced performance.

Caucasian	Target Present: Hits M	S.D.	Miss M	S.D.	False ID M	S.D.	Target Absent: Correct M	S.D.
Veridical	78.7	17.5	13.2	14.3	8.1	9.8	76.5	23.3
Transformed	68.4	21.7	16.9	17.1	14.7	14.1	65.4	29.5
African American	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Veridical	69.9	20.3	9.6	11.3	20.6	13.2	52.9	25.6
Transformed	62.5	22.5	18.4	18.3	19.1	18.3	47.1	27.4

Table 6.1. Evaluation of own race Caucasian and African American face matching within simultaneous arrays. Results are provided for veridical and race-transformed arrays and are broken down by the percentages of hits, misses and false identification for target present arrays, and the percentage of correct rejections for target absent arrays.

Correct scores were analysed using a repeated measures analysis of variance, with factors of target presence, race (Caucasian; African American), and image type (original images; race-transformed images). This showed no significant main effect of target presence, $p = .1$, but there were significant main effects of race, $F(1,16) = 24.2$, $p < .001$, $\eta_p^2 = .6$, and of image transformation, $F(1,16) = 8.2$, $p = .01$, $\eta_p^2 = .34$. Matching was significantly better for own-race Caucasian faces (72.5%, *s.e.* 4%), than for African American faces (58.5%, *s.e.* 3%), and was significantly impaired when the images were transformed; 61.3% (*s.e.* 3%), in comparison with 69.8% (*s.e.* 4%) for the unaltered images. There was a non-significant interaction between target presence and race, $F(1,16) = 3.8$, $p = .07$, $\eta_p^2 = .19$. No other interactions approached significance *(all p > .5).* When the target was present, matching tended to be easier for own-race faces, $t(16) = 1.9$, $p = .08$, $d = .4$; but when the target was absent, there were significantly more false identifications from African American arrays, $t(16) = 4.1$, $p = .001$, $d = .8$. As in memory studies

(Jackiw et al., 2008; Smith et al., 2001), cross-race matching elicited considerably more incorrect selections.

ORB has been associated with reduced discrimination and more lenient response criterion; if perceptual expertise can account for ORB, transforming African American images towards an average Caucasian face shape should make them more distinguishable to Caucasian observers. Signal detection analysis of hits and false positives were conducted to obtain measures of sensitivity *(d' prime),* and response bias *(criterion c).* Repeated measures analysis of variance on *d'* scores, with race and image transformation as factors, found a significant main effect of race, $F(1,16) = 21.2$, $p < .001$, $\eta_p^2 = .6$, and a significant main effect of image transformation, $F(1,16) = 6.4$, $p = .02$, $\eta_p^2 = .29$, with no significant interaction, $p = .6$. Caucasian observers were more sensitive to the variation in own race faces ($d' = 1.8$), than to African American faces ($d' = 0.42$). Sensitivity was not enhanced by transforming African American images towards the own race average, and was reduced for all of the transformed images, $d' = 0.77$ in comparison with unaltered images, $d' = 1.4$. Comparable analysis of response bias found a significant main effect of race, $F(1,16) = 7.1$, $p = .02$, $\eta_p^2 = .31$, but no main effect of transformation, $p = .9$, and no significant interaction, $p = .6$. As demonstrated by lower rates of incorrect selection, response bias was more conservative for Caucasian images ($c = 0.1$), than for African American faces ($c = -0.3$).

Discussion

The results were inconsistent with a straightforward perceptual expertise explanation of ORB: transforming African American faces towards the average Caucasian shape did not improve performance, although transformation of own race

Caucasian faces towards the average African American shape did follow the predicted pattern. As expected, sensitivity was better for own race versus other race faces, but rather than being enhanced by the transformation sensitivity for African American faces was significantly impaired. Simple categorisation of race as a feature (e.g. Levin, 2000) cannot explain the result either: the unaltered African American skin tones would have produced comparable effects for both the original and the race-transformed images, but the transformed images were less distinguishable. The results suggest that the race transformation was ineffective and that, rather than making African American faces more distinguishable to Caucasian observers, all of the transformed images became harder to discriminate. It is possible that categorisation of African American images as out-group prompted them all to be processed in a featural way (e.g. Hugenberg & Corneille, 2009), but the transformation reduced the distinctiveness of the faces and made this less effective.

Face matching requires that similar faces are discriminated, and within a simultaneous procedure, a number of images are regarded at the same time. Palermo and Rhodes (2002) evaluated the effect of divided attention in face matching, and concluded that attending to a number of faces disrupted holistic face processing. Paradoxically, the race-transformation is holistic and attempts to foster perceptual expertise: it is possible therefore, that featural task demands negated any holistic perceptual advantage from the Caucasian shape transformation. This would also suggest that if other race faces are processed in a more featural way (Rhodes, Hayward, & Winkler, 2006; Rhodes et al. 1989; Michel et al., 2006; Tanaka et al., 2004; Michel, Caldara, & Rossion, 2006), that a simultaneous matching task would be better suited to discrimination of the unaltered images. If however, sequential arrays promote holistic face perception (e.g. Palermo & Rhodes, 2002); this process might enable the race-transform to enhance discrimination of other race faces.

Experiment 20

Matching original and race transformed faces in sequential arrays

Experiment 20 evaluated a sequential face matching procedure using the same arrays that were employed in Experiment 19. Within the sequential procedure the array items were presented individually and remained on screen until an absolute matching decision was made. It was predicted that absolute judgements for each item would reduce false positive identification, but might also reduce the number of correct selections. It was also proposed that direct comparison of the target with each array item would be more compatible with expert holistic face processing and might enable transformation of African American faces towards familiar Caucasian dimensions to be beneficial.

Participants

Sixty-eight Caucasian students from the University of Stirling participated in return for course credit. Twenty-four were male and ages ranged from 18 to 46 years with a mean of 20 (*s.d.* 6.2). All had normal or corrected to normal vision.

Materials and design

Sequential face matching arrays that consisted of the probe image shown to the left of each item in turn were generated from the materials employed in Experiment 19. The trials were presented using E-Prime software (apparatus as before). Each pair of images was presented at the mid point of the screen and separated by 125 mm with an overall area of 225 mm x 80 mm; each head subtended a viewing angle of 2.5 ° at 70 cm viewing distance. A repeated measure design was employed with target presence, race (Caucasian; African American), and image type (original images; race-transformed images) as factors. Half of each race set portrayed transformed images with the target present in half of each subset. All conditions were counterbalanced across participants, and presentation of arrays was randomised for each participant.

Procedure

Testing was conducted individually and instructions were provided both verbally and with on screen prompts. The participants were informed that they would be shown pairs of faces and should decide whether the images were of the same person or of different people. They were advised that a number of faces could be shown for each target and that there may or may not be a matching face within each set. The participants were asked to respond in their own time: if they judged the faces to be of different people they were to type ‘n’, if they thought the images were of the same person they were to type ‘y’. In accordance with Lindsay and Wells (1985) each item was viewed once, the number of array items was not disclosed, an array was shown until a selection was made or until all of the items had been rejected, and decisions could not be changed. The order of items within the arrays was randomised with the constraint that in target present trials the matching image appeared in each position four times for each race.

Results

The means and standard deviations are shown in Table 6.2. Accuracy for African American images was again poorer than with Caucasian arrays. Transforming own-race Caucasian faces towards the average African American shape showed the expected decrement, while transforming African American images towards the average Caucasian shape produced noticeably better performance.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
Caucasian	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Veridical	76.3	*23.6*	11.9	*17.9*	11.8	*20.0*	83.1	*26.8*
Transformed	71.1	*24.2*	10.3	*15.4*	18.6	*22.4*	62.1	*29.6*
African American	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Veridical	57.9	*21.7*	11.0	*14.1*	30.9	*20.2*	48.0	*29.8*
Transformed	68.2	*25.4*	9.9	*16.0*	21.7	*23.4*	55.0	*29.5*

Table 6.2. Evaluation of own race Caucasian and African American face matching within sequential arrays. Results are provided for veridical and race-transformed arrays and are broken down by the percentages of hits, misses and false identification for target present arrays, and the percentage of correct rejections for target absent arrays,

Correct face matching scores were analysed using a 2 x 2 x 2 repeated measures analysis of variance. The analysis showed significant main effects of target presence, $F(1,67) = 8.3$, $p = .01$, $\eta_p^2 = .11$, and of race, $F(1,67) = 116.2$, $p < .001$, $\eta_p^2 = .63$, but no significant main effect of image transformation, $p = .1$. Main effects were qualified by two way interactions between target presence and race, $F(1,67) = 15.1$, $p < .001$, $\eta_p^2 = .18$, between target presence and image transformation, $F(1,67) = 10.9$, $p = .002$, $\eta_p^2 = .14$, and between race and image transformation, $F(1,67) = 55.3$, $p < .001$, $\eta_p^2 = .45$. There was also a significant three way interaction between target presence, race, and image transformation, $F(1,67) = 5.5$, $p = .02$, $\eta_p^2 = .08$.

Planned comparisons confirmed that transforming African American faces towards the average Caucasian face shape significantly improved correct rejection of target absent arrays, $t(67) = 2.5$, $p = .01$, $d = -0.2$, and correct identifications of matching images, $t(67) = 3.7$, $p < .001$, $d = -0.4$. When Caucasian images were transformed towards the African American average shape, correct rejections were reduced, $t(67) = 6.4$, $p < .001$, $d = 0.7$, while correct identifications also declined, $t(67) = 2.1$, $p = .04$, $d = 0.2$.

Repeated measures analysis of variance on d' scores with race and image transformation as factors, found significant main effects of race, $F(1,67) = 110.0$, $p < .001$, $\eta_p^2 = .62$, and of image transformation, $F(1,67) = 5.7$, $p = .02$, $\eta_p^2 = .08$, which were qualified by a significant interaction, $F(1,67) = 55.4$, $p < .001$, $\eta_p^2 = .45$. Planned comparisons showed that Caucasian observers were considerably more sensitive to own-race faces ($d' = 2.9$) than African American faces ($d' = 0.17$), $t(67) = 12.1$, $p < .001$, $d = 1.3$. Sensitivity to the variation in African American faces was enhanced by transformation towards the average Caucasian face shape ($d' = 1.0$), $t(67) = 3.9$, $p < .001$, $d = 0.4$, and was impaired when Caucasian images were transformed toward the African American average shape ($d' = 1.4$), $t(67) = 6.6$, $p < .001$, $d = -0.6$.

Repeated measures analysis of response bias also showed significant main effects of race, $F(1,67) = 11.7$, $p = .001$, $\eta_p^2 = .15$, and of image transformation, $F(1,67) = 13.0$, $p = .001$, $\eta_p^2 = .16$, which were also qualified by a significant interaction, $F(1,67) = 8.5$, $p = .01$, $\eta_p^2 = .11$. Response bias was more conservative for Caucasian images (c = 0.3) than for the African American faces (c = -0.16), $t(67) = 4.0$, $p < .001$, $d = 0.6$. Criterion was not significantly changed by transforming African American images toward the own race Caucasian face shape (c = 0.99, p = .7); but transforming Caucasian images towards the African American average shape substantially lowered own race criterion (c = -0.18), $t(67) = 4.3$, $p < .001$, $d = 0.6$. This was observed in significantly higher levels of false identification.

Discussion

Perceptual expertise predictions were fully supported: transforming African American faces towards the own race average shape significantly enhanced discrimination, while transforming Caucasian faces toward the average African American shape made performance much worse.

Own race face perception favours holistic and configural processing (Rhodes, Hayward, & Winkler, 2006; Rhodes et al. 1989; Michel et al., 2006; Tanaka et al., 2004; Michel, Caldara, & Rossion, 2006): it was impaired when faces were transformed towards another race, and the same pattern of results was obtained in simultaneous and sequential matching procedures. Other race face perception is proposed to be more featural (Rhodes et al.,1989; Tanaka, Kiefer, & Bukach, 2004; Michel et al, 2006a,2006b): transformation towards own race dimensions improved performance when images were presented sequentially but not within an array, and is consistent with the prediction that sequential presentation would foster holistic processing and enable the race-transform to improve discrimination. The results indicate that simultaneous and sequential face matching procedures can elicit different strategies and that own race Caucasian and African American faces are also processed differently. As such, they also provide support for a featural processing explanation of own race bias.

The findings support two distinct theories of ORB and are theoretically interesting; but in terms of application, discrimination of original African American faces was better within simultaneous arrays, perhaps by enabling multiple feature comparisons. To assess the potential of the race-transform to maximise other race face discrimination while also enabling featural comparison of multiple images, a procedure combining sequential and simultaneous presentation was developed.

Experiment 21

Matching original and race transformed faces in a combined procedure

Experiment 21 employed a hybrid procedure that would enable face matching of other race images in a simultaneous array, but might also reduce false identification and enable the race transform to enhance discrimination of individual items. The combined face matching procedure employed two distinct phases: in the first phase items would be inspected individually with the target; images that the viewer was sure did not match would be discarded; in the second phase, the most similar images would be presented as a simultaneous array for multiple comparison and an ultimate face matching decision. It was predicted that within the sequential phase, absolute judgements for each item would reduce false positives, while reducing the number of items presented in the simultaneous phase would enhance identification. Transformation of Caucasian images towards the African American shape was expected to impair performance, while individual inspection of African American images transformed to own race dimensions would be beneficial.

Participants

Thirty-six Caucasian students from the University of Stirling participated in return for course credit. Nine participants were male, and ages ranged from 16 to 50 years with a mean of 25.8 (*s.d.* 9.2). All had normal or corrected to normal vision.

Materials and design

The materials from Experiments 19 and 20 were used to generate combined sequential and simultaneous face matching arrays. Within the sequential phase the target image was shown to the left of each array item until a response was provided; for target present trials the matching image appeared in each position four times for each race. When each array item had been viewed, the potential matches were displayed in a simultaneous array: if all of the items were rejected a new trial commenced. An example of a simultaneous array following elimination of non-matching items is shown in figure 6.4. The experimental design, apparatus and image characteristics were as before.

Procedure

The instructions were provided verbally, and with on screen commands. Participants were recruited to participate in a face matching task and were informed that they would be shown pairs of faces and should decide whether the images were of different people, or whether they *could* be of the same person. If they thought the images were definitely of different people, they were to type ‘n’ on the keyboard, but if they thought the images could be of the same person, they were to press the space bar. At the end of each sequential array the images that were not rejected were displayed within a simultaneous array, i.e. the target image was shown above only the items for which the space bar response had been given. Participants were advised that there may or may not be a matching face within each set, and that to identify a match they should type the corresponding number, or press the space bar to reject the entire array. In summary, the new procedure allowed participants to reject obvious mismatches sequentially, before inspecting possible matches together for final comparison and matching decision.

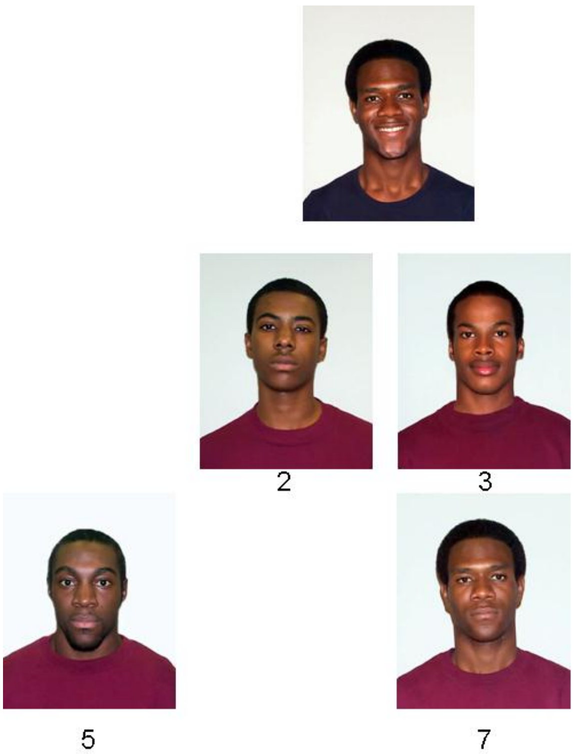


Figure 6.4. An example of a simultaneous face matching array following sequential elimination of non matching items. The sequential phase requires one to one correspondence of the target and each item. In the subsequent simultaneous array the target is shown above the most similar items for multiple comparison and a final matching decision.

Results

The means and standard deviations by condition are shown in Table 6.3. Performance was better than observed in the previous experiments and an advantage for the shape transform with the African American images was evident.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
Caucasian	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Veridical	88.8	*20.0*	2.4	*5.8*	8.7	*18.6*	91.4	*18.6*
Transformed	88.2	*18.4*	2.8	*6.1*	9.0	*18.1*	76.9	*23.7*
African American	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Veridical	81.6	*21.2*	4.2	*7.3*	14.2	*20.1*	62.7	*31.0*
Transformed	87.5	*19.8*	2.8	*7.4*	9.7	*18.0*	70.4	*31.8*

Table 6.3. Evaluation of own race Caucasian and African American face matching within combined sequential and simultaneous arrays. Results are provided for veridical and race-transformed arrays and the percentages of hits, misses and false identification for target present arrays are shown with the percentage of correct rejections of target absent arrays.

Correct face matching scores were analysed with a 2 x 2 x 2 repeated measures analysis of variance. There were significant main effects of target presence, $F(1,35) = 7.1$, $p = .01$, $\eta^2 = .17$, and of race, $F(1,35) = 38.1$, $p < .001$, $\eta^2 = .52$, but no significant main effect of image transformation, $p = .8$. Main effects were qualified by two way interactions between target presence and race, $F(1,35) = 12.8$, $p = .001$, $\eta^2 = .27$, between target presence and image transformation, $F(1,35) = 4.5$, $p = .04$, $\eta^2 = .11$, and between race and image transformation, $F(1,35) = 15.0$, $p < .001$, $\eta^2 = .30$. There was also a significant three way interaction between target presence, race, and image transformation, $F(1,35) = 8.3$, $p = .01$, $\eta^2 = .19$. Planned comparisons confirmed that transforming African American faces towards the

Caucasian average shape increased correct rejections, $t(35) = 2.0$, $p = .05$, $d = 0.2$, and correct matching judgements, $t(35) = 1.9$, $p = .06$, $d = 0.3$; while transformation of Caucasian faces towards the African American average reduced correct rejections, $t(35) = 5.4$, $p < .001$, $d = 0.7$, but did not influence accurate identification, $p = .8$.

Repeated measures analysis of variance of d' scores, with race and image as factors, found a significant main effect of race, $F(1,35) = 34.7$, $p < .001$, $\eta^2 = .50$. There was no significant main effect of image transformation, $p = .3$, but there was a significant interaction between image transformation and race, $F(1,35) = 13.6$, $p = .001$, $\eta^2 = .28$. Planned comparisons confirmed that Caucasian observers were more sensitive to own race faces ($d' = 4.4$), than to African American faces ($d' = 2.3$), $t(35) = 5.6$, $p < .001$, $d = 0.9$; and while sensitivity to African American images was enhanced by transformation ($d' = 3.0$), $t(35) = 2.0$, $p < .05$, $d = 0.3$, perception of own race images was significantly impaired ($d' = 3.3$), $t(35) = 3.4$, $p = .002$, $d = 0.5$.

Repeated measures analysis of response bias found significant main effects of race, $F(1,35) = 5.3$, $p = .03$, $\eta^2 = .13$, and of image transformation, $F(1,35) = 5.5$, $p = .03$, $\eta^2 = .14$, but no significant interaction, $p = .2$. Response criterion was more conservative for own race Caucasian faces, $c = -0.17$, than for African American faces, $c = -.45$, and transformation towards the other race shape substantially lowered criterion for all images. Interestingly, while this is evident in more false positives for Caucasian images, accuracy still improved for the other race images.

Discussion

In the face memory literature sequential line-ups are associated with lower false identification, but contingent reductions in correct identification as overall selections are reduced (Wells, 1984; Lindsay & Wells, 1985; Lindsay & Bellinger, 1999; Smith, Stinson, & Prosser, 2004). It was predicted that in the combined procedure individual judgements in the sequential phase would reduce false positives, while fewer items in the simultaneous phase would also improve accurate identification. The combined procedure was successful for own race Caucasian and cross race African American face matching; accuracy for both the original images and the race-transformed images was considerably better than previously observed in Experiments 19 and 20, indicating that it is possible to exploit absolute judgements while maintaining the benefits of multiple featural comparisons.

Within a procedure that combined sequential and simultaneous matching arrays, perceptual expertise predictions of ORB were again fully supported: transforming African American faces towards the own race average shape enhanced discrimination and performance, while transforming own race Caucasian faces towards an African American average shape reduced sensitivity and criterion and resulted in higher levels of false identification.

General discussion

Discrimination and face matching performance was consistently poorer for African American faces than for own-race Caucasian faces, and the criterion for accepting images as a match was less conservative throughout Experiments 19 - 21. The perceptual expertise account of ORB proposes that inexperience with other race facial characteristics causes poor discrimination (MacLin & Malpass, 2001) and as such, if other race images varied along familiar own race shapes and dimensions, identification would be substantially improved. The race-transform quantified the difference between an average Caucasian face shape and an average African American face shape. This was used to alter faces in a way that retained identity and racial markers (i.e. skin tone), but portrayed typical dimensions and variation of a contrasting race. Transformed African American faces remained overtly African American, but the shape of the facial information approximated familiar Caucasian dimensions. The transform thus enabled a study of perceptual expertise without memory, or confound of racial categorisation (e.g. MacLin & Malpass, 2001).

Perceptual expertise dictates that transformation of own race faces towards the facial variation of another race would impair discrimination and cause errors; this was observed in all matching procedures. A complimentary effect should show that transformation of African American faces towards own race dimensions enhances discrimination: this was observed when faces were regarded sequentially, providing evidence that perceptual expertise does contribute to ORB in face perception. However, original African American faces were matched better within simultaneous arrays. This contrast was not observed with own race faces, and is consistent with the proposal that other race faces are processed in a less holistic manner than is associated with own race expertise (Rhodes et al., 1989; Tanaka et al., 2004; Michel et al, 2006a, 2006b). Simultaneous presentation of multiple faces disrupts holistic face processing (Palermo & Rhodes, 2002), and would thus be most effective for featural discrimination of the original other race images: while prompting images to be processed in a sequential holistic way will only be beneficial if they vary according to familiar holistic dimensions.

The best levels of performance were achieved within the combined matching procedure when sequential presentation enabled perception of familiar holistic dimensions to be most effective, and simultaneous presentation of the most similar items facilitated featural comparison. These results support both perceptual expertise *and* featural processing accounts of ORB, and provide evidence that these theories are interdependent. Ability to utilise configural information effectively is dependent upon familiar facial variation; where facial variation has not been learned, holistic processing will be ineffective. Given that cross race identification can be poor; these findings suggest that race specific face transformations have the potential to improve accuracy in applied settings. However, it must be shown that the effects would be consistent across different races of face.

Evaluating ORB in Caucasians with Japanese faces

In the previous experiments transformation of own race faces towards an average African American face shape significantly impaired face discrimination; while transformation of African American faces to the own race average shape enhanced performance if the procedure was conducive to holistic face processing. If the task fostered featural comparison, discrimination of unaltered other race faces was better. To determine whether these effects would be consistent with other race faces in general, Experiments 22 - 24 evaluated matching of Japanese face images by Caucasian observers.

Experiment 22

Matching original and race transformed Japanese faces in simultaneous arrays

Experiment 22 evaluated discrimination and face matching of original and race-transformed Japanese faces within simultaneous arrays. Unfortunately, the Japanese data set could not be matched with Caucasian face images that were captured under the same lighting and camera conditions. If different face sets were employed within a repeated measures matching task it is likely that the Psychomorph race transformation would exaggerate image characteristics rather than facial variation, thus confounding comparative measures of sensitivity and bias. For this reason the following evaluations exclusively explore perception of Japanese faces by Caucasian observers.

Participants

Thirty-five Caucasian participants were recruited from visitors to Glasgow Science Centre. Fifteen were male, and ages ranged from 17 to 77 years with a mean of 39.5 (*s.d.* 14.7). All had normal or corrected to normal vision and none received payment.

Materials

Race transformed images were generated for a set of male Japanese faces $(n = 80)$ by generating a Psychomorph template for each image, from which the average Japanese shape reference was computed. Each individual image could then be transformed 100% away from the Japanese average shape towards the Caucasian average shape reference created for Experiment 19. Face matching arrays were generated for each Japanese target[6]. Probe images showed each target in a frontal smiling pose wearing street clothes. Arrays were constructed using photographs that bore resemblance to the targets shown in frontal pose with neutral expressions. As the same clothing was worn in both the neutral and smiling images, neutral items were cropped to remove clothing. All other details are as per Experiment 19. The items were arranged in two rows and numbered 1-8, and on target present trials, target position was equally sampled. The head within each image gave a viewing angle of approximately 2.5° and within race-transformed arrays both the target and the array items were altered using the same transformation and reference images. An example of the race-transformation with Japanese face images is shown in Figure 6.5, and an example of a simultaneous Japanese face matching array is shown in Figure 6.6.

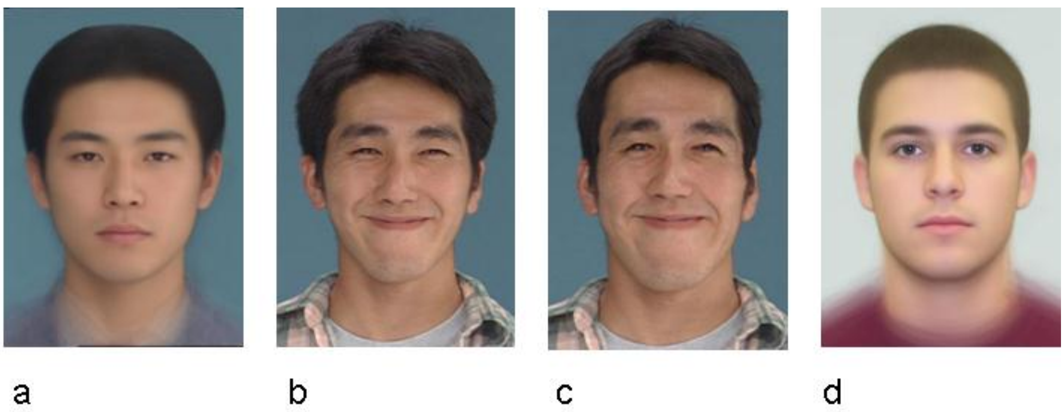


Figure 6.5. The average templates were used to quantify the difference between (a) an average Japanese face shape and, (d) an average Caucasian face shape. The difference could then be applied to each Japanese face image: (b) an example of an original Japanese image; (c) the Japanese face image when it is transformed 100% away from the average Japanese face shape towards the average Caucasian face shape.

[6] Line-ups were constructed for all eighty targets but to fully counterbalance a design within an experiment of acceptable duration for Glasgow Science Centre, pilot testing with Caucasian participants was employed to identify a subset of the 20 most difficult arrays.

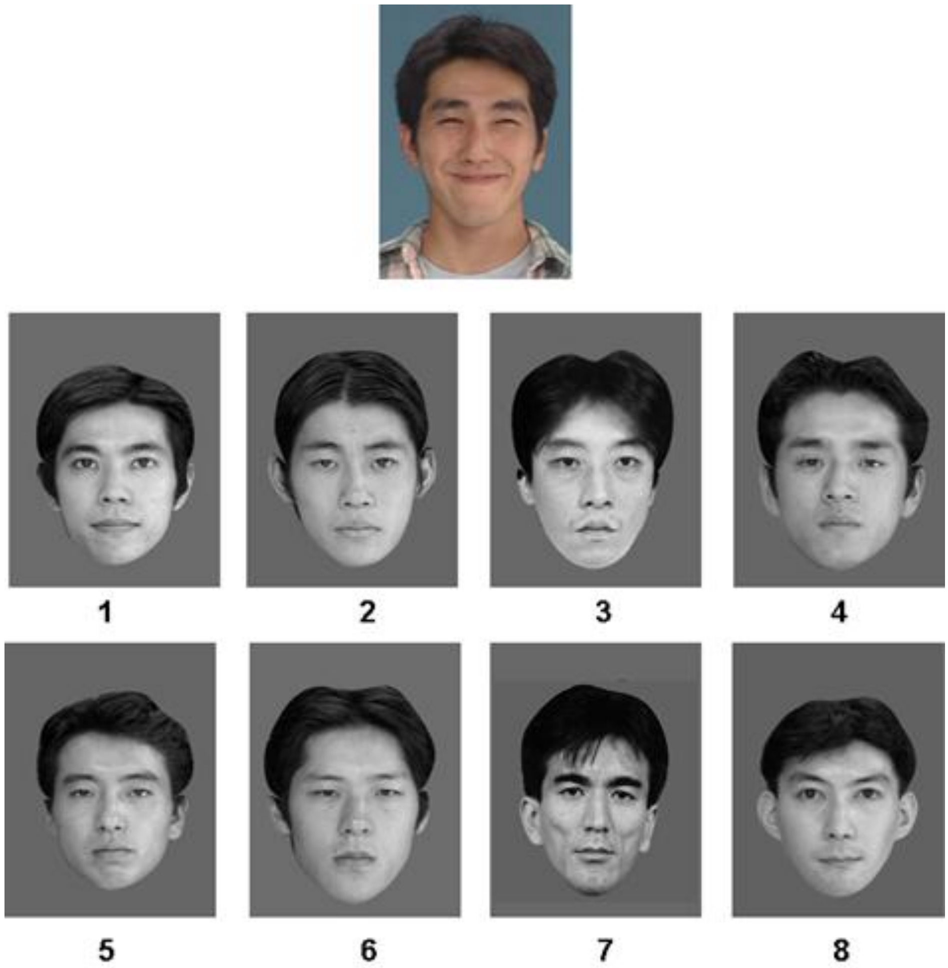


Figure 6.6. An example of a simultaneous face matching array comprising original Japanese faces. The target is shown smiling and wearing street clothes. The array shows eight Japanese males with neutral expressions and clothing information removed. Within race-transformed arrays the target and the array items were altered using the same transformation and reference images.

Design

A 2 x 2 repeated measure design employed target and image type (original images; race-transformed images) as factors. Half of the arrays portrayed race-transformed images, and the target was present in half of each image set. Conditions were counterbalanced across participants and presentation of trials was randomised for each participant.

Procedure

Participants were tested individually within an enclosed demonstration area at Glasgow Science Centre and the experiment was conducted using E-Prime software on a Dell Inspiron 6400 laptop with a 17 inch LCD monitor at 1024 x 786 pixels resolution. Instructions were provided both verbally and with on screen commands: participants were informed they would complete a face matching task and that on each trial a second picture of the target may or may not be shown in the line-up array. Participants were told that there was no time limit and to identify a match from the array they should type the corresponding number; if no match was present they should press the space bar.

Results

The means and standard deviations are provided in Table 6.4 and are broken down by types of response. Accuracy was poorer than previously found with African American and Caucasian arrays, but as the arrays were selected on the basis that they were difficult, this is not evidence of ORB. Replicating the effects that were found with the African American images, transforming Japanese faces towards an own race average shape reduced face matching performance.

Correct face matching scores were analysed with a repeated measures analysis of variance with factors of target presence and image type. There was a significant main effect of target presence, $F(1,34) = 18.8$, $p < .001$, $\eta_p^2 = .36$, and a significant main effect of image transformation, $F(1,34) = 6.6$, $p = .02$, $\eta_p^2 = .16$, with no interaction between these factors, $p = .6$. Accuracy was significantly better when the target was present, 52.9% (*s.e.* 3.4%), with high levels of false positive identification when a correct match was not available, 66% (*s.e.* 3.4%). Overall, performance was not high 47.4% (*s.e.* 3.2%), but when Japanese images were transformed towards the Caucasian average shape, differentiation was significantly worse, 39.4% (*s.e.* 2.8%). Evaluation of signal detection measures of discrimination and bias found that sensitivity to Japanese images (d' = -.2) was reduced by transformation towards the Caucasian average shape (d' = -1.1), $t(34)$ = 2.6, p = .01, d = 0.5, although response criterion (c = -.25) was not significantly reduced by the race-transformation (c = -.47), p = .1.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
Japanese	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Veridical	56.0	*25.6*	16.0	*15.9*	28.0	*22.9*	38.9	*29.1*
Transformed	49.7	*21.3*	13.1	*15.3*	37.1	*22.3*	29.1	*24.4*

Table 6.4. Evaluation of cross race Japanese face matching within simultaneous arrays. Results are provided for veridical and race-transformed arrays and are broken down by percentages of hits, misses and false identification for target present arrays, and correct rejection of target absent arrays.

Discussion

Face matching rates and levels of sensitivity were low in comparison with the African American images, but as the arrays were not equated for difficulty and only a subset of difficult Japanese arrays were employed, absolute levels may reflect a harder stimulus set in this experiment. The Japanese face matching results did, however, replicate the effects observed in Experiment 19; transforming other race African American and Japanese faces towards an own race average Caucasian face shape impaired discrimination within simultaneous matching arrays.

Simultaneous arrays impair holistic face processing (Palermo & Rhodes, 2002) and thus enable featural comparison: these results confirm that they are more effective for discrimination of original other race faces than faces transformed to show own race dimensions. However, the proportion of variability that can be explained by the race-transform was much lower for the Japanese faces (η_p^2 = .16) than for the African American faces (η_p^2 =.34): if the race-transform makes featural discrimination more difficult by reducing the distinctiveness of other race characteristics, this might indicate that the Japanese faces were less distinctive or discriminable than the African American images.

Experiment 23

Matching original and race transformed Japanese faces in sequential arrays

Experiment 23 evaluated the effect of the race-transform on discrimination of Japanese faces within a sequential matching procedure. When the array items were presented individually, discrimination of race-transformed African American images was significantly better than discrimination of the original images and it was predicted that comparable effects would be observed with original and race-transformed Japanese images.

Participants

Seventy-four Caucasian participants were recruited from visitors to Glasgow Science Centre. Thirty-seven were male and ages ranged from 18 to 70 years with a mean age of 36.5 (*s.d.* 11.3). All had normal or corrected to normal vision and none received payment.

Materials and design

The materials from Experiment 22 were used to generate sequential face matching arrays that consisted of the probe image shown to the left of each array item. All other details replicate Experiment 20: each pair of images was separated by a distance of 125 mm, each item measured 50 mm x 80 mm and each head subtended a viewing angle of approximately 2.5 °. A repeated measure design employed target presence and image type (original images; race-transformed images) as factors. Half of the arrays portrayed original Japanese images, and half the race-transformed images with the target present in half of each set. All conditions were counterbalanced across arrays and participants.

Procedure

Participants were tested individually within an enclosed demonstration area at Glasgow Science Centre (apparatus as before). Each person received instruction both verbally and visually with on screen commands: they were informed that they would be shown pairs of faces and should decide whether the images were of the same person or of different people. The participants were advised to respond in their own time, that a number of faces could be shown for each target, and that there may or may not be a matching face in the set. If they judged the faces to be of different people they were to type ‘n’, but if they thought the images were of the same person they were to type ‘y’. The arrays were presented in randomised order and the target position was equally sampled. Within an array the number of array items was not disclosed, each item was shown until a decision was made and decisions could not be changed (cf. Lindsay & Wells, 1985).

Results

The means and standard deviations are shown in Table 6.5 and are broken down by type of response. Transforming Japanese images towards the average Caucasian shape did not produce any noticeable effect, and thus failed to replicate the effects reported in Experiment 20. Accuracy for Japanese images was again poorer than previously observed with African American and Caucasian arrays.

	Hits		Target Present Miss		False ID		Target Absent Correct	
Japanese	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Veridical	47.3	*28.4*	11.9	*17.8*	40.8	*32.2*	39.7	*32.3*
Transformed	47.9	*23.1*	15.4	*18.0*	36.8	*24.6*	37.0	*24.4*

Table 6.5. Evaluation of cross race Japanese face matching within sequential arrays. Results are provided for veridical and race-transformed arrays and are broken down by percentages of hits, misses and false identification for target present arrays, and Correct rejection of target absent arrays.

Correct face matching scores were analysed using a 2 x 2 repeated measures analysis of variance. The results showed a significant main effect of target presence, $F(1,73) = 7.6$, $p = .01$, $\eta_p^2 = .09$, but no significant main effect of image transformation, $p = .6$, and no significant interaction between these factors, $p = .5$. Accuracy was significantly better when the target was present, 47.6% (*s.e.* 2.4%); when a correct match was not available the Japanese arrays again generated high levels of false positive identification, 61.6% (*s.e.* 3.2%). Overall performance for sequential Japanese arrays was poor, 43.5% (*s.e.* 2.9%) and it was not improved when images were transformed towards the average Caucasian shape, 42.4% (*s.e.* 2.3%). Paired sample comparison of *d'* prime scores showed that sensitivity to original Japanese images (*d'* = -.7), was not enhanced by transformation towards an average Caucasian shape (*d'* = -.53), $p = .5$; while analysis of response bias found that the response criterion for the Japanese images (*c* = -.23) was also unaffected by race-transformation (*c* = -.24), $p = .9$

Discussion

In contrast to predictions the Japanese sequential face matching results did not replicate the effects observed in Experiment 20: transforming African American faces towards an own race average shape significantly enhanced discrimination, and transforming Caucasian faces toward the average African American shape made performance much worse; but performance for Japanese images was generally poor and failed to show any influence of the race-transform manipulation. However, it should be noted that although transforming Japanese faces towards an average Caucasian face shape did not enhance discrimination, the race transformation did not impair performance.

The results of Experiment 20 indicated that own race Caucasian and African American faces are processed differently and that simultaneous and sequential procedures can elicit different discrimination strategies for other race faces. As the race transform did not impair discrimination in the sequential procedure but did in the simultaneous format, these results are consistent with this theory. Because of experimental constraints (Glasgow Science Centre), the Japanese face matching materials were restricted to a small set of arrays and were selected to avoid ceiling effects. It seems likely that this control produced arrays that were less distinctive and contained less variability than the African American and Caucasian face sets, and which may have weakened any effects of transformation. It is thus possible that with a more variable Japanese face set enhanced performance for race-transformed images might be observed. The final experiment explored the race transform and Japanese face matching within combined simultaneous and sequential arrays and was intended to replicate the effects observed with African American images in Experiment 21.

Experiment 24

Matching original and race transformed Japanese faces in combined arrays

Experiment 26 assessed the potential of the combined sequential and simultaneous face matching procedure to facilitate the race-transform and maximise Japanese face discrimination. Within this format transforming African American faces towards an own race average shape enhanced performance, while transforming own race faces towards the African American average shape resulted in higher levels of false identification. Given the results of Experiments 22 and 23 it was predicted that the effects of the race transform in African American images would not be replicated with Japanese faces, but that the race transform would not impair Japanese face discrimination and overall performance for Japanese images might also be enhanced.

Participants

Forty-two visitors to Glasgow Science Centre volunteered to act as participants. Fifteen participants were male, and ages ranged from 20 to 62 years with a mean of 39.4 (*s.d.* 11.5). All had normal or corrected to normal vision and none received payment.

Materials and design

Combined sequential and simultaneous arrays were generated from the materials employed in Experiments 22 and 23, and the procedure is a replication of Experiment 21. Within the sequential phase the target was shown with each array item until a response was provided and when each item had been viewed, potential matches were displayed simultaneously. Materials, design, and apparatus as previously described.

Procedure

The participants were tested individually within an enclosed demonstration area at Glasgow Science Centre. Instructions were provided verbally and visually with on screen commands. Participants were told that they would see pairs of faces and should decide whether the images could be of the same person, or were of different people. If they judged the images to be of different people they were to type ‘n’, but if they thought it possible that the images were of the same person they were to press the space bar. Once a response had been provided for each item, the target image was shown above the possible matches and participants were asked to make a face matching decision. At this time they were advised that there may or may not be a matching face within each set, to select an image they should type the corresponding number, but to reject all of the items they should press the space bar.

Results

The means and standard deviations by condition are shown in Table 6.6. Accuracy for Japanese images was again poor in comparison with the results for African

American faces and there was no apparent effect of race-transform for Japanese faces within the combined procedure.

	Target Present						Target Absent	
	Hits		Miss		False ID		Correct	
Japanese	M	*S.D.*	M	*S.D.*	M	*S.D.*	M	*S.D.*
Veridical	59.5	*28.5*	13.3	*22.8*	27.1	*22.0*	40.0	*31.2*
Transformed	58.1	*27.9*	16.2	*23.1*	25.7	*23.5*	35.7	*31.7*

Table 6.6. Evaluation of cross race Japanese face matching within combined sequential and simultaneous arrays. Results are provided for veridical and race-transformed arrays and are broken down by percentages of hits, misses and false identification for target present arrays, and correct rejection of target absent arrays.

Correct matching scores were analysed with a 2 x 2 repeated measures analysis of variance with target presence and image transformation as factors. There was a significant main effect of target presence, $F(1,41) = 13.2$, $p < .001$, $\eta^2 = .24$, but no significant main effect of image transformation, $p = .4$, and no significant interaction, $p = .6$. Accuracy was better when the target was present. The original Japanese arrays elicited an overall accuracy rate of 49.8% (*s.e.* 2.9%), which was not altered by transformation towards the average Caucasian shape, 46.9% (*s.e.* 3.5%). Paired sample comparison of *d'* scores found that sensitivity to original Japanese images ($d' = .19$) was not affected by transformation towards the average Caucasian shape ($d' = -.33$), $p = .5$; while analysis of response bias showed that response criterion for the Japanese images ($c = -.45$) also did not change ($c = -.56$), $p = .2$.

Discussion

In line with predictions the race transform did not impair discrimination of Japanese faces, but performance was not enhanced. The results therefore failed to replicate the identification advantage observed with race-transformed African American faces in Experiment 21. The combined matching procedure was successful for Caucasian and African American face discrimination but the prediction that Japanese face matching would also be enhanced was not supported: accuracy was somewhat better than observed with the sequential procedure but comparable with Experiment 22. In summary it was not possible to exploit own race holistic face perception with this particular set of images.

Summary and conclusions

The aim of this chapter was to evaluate a novel cross race transformation within different face matching procedures, and to determine whether perceptual inexperience with unfamiliar races contributes to ORB. Face matching ability was better for own-race Caucasian faces than for African American faces, and performance with Japanese faces was poor in comparison with African American faces. However, the sets of arrays were not rated for distinctiveness and were not equated for difficulty; therefore, overall levels of accuracy cannot be regarded as evidence of own race bias.

Theoretical considerations

Perceptual expertise was explored by way of a race-transform that manipulated face images to portray the facial dimensions and variation of a contrasting race. In line with predictions, transforming own race faces towards an average African American face shape impaired discrimination in all procedures, but to support perceptual learning theories, transforming other race faces towards own race dimensions must also enhance discrimination. These predictions were only partially supported: the race-transform enhanced discrimination of African American faces in sequential and combined matching procedures, but original African American faces were matched better in simultaneous arrays. This suggested that the holistic race-transform was only beneficial when the faces were inspected individually, and that own race and African American face matching was accomplished in different ways.

The perceptual encoding account of ORB proposes that other race faces are processed in a more featural manner than own race faces (Rhodes et al.,1989; Tanaka, Kiefer, & Bukach, 2004; Michel et al, 2006a,2006b). As simultaneous arrays impair holistic face processing (Palermo & Rhodes, 2002), it is consistent with this theory that this procedure would be most effective for featural discrimination of original African American images; and as observed, to enable expert holistic perception of the race-transformed images, it would be necessary to regard images individually. It was concluded that perceptual expertise contributes to ORB but that other race faces will be processed in a featural manner; moreover, prompting images to be processed in a holistic way will only be beneficial if they vary according to familiar facial dimensions. The theories are therefore interdependent: because we lack perceptual expertise with unfamiliar facial variation, we are unable to process these images effectively in a holistic manner.

Evaluation of Japanese face matching intended to replicate the results observed with African American images, but once again the predictions were only partially supported. In line with African American face matching the race-transform impaired Japanese face discrimination within the sequential procedure, but within the sequential and combined matching formats no effect of race-transform was observed. While this failed to replicate the race-transform advantage in African American face matching, performance was not impaired and perception of the own race holistic dimensions was enhanced relative to the simultaneous format. As with African American face matching, the results indicate that Japanese images were processed differently in simultaneous and sequential matching procedures: unaltered other race images will be matched more efficiently within a simultaneous multiple comparison format and individual analysis will enable extraction of own race holistic dimensions from race-transformed images.

A question remains regarding why no advantage was obtained for race-transformed Japanese images within the sequential matching procedures and a number of possible explanations should be addressed. Categorisation of identical faces on the basis of race cues has been shown to affect perceptual judgement and memory (e.g. Pauker & Ambady, 2009) and indications that other race faces were processed in a different manner to Caucasian faces suggests that racial categorisation will determine processing strategy. Categorisation of race will be influenced by how different the other race face is perceived to be. Race-transformed African American images retained their skin tone and remained unequivocally other race, but it is possible that transformation of Japanese faces would have generated items that were more ambiguous and may have diluted the effects. See figure 6.7 for a comparison of original and race transformed images.

Figure 6.7. An example of (a) an original African American image; (b) the race-transformed African American image; (c) an original Japanese image; (d) the race-transformed Japanese image; (e); an original Caucasian image; and (f) the race-transformed Caucasian image.

Alternatively, distinctiveness and the inherent difficulty of the face matching arrays may have created apparent differences. The Japanese face matching data was collected at Glasgow Science Centre and materials had to be restricted to just 20 experimental trials *(with larger sample sizes)*; to this end and to avoid ceiling effects, a pilot study determined the most difficult arrays for inclusion in the study and it is possible that this subset were less distinctive and contained less variability than the African American or Caucasian materials. As the race transform could have reduced distinctiveness, discrimination of the images would be more difficult and would weaken any observed effects. To establish the true nature of Japanese face matching with original and race-transformed images, it would be necessary to repeat this study with a larger more variable array set and to employ ratings to control for perceived distinctiveness.

A final consideration concerns the image characteristics of the different face sets. Any transformation of visual characteristics will be dependent upon image quality, and image capture conditions. The African American and Caucasian images were captured in identical conditions and we can conclude that the transformation between these averages has a high probability of reflecting racial differences. The Japanese face set were obtained from a different source and cannot be matched for image characteristics. Essentially this means that the image transformation between Japanese and Caucasian averages will also reflect a degree of imaging noise, and thus may not reflect true racial differences.

Own race bias and Multi-Dimensional Face Space

The race-transformation employed in this study evolved within the concept of multi-dimensional face space (MDFS) (Valentine, 1991) and perceptual expertise theories of ORB. Perceptual expertise dictates that ability to discriminate faces is determined by experience and that infrequent exposure to other race faces means we don't learn the range of facial variation needed to distinguish them (MacLin & Malpass, 2001). Within a MDFS other race faces will be distinctive and will be encoded in a race specific cluster some distance from the central tendency of own race faces (Rhodes & McLean, 1990). The faces will be close together and appear more similar than faces that are further apart and will thus be difficult to differentiate. Evaluation of MDFS typically employs recognition tasks (e.g. Byatt & Rhodes, 1998; Stevenage, 1995) but stored knowledge should provide the basis from which to interpret any novel signal, such as discrimination of unfamiliar faces. If the concepts of MDFS and perceptual expertise are correct, the organisation of face memory will not be suitable to differentiate other race faces, but by making them vary in the same way as own race images they would be shifted to the area of learned variation and become easier to discriminate.

Performance for race-transformed Caucasian images consistently declined and discrimination of transformed African American images improved within sequential and combined procedures, providing support for both MDFS and perceptual expertise predictions. However within the simultaneous procedure, matching ability declined for both transformed Japanese and transformed African American images indicating that a purely perceptual account of ORB is not supported and that cognitive allocation of processing strategies may also be involved.

Applied considerations

The UK Home Office Scientific Development Branch identifies Facial Image Comparison (FIC) as a growth area in the security sectors, and is in the process of developing innovative training methods for live face matching, face image comparison, and biometric application use for security personnel. The work presented here is therefore important, and at the most basic level can foster awareness that cross-race identity verification will produce comparatively higher rates of error than within-race authentication. To appreciate the potential impact, at border controls the risk is greater that a person of another race will be able to use a fraudulent passport; conversely there is also a higher probability of ethnic minorities being erroneously linked to CCTV or security footage. Both types of error are costly but understanding the cognitive factors may ameliorate such effects.

Security services maintain surveillance and assimilate security footage with existing databases; it is here that the work presented in this chapter is most applicable. Cross race face matching is prone to error (Megreya et al., 2011; Sporer et al., 2007) but can be improved with knowledge and understanding of the causes and effects. For example, security personnel tasked with identity verification from official documents can be trained to individuate faces of another race (e.g

Hugenberg, Miller, & Claypool, 2007; Lebrecht, Pierce, Tarr, & Tanaka, 2009; Rhodes et al., 2009), while an understanding that featural comparison is more effective for cross race judgements can induce personnel to attend to facial elements rather than to the images as a whole (Rhodes, Hayward, & Winkler, 2006; Rhodes et al. 1989; Michel et al., 2006; Tanaka et al., 2004; Michel, Caldara, & Rossion, 2006). What is more, in a social climate that strives to avoid racial discrimination, education and research is necessary to support adoption of race as a factor in personnel selection and training.

The scale of modern security risks have fuelled advancement of biometric face recognition, but while increasingly effective, they can only display potential matches and the task remains subject to human bias. In the future biometric applications will become more common creating a demand for operators to intensively perform identity checks with the assistance of a computer interface. Within the biometric face models race-transforms will be able to alter other race faces to approximate the racial variation of the viewer, and there is a real potential for such manipulations to enhance the safety of matches. Before such implementations can be adopted, however, it remains to be shown that effects are consistent.

The face sets employed in this study were not equated for distinctiveness, difficulty, or image quality; therefore it is not possible to draw strong conclusions regarding the effect of the race–transform across different races. Future work should control for distinctiveness (as indexed by the race of observer) and match other race arrays for task difficulty. It is also possible that other race characteristics must be distinctive for a race transform to be effective, or that given a preference for featural discrimination a caricature manipulation might enhance other race discrimination more efficiently. Image characteristics will continue to undermine accuracy in face matching judgements; however, with the introduction of 3-dimensional face models, it may be possible to reduce many such effects. Finally, it would be useful to study face discrimination strategies and ability of racial minorities. If they do individuate majority out group faces by expert holistic face processing they might apply holistic processing to all face images, in which case one would expect superior face matching ability overall.

Conclusions

Perceptual expertise was important for own race face perception and discrimination of Caucasian faces was impaired when the images were transformed towards African American facial attributes. The same pattern of results was obtained in all face matching tasks, but performance was best within the combined sequential and simultaneous procedure. Perceptual expertise also influenced discrimination of African American images, and transformation towards own race variation was successful when the images were matched one at a time. Conversely, original African American images were matched better within simultaneous arrays, indicating that the faces were processed in a more featural manner. When the procedures were combined, performance for both original and race transformed images improved. The same pattern of results was obtained for Japanese face matching: discrimination of unaltered images was best within a simultaneous procedure and perception of race transformed images was better within the sequential process. However the effects were much weaker and the transform did not enhance rates of cross race identification.

Overall, the combined procedure was most successful for face discrimination, indicating that it is possible to exploit absolute holistic judgements while maintaining the benefits of multiple featural comparisons. It was concluded that although perceptual expertise contributes to ORB, other race faces will be processed in a featural manner and prompting images to be processed in a holistic way will only be beneficial if they vary according to familiar facial dimensions.

7

General Discussion

Forensic issues around recognition of unfamiliar people have received considerable attention, yet perception and discrimination of unfamiliar faces has been relatively overlooked. The present research examined the effects of face perception on identification of facial composites, and discrimination of unfamiliar faces in facial image comparison procedures. The introduction provided an overview of how faces are typically perceived and in what ways familiar and unfamiliar face processing differ. Two influential models of face memory were described and have been referred to throughout this work; these were employed to conceptualise how unfamiliar face images may be interpreted, and how the nature of such representations might be manipulated to enhance perception and identification. A brief review of the development of UK forensic identification guidelines provided an applied frame of reference within which perceptual discrimination of unfamiliar faces has yet to be addressed. This chapter describes the main findings of this research, suggests areas for future study and offers some recommendations for unfamiliar face perception in forensic applications.

Facial composite perception and identification

The split facial composite effect

The differences in the way that familiar and unfamiliar faces are perceived have important consequences for forensic facial identification, and the efficacy of facial composites, in particular. Faces are processed holistically and for familiar faces there is great sensitivity to the configural properties (Haig, 1984). This, however, means that recognition of facial features can be impaired if a facial configuration is altered in the production of a facial composite (Tanaka & Sengco, 1997; Young et al., 1987). Facial composites must be recognised by familiar people, but as they are pictorial representations of unfamiliar face memories, they will be reconstructed from a limited, context specific memory, and will, to some extent, be inaccurate.

Chapter 2 investigated whether inaccuracies in the configural arrangement of a facial composite could interfere with identification of accurate composite features. It was shown that configural inaccuracy does significantly impair identification; but that presenting the composites in a way that prevents holistic face perception, will enable identification of accurate composite elements. Replication of the effect in this work, and by another researcher with a different set of composites (Frowd, unpublished), confirms that the split composite advantage is reliable. Simply splitting the face images was not sufficient to enhance recognition, the face parts also had to be misaligned, causing the images to be interpreted in a slower non-holistic process.

Facial composites portray imperfect reconstructions of unfamiliar face memories, and will be difficult to match with familiar face representations that are stored in memory. Holistic perception of an inaccurate facial configuration will impair identification of any accurate composite elements, but when the image is split and misaligned, holistic analysis will be precluded and the facial features can be extracted and matched against stored face representations, with a greater chance of recognition. The results suggest that police forces should consider employing split and misaligned facial composite images. They might be reluctant to employ split images instead of conventional composites, but an evaluation by Frowd (unpublished) found that if split composites were shown together with the complete composite images, participants attended to the original image and there was no identification gain. A future experiment could consider the potential of showing a split image a short time before a complete composite, or on newspaper-like pages that are not viewed at the same time. However, Singer and Scheinberg (2006) demonstrated that when face halves were viewed consecutively the face image was perceptually reconstructed; therefore, temporal and spatial distance between a whole composite and a split composite would be necessary to produce the desired effect.

A second forensic application of the split composite effect may also be found in the composite construction process. Featural composite systems require the witness to select features within the context of a whole face (e.g. Davies & Christie, 1982; Tanaka & Sengco, 1997), but it would be useful to determine if holistic perception during construction is really beneficial, or whether features should be selected in isolation before generating an overall composite likeness. Police artists employ feature selection from a manual before generating a sketch, and can produce better likenesses than the featural composite systems (Frowd et al., 2004; Gibson, 2008). It might be possible to draw on both processes and present all of the composite information in a split format that would provide facial context and enable some spatial relations to be beneficial, but would avoid the dominance of holistic face perception. It could also be beneficial to 'toggle' between holistic and split face perception, such that splitting would enable the witness to attend to featural information, but the holistic phase would allow them to combine the elements as effectively as possible. As a final consideration, Gibson (2008) notes that a common error in face image reconstruction is to exaggerate nose length and the distance between the nose and the mouth, it would be useful to determine which configural flaws are most common in facial composite generation as such errors could then be addressed within the construction process.

The positive facial composite effect

When facial identification is not easy, affective information will influence recognition judgements, such that smiling faces will be perceived as more familiar (Baudouin, Gilibert, Sansone & Tiberghien, 2000; Gallegos & Tranel, 2005; Kottoor, 1989; Davies & Milne, 1982; Endo et al. 1994), and faces with negative expressions as less familiar (Lander & Metcalf, 2007). Facial composites are imperfect representations and recognition by familiar people will be difficult. Chapter 3 therefore aimed to establish whether affective information was also important for facial composite identification, and whether affective quality could be enhanced to improve familiar face perception.

A subtle smile transformation was on the threshold of detection yet significantly improved identification; therefore, affective signals are important for facial composite recognition, and facial expression need not be explicit to exert a powerful influence on identification judgements. The effect was also much larger for composites of personally familiar people, which indicated that enhanced identification was less likely to be mediated by memories for smiling faces, than by an association between positive affect and familiarity. What is more, false identification of facial composites did not increase, suggesting that the memory representation of a specific person was activated before perception of positive affect influenced a recognition response. The results obtained for explicit expressions were also somewhat different: at threshold levels of expression, perception was enhanced only for manipulated images; while at explicit levels of expression, any face image with the potential to signal familiarity may be better attended, which raises the possibility that the smiling face effect might operate at two distinct levels.

Facial composites are difficult to match with representations of familiar faces in memory; therefore, affective information will be an important factor in achieving an identification judgement (Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2004; Kaufmann & Schweinberger, 2004). Absence of positive affect will signal a lower probability of familiarity, reducing attention and the likelihood that a name will be offered (Lander & Metcalfe, 2007). The Bruce and Young model (1986) proposes that expression and identity are processed in separate parallel routes, but there is substantial evidence that these processes can interact (e.g. Baudouin, Gilibert, Sansone & Tiberghien, 2000; Gallegos & Tranel, 2005; Kottoor, 1989; Davies & Milne, 1982; Endo et al. 1994). Endo et al (1992) reported that familiarity decisions for unfamiliar faces took longer if they were smiling, and Ganel and Goshen-Gottstein (2004) found that expression interfered with identification judgements if discrimination was difficult. These findings are consistent with a dual process feedback mechanism, which *would* be permissible within the Bruce and Young (1986) model. At a basic level, positive affect may signal familiarity and be sufficient to prolong structural analysis of the face image, but as difficult recognition takes longer, this would allow time for affective feedback from an expression analysis route to supplement sub-threshold activation at person identity nodes (PINs) *(see figure 1.1).*

While it is beyond the scope of this research, it is useful to consider neural mechanisms that would facilitate such interactive processes. Studies have shown that expressions of fear activate the amygdala more strongly than happy expressions (Morris et al. 1996), or neutral expressions (Breiteret al. 1996a), and novel faces also produce stronger activation than familiar faces (Schwartz et al., 2003). Although speculative, these findings could indicate that reduced amygdala activation accounts for the smiling face bias in familiarity perception. In conditions of doubt, a smiling expression would reduce activation to a novel face, thereby increasing perceived familiarity. Likewise, negative affect (Lander & Metcalfe, 2007) would increase amygdala activation and reduce perceived familiarity. Where identification is not difficult, affective feedback connections will not influence judgements; but in facial composite recognition, secondary information would be very important. Facial composites are processed holistically and may be perceived as novel faces. Novel face perception will heighten amygdala responses and if the composites also exhibit negative affect, activation would be even greater and chances of identification substantially reduced. By this reasoning, any reduction of negative affect would lessen the amygdala response and be beneficial for facial composite recognition.

The facial composites employed in this study were not generated by victims of crime and are therefore unlikely to embody negative emotional affect. The results of these experiments strongly suggest that real life composites will suffer more from negative affective bias. While it is important to employ context reinstatement to recreate emotion and enhance memory recall (Geiselman et al., 1987; Ucros), and it is important that the information provided by the witness is accommodated, the witness has a vested interest in the composite image being as identifiable as possible. When a composite likeness has been created, the witness should be asked to allow affective enhancement in order to improve chances of identification. It is likely that level of negative affect, and hence the extent of acceptable transformation, would be specific to each individual circumstance: incorporation of an affective transform within facial composite systems would enable the witness to set a level of manipulation, with which they are satisfied. As extremely subtle transformations were found to produce powerful effects, even modest alterations have the potential to significantly improve rates of identification.

The research that inspired this series of experiments (e.g. Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2004; Kaufmann & Schweinberger, 2004) explored the smiling face bias with explicit facial expressions, and within the expression literature most studies have employed a limited selection of standardised materials that portray unambiguous positive, neutral, or negative affect (e.g. Pictures of facial Affect, Ekman, Friesen, & Press, 1976; Karolinska Directed Emotional Faces, Lundqvist, Flykt, & Öhman, 1998). However, the results reported here demonstrate that barely perceptible expression can significantly alter behaviour, and can produce *different* effects from explicit expression. This means that the reported effects of 'gross' expression may mask more important results, and that our understanding of the effects of emotional transmission may, in fact, be somewhat limited.

The expression transform employed here was generated from the average of one hundred posed smiles and highlights a second problem in the facial affect literature. Most studies have employed a limited range of datasets (see above) that portray posed expressions, or ‘natural’ expressions that are posed according to Ekman’s (Ekman & Friesen, 1976) Facial Action Coding System (FACS). Pilot data collected by the author found that affective responses to posed expressions, measured using facial electromyography, were significantly weaker than responses to spontaneous expressions. The implication is that a smiling face transform would be more powerful if it was based on the average of spontaneous rather than posed smiles. It also suggests that knowledge regarding perception of facial affect should be updated to accommodate reactions and cognitive effects in response to explicit and threshold spontaneous facial expressions of emotion. At this time, there is no suitable face dataset to facilitate this; a preliminary step would be generation of such materials.

Social communication is mediated most effectively by face to face communication, and these results show that even subtle emotional signals will produce important effects. While affective communication is transmitted visually it is also interpreted by mirroring and efferent feedback from facial muscles (Moody et al., 2007). Extreme sensitivity to this information might have important implications for social communication in modern culture. Cosmetic procedures that immobilise facial muscles have become increasingly common and their effect on communication and decision making should be established. For example, communicating with a ‘botoxee’ will provide fewer affective cues, while a ‘botoxee’ will be unable to mirror, or receive, a full ‘dialogue’ of affective information. How this plays out in social interaction and decision making would be of interest. It would also be interesting to assess whether absence of efferent signals will attenuate the social transference of affect?

Facial Image Comparison

Despite being the most common security procedure there has been remarkably little study of accuracy and effects of facial image comparison (FIC) within forensic application (e.g. Henderson et al, 2001, Davies & Valentine, 2010), and at this time there are no official guidelines or training protocols in the UK. In light of research that shows unfamiliar face matching produces high rates of error (Bruce et al., 1999, 2001; Megreya & Burton, 2006, 2007; Kemp et al, 1997) the remainder of this thesis concerns perceptual discrimination of unfamiliar faces and explores methods designed to enhance performance.

The smiling face bias and facial discrimination

Where facial identification requires effort, affective information will influence judgements of familiarity. In Chapter 3 a smiling face advantage in facial composite recognition was associated with increased sensitivity and it was proposed that participants were induced to attend better to images that showed positive affect (cf. Endo et al., 1992). Chapter 4 evaluated whether positive affect could also induce more attention and improve performance in face discrimination tasks.

However, in face matching sensitivity declined, providing evidence that the smiling face bias involves activation of face representations in memory *(or the belief that this has been achieved)*, but is not a direct consequence of perception and attention. Bruce et al. (1999) found that face matching is poorer if the target face and arrays have different expressions. These results show that discrimination will also be worse when all of the images show positive affect, indicating that expression may be distracting and that identification documents and security procedures should employ neutral expressions. Simultaneous comparison of multiple images was also more effective than comparing each item to the target; but while this may be useful when an image is matched to a database, face image comparison is most commonly conducted by comparing one person to one image. Experiment 14 evaluated presentation of a target image, followed by two faces, one showing positive affect. Accurate identification wasn't influenced by expression, but faces showing positive affect were incorrectly identified less often. In summary, although positive affect can enhance facial identification from memory, sensitivity will not be enhanced in perceptual discrimination and accurate face matching judgements may be significantly impaired.

The caricature effect in face discrimination

Caricatures are highly identifiable because they exaggerate distinctive features (Stevenage, 1995) and computerised caricatures have been shown to enhance recognition of familiar faces (Benson & Perrett, 1994; Rhodes et al, 1987). Perceptual discrimination of unfamiliar faces is difficult, and it was proposed that a systematic caricature transform that could increase distinctiveness would also enhance unfamiliar face matching. Within face matching arrays the target and the array images were caricatured relative to the same norm face; if the face had a larger nose than the norm face, this would become even larger, which would distinguish it from the other images. In the concept of MDFS this would increase the distance between each face and the norm, and make them more dissimilar. By exaggerating the faces it was predicted that incorrect identifications would decline, but if the target and matching faces became too dissimilar, correct identification would also be reduced. Signal detection measures of sensitivity were enhanced by all levels of transformation, while response bias became more conservative. Identification of correct matches was not enhanced, and if the caricature was too strong performance was much worse. Correct rejection of target absent arrays was, however, enhanced at all levels.

The results from the simultaneous arrays supported MDFS predictions: discrimination of foils was better at all levels of transformation, although for the 30% transformation, this was confined to the most similar items. When the target was present there was no consistent effect for identification of 30% and 50% caricatures, but at the 70% level of manipulation performance was significantly worse, indicating that the matching items had become too distant in the face space for correspondence to be achieved. However, with the sequential procedure the effects were reversed: identification was enhanced and discrimination of foils was not. This suggests that the images may have been processed in different ways. The simultaneous procedure is associated with multiple featural comparisons, while the sequential format is associated with holistic processing and should therefore be more likely to support perceptual expertise predictions; but it was the opposite pattern that was observed and the results do not sit well with MDFS models. While it is consistent that caricature would increase distinctiveness and produce greater levels of sensitivity, it is not clear why the sequential procedure should reduce levels of response bias when the contrasting effect is typically observed. It would be useful to repeat this experiment in order to establish that this is a reliable effect.

With this particular set of images, caricature enhanced discrimination of non-matching images and was most effective at a 50% level of transform, but an optimal caricature level is likely to be specific to a given set of images and the level of disparity between them. It would be useful to replicate these caricature effects with different image sets and different levels of matching difficulty. The images employed in the experiments were captured on the same day in good lighting but caricature would also exaggerate superficial differences in images captured in more variable conditions. Three dimensional modelling will be able to reconstruct images and match lighting effects, and this holds the solution to such imaging artefacts.

Biometric face recognition applications quantify the correspondence between face images and can caricature face images automatically; a useful focus of future study would be to determine whether biometric correspondence between images can be used to determine the most useful level of caricature. In this way images with close correspondence would be more difficult to discriminate and would benefit from stronger levels of manipulation, while levels of transformation that are too powerful would could also be avoided.

For use within face image comparison the point of reference for caricature transformation should also be explored. Within caricature and distinctiveness studies, the point of reference is always an average face (e.g. Stevenson; Benson & Perrett, 1994; Rhodes et al, 1987). While this enables distinctiveness to be evaluated relative to a norm, for face image comparison the most useful metric is not difference from the norm but difference from the target or probe image. If an array of images is caricatured relative to the target image, differences will also be exaggerated but the resulting disparity should, on balance, be greater for images of different people than for another image of the same person. A target specific caricature could, if effective, be incorporated within biometric systems to enhance both successful rejection of non matching images and identification of those that originate from the same individual.

Sequential face matching was again shown to be less effective than matching with simultaneous arrays but as the most common face image comparison practice involves inspection of individual images, sequential face matching should also be the focus of future study. One avenue of research could explore the utility of caricature to enhance perception of photographic ID. These images are tiny in comparison with the bearer, but because of this disparity, caricature might usefully highlight diagnostic facial characteristics and improve live identify verification.

Own Race Bias in face discrimination

Discrimination of unfamiliar faces is difficult (e.g. Bruce et al., 1999; Henderson et al., 2001), but the task is even harder when the faces are of another race (e.g. Megreya, White, & Burton, 2011). The perceptual expertise theory of ORB was investigated within a series of face matching tasks. African American faces were manipulated to retain identity and skin tone, but to vary along familiar Caucasian facial dimensions; likewise, the shape of Caucasian faces were manipulated to vary along African American facial dimensions. To demonstrate perceptual expertise in own race and other race face processing, it would be necessary to show that discrimination of own race faces is poorer when they vary along other race dimensions, while discrimination of other race faces is enhanced when they vary along familiar own race dimensions.

Discrimination was better for own race faces and making them vary along other race dimensions reduced performance. However, making African American faces show own race Caucasian dimensions was only effective if the images were matched individually, while the original faces were matched better in a multiple item array. These procedures are conducive to different types of face processing:

when a simultaneous procedure was used, discrimination of the original African American faces was better, indicating that perception of other race faces is best within a featural discrimination task, and that perception of own race dimensions would only be effective in a process that promotes holistic face perception. The results support both perceptual expertise and featural processing accounts of ORB, and suggest that perceptual learning of facial dimensions is required to enable effective holistic interpretation. Performance with the Japanese face set was generally poor and no significant effects were found. However, it was notable that the same pattern of effects by procedure and transformation was observed.

MDFS models are constructed based on experience and knowledge of the faces that one encounters, and are therefore not equipped to differentiate faces that differ along unlearned dimensions. These models would predict that making other race faces show own race dimensions would enhance discrimination, while making own race faces vary along unlearned dimensions would produce the opposite effect. For perception of own race faces these predictions were supported, but for other race faces the effects were only observed when a sequential procedure promoted holistic face processing, and the efficacy of holistic face processing depended on the presence of familiar facial variation. The findings are, therefore, problematic for a purely structural understanding of how faces are stored in memory, and suggest that a form of binding for facial information is also important.

These results indicate that in forensic and security settings cross race identity verification will produce comparatively higher rates of error (e.g. Megreya and Burton, 2011; Sporer et al., 2007). An understanding that featural comparison will be more effective for cross race individuation can help operators to attend to facial elements rather than the whole image, but security personnel should preferably receive specific training to improve individuation of own race and other race faces (e.g Hugenberg, Miller, & Claypool, 2007; Lebrecht, Pierce, Tarr, & Tanaka, 2009; Rhodes et al., 2009). In a number of studies it was shown that racial minorities did not exhibit ORB and may therefore be more effective at identity verification. It would be useful to determine whether this only applies to discrimination of the majority race in the populations studied, or whether the effect might transfer to discrimination of faces of any race. In a similar vein, if individuals can be free of race bias, it would be useful to identify face discrimination tests that can identify operators that are not be susceptible to ORB.

Biometric face recognition is increasingly common and there is potential to incorporate face image manipulations that will enhance identification. For example, the models would be able to alter the other race face images to approximate the racial variation of the viewer. The race transformation can be effective, but discrimination will still rely on perceived distinctiveness of the face images, and if this is lacking, race transformation may be ineffective and the use of caricature to increase facial distinctiveness would be better. Future work should replicate these experiments with larger cross race array sets that are equated for difficulty and rated for distinctiveness by participants of the same race. Races or face sets rated high on distinctiveness may benefit from race transformation and a sequential discrimination process, while other race faces rated low for distinctiveness might benefit from caricature within a simultaneous comparison format. Alternatively caricature could be employed to enhance distinctiveness before a race transformation is employed.

Imaging methodology

The theme throughout the thesis has been the use of imaging software to enhance perceptual discrimination of face images. The software can be used to generate reference images by morphing together a number of exemplars; for example a set of female faces might be combined to generate an average female face. The resulting image will be typically female, and the morphing process will have averaged out any distinctive characteristics. Average reference faces were used in several of the studies to determine the image transformations. It should be noted, however, that with the exception of the morphed composites in chapter 2, none of the experimental stimuli were combined with other images.

The composite morphing procedure didn't use averages, but warped the shape of facial composites towards photographs of the people they portrayed. The transform made it possible to evaluate composite identification with accurate configurations, but as morphing would also have improved features to some extent, this may have contributed to the identification improvement. What can be concluded is that when holistic face processing was removed by splitting the images, the morphed composites provided no advantage over original composites; therefore, any enhancement to the individual features was not easy to perceive. When an average reference was used, the simplest manipulation involved warping the shape of individual faces away from the shape of the average face. This exaggerated the way that each image differed from the average, making the faces more distinctive and discriminable.

The other manipulations involved transformation between two average reference images that served as anchors, or end points, of facial dimensions. In the cross race studies the transform quantified how the shape of an average African

American or Japanese face differed from the shape of an average Caucasian face. By warping the shape of individual face images away from the same-race reference face toward the average shape of another race, the shape of the faces was transformed from the typical variation of one race to the other. It is important to note that such transforms do not specify the absolute location of any face image in the perceptual space, only the trajectory in the space of the perceptual alteration. This allowed investigation of the perceptual expertise theory of ORB. Faces should have been more discriminable to Caucasian viewers when they showed the typical variation of Caucasian faces, and less discriminable when showing the variation of another race.

However, the nature of this transform also meant that images transformed towards the own race Caucasian average should become less distinctive, while faces transformed away from the own race average should be more distinctive. Essentially this would entail contrasting predictions to the perceptual expertise account: Caucasian faces transformed towards the variation of other race faces would be more distinctive and discriminable, while other race faces transformed towards the own race average would be less distinctive and discriminable.

For own race Caucasian faces the results consistently supported perceptual expertise theories: although the faces were less like the own race average and would have been more distinctive, they were harder to discriminate. For African American faces the perceptual expertise theory was also supported: faces were more discriminable when they became less distinctive, but only when a sequential procedure was conducive to holistic face perception. When a simultaneous procedure, which is associated with featural processing, was used, the faces were less discriminable, indicating that they were indeed less distinctive. The effects for

Japanese faces showed the same pattern, although discrimination was never enhanced by transformation. Thus indicating that holistic face perception was beneficial, but transformation towards the own race average variation reduced discrimination. The methodology was therefore effective in assessing own race expertise, but would have less efficacy as means of enhancing other race face discrimination.

In the second facial composite study neutral reference faces and smiling reference faces were used to quantify how, on average, a face would change when a person smiled. By warping the shape of individual images away from the neutral average face toward the average smiling face, the image was shifted from one area in perceptual space to another according to the trajectory of the average smile transform. Visually, the face image was altered according to how, on average, it would change if the person smiled.

Conclusions

A substantial body of literature has evaluated unfamiliar *face recognition* within forensic applications (e.g. Cattell, 1893; Loftus 1992; Rattner 1988) and as a direct consequence of research and formal enquiry, the Turnbull Guidelines require that jurors be cautioned to the risk of recognition error (Devlin, 1976; R v Turnbull, 1977), and the Police and Criminal Evidence Act (PACE, 1984) provides a detailed code of practice for the use of formal facial recognition procedures. Recall and reconstruction of facial information has also received considerable attention (e.g. Brace et al., 2000: Bruce et al., 2002: Davies et al., 2000: Frowd et al. 2004a,2004b; Geiselman, Fisher, MacKinnon, & Holland, 1987; Shepherd & Ellis, 1996; Wells,

1985) and guidelines for interviewing and imaging techniques are provided in the Facial Identification Guidance (2009) document produced by the National Policing Improvement Authority (NPIA) on behalf of the Association of Chief Police Officers (ACPO).

There have, however, been few evaluations of *face perception* within forensic application (e.g. Bruce et al, 1999, Henderson et al, 2000; Kemp et al, 1997; Strathie, 2010) with the consequence that in legal and operational settings there is little awareness of how unfamiliar faces are interpreted (Hancock & McIntyre, 2011) and no acknowledgement of difficulty in unfamiliar face perception within any formal recommendations (e.g. Attorney General's Reference No. 2 of 2002). The aim of this thesis was to provide a better understanding of how unfamiliar face perception influences identification of facial composites, and identity verification from facial image comparison, and to establish whether facial imaging techniques could be employed to develop more effective identification procedures.

It was demonstrated that unfamiliar face perception will influence the quality of facial composites but that an appreciation of these effects can produce identification protocols that facilitate familiar face recognition. It is recommended that facial composites are presented in a split and misaligned format to enhance recognition of accurate composite features. Future study should determine whether this can be achieved in conjunction with presentation of original composite images, and whether a split format during composite construction can enhance the quality of facial information. Affective information was shown to have a powerful influence on whether recognition judgements for facial composites will be achieved; it is recommended that the evolutionary facial composite systems should incorporate affective transformations so that informed witnesses may agree an acceptable image modification. With appropriate field trials and police service collaboration, such implementations could usefully be included within formal imaging guidelines.

Identity verification from face images is common practice but there is scant awareness in operational policing or the courts that unfamiliar facial comparison is problematic. The Home Office Scientific Development Branch recognises that appropriate training and guidelines are required, but in view of the available technology, it is also requisite that imaging techniques be developed to enhance face matching performance. Discrimination of unfamiliar faces is difficult but will be facilitated by distinctiveness; this can be enhanced with a computerised caricature manipulation. Modest levels of transformation can significantly reduce false identification with no loss to accurate identification. To confirm the reliability of this effect, future work should replicate these findings with different and more variable image sets. The use of caricature should also be evaluated within biometric face recognition applications, such that level of image correspondence may usefully determine an appropriate level of caricature transformation.

Cross race identification is particularly problematic and will produce more errors in unfamiliar facial image comparison. Perceptual inexperience with other race facial variation is a significant factor, which means that other race faces cannot efficiently be processed in a holistic manner. Cross race identity verification will be more effective if a procedure that facilitates featural comparison is employed, i.e. simultaneous comparison of a number of images. Where individual images must be inspected, transforming other race faces towards the facial variation of the operator will allow holistic face processing and discrimination to be more effective. For both own race and other race image comparison, a matching procedure that employed sequential and simultaneous presentation produced the most accurate results. It is recommended that the race transformation effects are replicated with different race faces, and that the influence of distinctiveness and presentation format should be explored.

The costs of incorrect facial identification are high and while the Home Office recognises a need for training and formal guidelines, such implementation must be based on knowledge of how unfamiliar face images are perceived. The current state of knowledge within policing and the courts is negligible and will contribute to facial identification and legal error. Formal enquiry and dissemination is needed to provide an informed frame of reference, and where appropriate to develop safer identification protocols. The research presented in this thesis provides evidence that unfamiliar face perception is prone to error but has demonstrated that an understanding of the effects can be used with imaging technology to develop more effective facial identification applications.

References

Adini, Y., Moses, Y., & Ullman, S. (1994). Face recognition: The problem of compensating for changes in illumination direction. In J. Eklundh (Ed.), *Lecture Notes in Computer Science: Vol. 800. Computer Vision—ECCV'94*, (pp. 286–296). Berlin, Germany: Springer-Verlag. doi: 10.1007/3-540-57956-7_33..

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and Nurture in Own-Race Face Processing. *Psychological Science, 17*(2), 159-163. doi:10.1111/j.1467-9280.2006.01679.x

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*(3), 281-316.

Barton, J. J., Cherkasova, M., & O'Connor, M. (2001). Covert recognition in acquired and developmental prosopagnosia. *Neurology*, *57*(7), 1161-1168.

Baudouin, J. Y., Gilibert, D., Sansone, S., & Tiberghien, G. (2000). When the smile is a cue to familiarity. *Memory*, *8*(5), 285–292.

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, *25*(5), 475–491.

Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, *3*(1), 105-135.

Benson, P. J., & Perrett, D. I. (1994). Visual processing of facial distinctiveness. *Perception*, *23*(1), 75-93.

Blackburn, R. (1996). What is forensic psychology? *Legal and Criminological Psychology*, *1*(1), 3–16.

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*(5), 527-536. doi:10.1080/13506280244000168

Bonner, L., Burton, A. M., Jenkins, R., McNeill, A., & Bruce, V. (2003). Meet The Simpsons: Top-down effects in face learning. *Perception*, *32*(10), 1159–1168.

Brace, N. A, Pike, G. E., & Turner, J. A. (2008b). Holistic facial composite systems: are they compatible with witness recall? *Cognitive Technology Journal*, *13*(2), 42–49.

Brace, N. A., Pike, G. E., Allen, P., & Kemp, R. I. (2006). Identifying composites of famous faces: Investigating memory, language and system issues. *Psychology, Crime and Law*, *12*(4), 351–366.

Brace, N., Pike, G., Kemp, R., Turner, J., & Bennett, P. (2006). Does the presentation of multiple facial composites improve suspect identification? *Applied Cognitive Psychology*, *20*(2), 213–226.

Brace, N. A., Hole, G. J., Kemp, R. I., Pike, G. E., Van Duuren, M., & Norgate, L. (2001). Developmental changes in the effect of inversion: Using a picture book to investigate face recognition. *Perception*, *30*(1), 85-94. doi:10.1068/p3059

Bredart, S., Brennen, T., Delchambre, M., McNeill, A., & Burton, A. M. (2005). Naming very familiar people: When retrieving names is faster than retrieving semantic biographical information. *British Journal of Psychology*, *96*(2), 205-214.

Brennan, S. (1985). Caricature Generator, the Dynamic Exaggeration of Faces by Computer + Illustrated Works. *Leonardo*, *18*(3), 170-178.

Brigham, J. C., & Ready, D. J. (1985). Own-race bias in lineup construction. *Law and Human Behavior*, *9*(4), 415-424. doi:10.1007/BF01044480

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

Bruce, V., Ness, H., Hancock, P. J. ., Newman, C., & Rarity, J. (2002). Four heads are better than one: combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, *87*(5), 894.

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

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

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

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243-248.

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

Caharel, S., Courtay, N., Bernard, C., Lalonde, R., & Rebai, M. (2005). Familiarity and emotional expression influence an early stage of face processing: an electrophysiological study. *Brain and Cognition*, *59*(1), 96–100.

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

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*(2), 527-551. doi:10.1037/0096-1523.26.2.527

Calderwood, L., & Burton, A. M. (2006). Children and adults recall the names of highly familiar faces faster than semantic information. *British Journal of Psychology*, *97*(4), 441–454.

Campbell-Tiech, A. (2005). “Stockwell” revisited: The unhappy state of facial mapping. *Archbold News*, *6*, 4–6.

Carey, S. (1981). The development of face perception. In G. Davies, H. Ellis, & J. Shephard (Eds.), Perceiving and remembering faces. London: Academic Press.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition*, *2*(3), 253-274. doi:10.1080/13506289408402302.

Carson, D. R., Burton, A. M., & Bruce, V. (2000). Putting names to faces: A review and tests of the models. *Pragmatics & Cognition*, *8*(1), 9-62.

Cattell, J. M. . (1893). On errors of observation. *The American Journal of Psychology*, *5*(3), 285–293.

Chance, J. E., & Goldstein, A. G. (1979). Reliability of face recognition performance. *Bulletin of the Psychonomic Society*, *14*(2), 115-117.

Chance, J. E., & Goldstein, A. G. (1981). Depth of processing in response to own- and other-race faces. *Personality and Social Psychology Bulletin*, *7*(3), 475-480.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *48A*(4), 879-894.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, *29*(8), 893-909. doi:10.1068/p2949

Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by Juries, Exonerated by Science*. Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, *111*(2), 219-247. doi:10.1016/j.cognition.2009.02.004

Davies, G., & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, *67*(1), 103.

Davies, G., & Griffiths, L. (2008). Eyewitness Identification and the English Courts: A Century of Trial and Error. *Psychiatry Psychology and Law*, *15*(3), 435-449.

Davies, G., & Milne, A. (1982). Recognizing faces in and out of context. *Current Psychological Reviews*, *2*(4), 235–246.

Davies, G., & Milne, A. (1985). Eyewitness Composite Production. *Criminal Justice and Behavior*, *12*(2), 209.

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. doi:10.1037/0021-9010.63.2.180

Davies, G., van der Willik, P., & Morrison, L. J. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology*, *85*(1), 119-124. doi:10.1037/0021-9010.85.1.119

Davis, J. P., & Valentine, T. (2009). CCTV on Trial: Matching Video Images with the Defendant in the Dock. *Applied Cognitive Psychology*, *23*(4), 482-505.

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied*, *14*(2), 139-150.

de Haan, E. H., Young, A. W., & Newcombe, F. (1992). Neuropsychological impairment of face recognition units. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *44A*(1), 141-175.

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology*, *96*(1), 57-70.

Devine, P. G., & Malpass, R. S. (1985). Orienting Strategies in Differential Face Recognition. *Personality and Social Psychology Bulletin*, *11*(1), 33 -40. doi:10.1177/0146167285111003

Devlin, P. D. (1976). *Report to the Secretary of State for the Home Department of the Departmental Committee on evidence of identification in criminal cases*. London: HMSO.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107-117. doi:10.1037//0096-3445.115.2.107

Dimberg, U., & Lundquist, L. O. (1990). Gender differences in facial reactions to facial expressions. *Biological Psychology*, *30*(2), 151–159.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition*, *25*(2), 129–139.

Dwyer, J., Neufeld, P. J., & Scheck, B. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday. doi:10.1111/j.1467-9280.2007.01879.x.

Edmond, G., Biber, K., Kemp, R., & Porter, G. (2009). Law's Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images. *Current Issues in Criminal Justice, 20*(3), 337–377.

Eich, E. (1995). Searching for mood dependent memory. *Psychological Science*, *6*(2), 67.

Eich, E., Macaulay, D., & Ryan, L. (1994). Mood dependent memory for events of the personal past. *Journal of Experimental Psychology: General*, *123*(2), 201.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Journal of Nonverbal Behavior*, *1*(1), 56–75.

Ellis, H. D. (1990). Developmental trends in face recognition. *The Psychologist*, *3*, 114-119.

Ellis, H. D., & Young, A. W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry, 157: 239-248*.

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*(4), 431-439.

Endo, N., Endo, M., Kirita, T., & Maruyama, K. (1992). The effects of expression on face recognition. *Tohoku Psychologica Folia, Vol. 51,* 37-44.

Fallshore, M., & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1608.

Farah, M. J. (1994). Specialization within visual object recognition: Clues from prosopagnosia and alexia. In M. J. Farah & G. Ratcliff (Eds). *The neuropsychology of high-level vision: Collected tutorial essays. Carnegie Mellon symposia on cognition.* (pp. 133-146). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 628-634.

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, *35*(14), 2089-2093.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*(3), 482-498.

Fischer, A. H. (1993). Sex differences in emotionality: Fact or stereotype? *Feminism & Psychology*, *3*(3), 303.

Fisher, R. P., Geiselman, R. E., Raymond, D. S., Jurkevich, L. M., & others. (1987). Enhancing enhanced eyewitness memory: Refining the cognitive interview. *Journal of Police Science & Administration, Vol. 15*(4), 291-297. .

Fletcher, K. I., Butavicius, M. A., & Lee, M. D. (2008). Attention to internal face features in unfamiliar face matching. *British Journal of Psychology*, *99*, 379-394.

Freire, A., & Lee, K. (2001). Face Recognition in 4- to 7- Year-Olds: Processing of Configural, Featural, and Paraphernalia Information. *Journal of Experimental Child Psychology*, *80*(4), 347–371.

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

Frowd, C. D, Bruce, V., Smith, A. J., & Hancock, P. J. . (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied, 14*(3), 276.

Frowd, C. D., Bruce, V., McIntyre, A. H., & Hancock, P. J. B. (2006). Adding holistic dimensions to a facial composite system. *Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition,* 183-188.

Frowd, C. D., Carson, D., Ness, H., Quiston-Surrett, D., Richardson, J., Baldwin, H., & Hancock, P. J. B. (2005). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, *10*, 63-81.

Frowd, C. D., Bruce, V., McIntyre, A. H., & Hancock, P. J. B. (2007). The relative importance of external and internal features of facial composites. *British Journal of Psychology*, *98*, 61-77.

Frowd, C. D., Bruce, V., Ross, D., McIntyre, A. H., & Hancock, P. J. B. (2007). An application of caricature: How to improve the recognition of facial composites. *Visual Cognition*, *15*(8), 954-984.

Frowd, C. D., Bruce, V., Smith, A. J., & Hancock, P. J. B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology-Applied*, *14*(3), 276-287.

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., & Hancock, P. J. B. (2005). A forensically valid comparison of facial composite systems. *Psychology Crime & Law*, *11*(1), 33-52.

Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception*, *1*(1), 19-39. doi:10.1145/1008722.1008725

Gallegos, D. R., & Tranel, D. (2005). Positive facial affect facilitates the identification of famous faces. *Brain and Language*, *93*(3), 338–348.

Ganel, T., & Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: the parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 583.

Garcia-Marques, T., Mackie, D. M., Claypool, H. M., & Garcia-Marques, L. (2004). Positivity can cue familiarity. *Personality and Social Psychology Bulletin*, *30*(5), 585.

Geiselman, R. E., & others. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police Science & Administration, Vol. 12*(1), 74-80..

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, *70*(2), 401.

Gibson, L. (2008). *Forensic art essentials: A manual for law enforcement artists*. Burlington: Academic Press.

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 995-1002.

Goldstein, A. G. (1979a). Facial feature variation: Anthropometric data II. *Bulletin of the Psychonomic Society*, *13*(3), 191-193.

Goldstein, A. G. (1979b). Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society*, *13*(3), 187-190.

Goldstein, A. G., & Chance, J. (1978). Judging face similarity in own and other races. *Journal of Psychology: Interdisciplinary and Applied*, *98*(2), 185-193.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Oxford: Robert E. Krieger.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception*, *13*(5), 505–512.

Hancock, P. J. B., & McIntyre, A. H. (n.d.). *Training methods for facial image comparison of unfamiliar faces: review and recommendations*.

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

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, *106*(2), 1017-1027.

Hehman, E., Mania, E. W., & Gaertner, S. L. (2010). Where the division lies: Common ingroup identity moderates the cross-race facial-recognition effect. *Journal of Experimental Social Psychology*, *Vol. 46*(2), 445-448.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*(4), 445-464. doi:10.1002/acp.718

Hess, U., & Bourgeois, P. (2010). You smile-I smile: Emotion expression in social interaction. *Biological Psychology*, *84*(3), 514–520.

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

Hills, P. J., & Lewis, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *The Quarterly Journal of Experimental Psychology*, *59*(6), 996-1002.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, *23*(1), 65-74. doi:10.1068/p230065

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, *31*(10), 1221-1240. doi:10.1068/p3252

Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia*, *34*(4), 247-261.

Hosie, J. A., Ellis, H. D., & Haig, N. D. (1988). The effect of feature displacement on the perception of well-known faces. *Perception*, *17*(4), 461-474. doi:10.1068/p170461

Hugenberg, K., & Corneille, O. (2009). Holisitic processing is tuned for in-group faces. *Cognitive Science: A Multidisciplinary Journal, Vol. 33*(6), 1173-1181.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334-340. doi:10.1016/j.jesp.2006.02.010

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, *117*(4), 1168-1187. doi:10.1037/a0020463

Itier, R. J., & Taylor, M. J. (2004). Face inversion and contrast-reversal effects across development: in contrast to the expertise theory. *Developmental Science, 7*(2), 246–260.

Jackiw, L. B., Arbuthnott, K. D., Pfeifer, J. E., Marcon, J. L., & Meissner, C. A. (2008). Examining the cross-race effect in lineup identification using Caucasian and First Nations samples. *Canadian Journal of Behavioural Science*, *40*(1), 52-57.

Johnston, R. A., & Bruce, V. (1990). Lost properties? Retrieval differences between name codes and semantic codes for familiar people. *Psychological Research*, *52*(1), 62-67.

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., Liu, Q., et al. (2007). Cross-Race Preferences for Same-Race Faces Extend Beyond the African Versus Caucasian Contrast in 3-Month-Old Infants. *Infancy*, *11*(1), 87. doi:10.1080/15250000709336871

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, *104*(1), 105-114. doi:10.1016/j.jecp.2009.01.006

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084-1089.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L., et al. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*(6), F31-F36. doi:10.1111/j.1467-7687.2005.0434a.x

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception*, *19*(4), 531–543.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, *11*(3), 211-222.

Kenealy, P. M. (1997). Mood state-dependent retrieval: The effects of induced mood on memory reconsidered. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(2), 290–317.

Kottoor, T. M. (1989). Recognition of faces by adults. *Psychological studies, Vol. 34*(2), 102-105.

Lander, K., & Metcalfe, S. (2007). The influence of positive and negative facial expressions on face familiarity. *Memory, Vol. 15*(1), 63-69.doi: 10.1080/09658210601108732..

Laughery, K. R., & Fowler, R. H. (1980). Sketch Artist and Identi-Kit Procedures for Recalling Faces. *Journal of Applied Psychology, 65*(3), 307-316.

Le Gal, P. M. (1999). Cognitive aspects of emotional expression processing. (Doctoral thesis). Stirling Online Research Repository. http://hdl.handle.net/1893/1772

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PloS one*, *4*(1), 4215.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *51*(3), 449-473.

Leder, H., & Bruce, V. (2000). When inverted faces are recognised: the role of configural information in face recognition. *Quarterly Journal of Experimental Psychology*, *53A*, 513-536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception*, *30*(1), 73-83.

Lee, K. J., & Perrett, D. I. (1997). Presentation-time measures of the effects of manipulations in colour space on discrimination of famous faces. *Perception*, *26*(6), 733-752.

Lee, K. J., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, *11*(5), 379-385.

Lee, W. J., Wilkinson, C., Memon, A., & Houston, K. (2009). Matching Unfamiliar Faces from Poor Quality Closed-Circuit Television (CCTV) Footage. *Axis: The Online Journal of CAHId*, *1*(1), 19–28.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General*, *129*(4), 559-574. doi:10.1037/0096-3445.129.4.559

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology*, *84*(3), 315-321. doi:10.1037/0021-9010.84.3.315

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556-564. doi:10.1037/0021-9010.70.3.556

Loftus, E. F. (1992). When a lie becomes memory’s truth: Memory distortion after exposure to misinformation. *Current Directions in Psychological Science*, *1*(4), 121–123.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces [Database of standardised facial images]. (Available from Psychology

Section. Department of Clinical Neuroscience, Karolinska Hospital, S-171 76 Stockholm, Sweden)

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law*, *7*(1), 98–118.

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law*, *7*(1), 134-152. doi:10.1037/1076-8971.7.1.134

Malone, D. R., Morris, H. H., Kay, M. C., & Levin, H. S. (1982). Prosopagnosia: A double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurology, Neurosurgery & Psychiatry*, *45*(9), 820-822. doi:10.1136/jnnp.45.9.820

Malpass, R. S. (1990). An Excursion Into Utilitarian Analysis. *Cross-Cultural Research*, *24*(1-4), 1 -15. doi:10.1177/106939719002400101

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

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

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics*, *69*(7), 1175-1184.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364-372.

Megreya, A. M., White, D., & Burton, A. M. (2011). The other race effect does not rely on memory: evidence from a matching task. *Quarterly Journal of Experimental Psychology, Vol. 64*(8), 1473-1483. doi: 10.1080/17470218.2011.575228.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3-35.

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for Own- and Other-race Faces: A Dual-process Approach. *Applied Cognitive Psychology*, *19*(5), 545-567.

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face encoding. *Cognitive Science*, *31*(5), 911–924.

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*(1), 55-73.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic Processing Is Finely Tuned for Faces of One's Own Race. *Psychological Science*, *17*(7), 608-615. doi:10.1111/j.1467-9280.2006.01752.x

Mondloch, C. J., Le Grand, R., & Maurer, D. (2010). Development of expertise in face recognition. in I. Gauthier, M. J. Tarr, & D. Bub (Eds.) *Perceptual expertise: Bridging brain and behavior* (pp.67-106). New York: Oxford University Press.

Mondloch, C., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*(5), 553-566. doi:10.1068/p3339

Mondloch, C.J., Dobson, K. S., Parsons, J., & Maurer, D. (2004). Why 8-year-olds cannot tell the difference between Steve Martin and Paul Newman: Factors contributing to the slow development of sensitivity to the spacing of facial features. *Journal of Experimental Child Psychology*, *89*(2), 159-181.

Mondloch, C. J., Pathman, T., Maurer, D., Le Grand, R., & de Schonen, S. (2007). The composite face effect in six-year children: Evidence of adult-like holistic face processing. *Visual Cognition*, *15*(5), 564-577.

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

Munsterberg, H. (1908). *On the Witness Stand.* New York: McClure.

Ng, W. J., & Lindsay, R. C. L. (1994). Cross-Race Facial Recognition. *Journal of Cross-Cultural Psychology*, *25*(2), 217 -232. doi:10.1177/0022022194252004

Nishimura, M., Maurer, D., & Gao, X. (2009). Exploring children’s face-space: A multidimensional scaling analysis of the mental representation of facial identity. *Journal of Experimental Child Psychology*, *103*(3), 355–375.

O’Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, *30*(6), 755-764.

Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition, 82*(3), 225-257.

Pauker, K., & Ambady, N. (2009). Multiracial faces: How categorization affects memory at the boundaries of race. *Journal of Social Issues*, *65*(1), 69-86.

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B. J., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology*, *96*(4), 795-810. doi:10.1037/a0013265

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening line-up most likely to affect eyewitness identification accuracy? *Legal and Criminological Psychology*, *10*, 247-263. doi:10.1348/135532505X49846

Pike, G., Brace, N., Turner, J., & Kynan, S. (2005). Making faces with computers: Witness cognition and technology. *Pragmatics & Cognition*, *13*(3), 459–479. doi:10.1075/pc.13.3.04pik.

Platz, S. J., & Hosch, H. M. (1988). Cross-Racial/Ethnic Eyewitness Identification: A Field Study. *Journal of Applied Social Psychology*, *18*(11), 972-984. doi:10.1111/j.1559-1816.1988.tb01187.x

Rahman, R. A., Sommer, W., & Olada, E. (2004). I recognize your face, but I can't remember your name: A question of expertise? *Quarterly Journal of Experimental Psychology Section A*, *57*(5), 819–834.

Rakover, S. S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology*, *93*, 1-30.

Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American*, *259*(2), 76–83.

Rattner, A. (1988). Convicted but innocent. *Law and Human Behavior*, *12*(3), 283–293.

Rattner, A., Weimann, G., & Fishman, G. (1990). Cross ethnic identification and misidentification by Israelis. *Sociology and Social Research*, *74*(2), 73-79.

Rhodes, G., & McLean, I. G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding. *Perception, 19*(6), 773 – 794. doi:10.1068/p190773

Rhodes, G., & Tremewan, T. (1994). Understanding face recognition: Caricature effects, inversion, and the homogeneity problem. In V. Bruce & G. Humphreys (Eds.) *Object and face recognition. Special issue of Visual Cognition, Vol. 1,* No. 2/3, (pp. 275-311). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1995-97446-005&site=ehost-live

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80*(3), 313-331.

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

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review*, *13*(3), 499-505.

Rhodes, G., Locke, V., Ewing, L., & Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception, 38*(2), 232-241.

Rhodes, G., Tan, S., Brake, S., & Taylor, K. (1989). Expertise and configural coding in face recognition: Erratum. *British Journal of Psychology, 80*(4), 526.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision.

Proceedings of the Royal Society B: Biological Sciences, *271* (Suppl 6), S448-S450.

Rodin, M. J. (1987). Who is memorable to whom? A study of cognitive disregard. *Social Cognition*, *5*(2), 144-165.

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

Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica, 132*(3), 300-312.

Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., & Guérit, J. M. (1999). Spatio-temporal localization of the face inversion effect: An event-related potentials study. *Biological Psychology*, *50*(3), 173–189.

Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision*, *8*(4), 1-13. doi:10.1167/8.4.3

Sangrigoli, S., & de Schonen, S. (2004). Effect of visual experience on face processing: a developmental study of inversion and non-native effects. *Developmental Science*, *7*(1), 74-87. doi:10.1111/j.1467-7687.2004.00324.x

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the Other-Race Effect in Face Recognition During Childhood. *Psychological Science*, *16*(6), 440-444. doi:10.1111/j.0956-7976.2005.01554.x

Scanlan, L. C., & Johnston, R. A. (1997). I Recognize Your Face, But I Can't Remember Your Name: A Grown up Explanation. *The Quarterly Journal of Experimental Psychology A, 50*(1), 183–198.

Schiltz, C., Dricot, L., Goebel, R., & Rossion, B. (2010). Holistic perception of individual faces in the right middle fusiform gyrus as evidenced by the composite face illusion. *Journal of Vision*, *10*(2), 1-16. doi:10.1167/10.2.25

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36-71. doi:10.1016/0010-0285(90)90003-M

Schwaninger, A, Lobmaier, J., Wallraven, C., & Collishaw, S. (2009). Two Routes to Face Perception: Evidence From Psychophysics and Computational Modeling. *Cognitive Science*, *33*(8), 1413-1440. doi:10.1111/j.1551-6709.2009.01059.x

Schwarzer, G., Zauner, N., & Jovanovic, B. (2007). Evidence of a shift from featural to configural face processing in infancy. *Developmental Science*, *10*(4), 452–463.

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*(6), 1748.

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception & Psychophysics*, *61*(6), 1102-1115.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology: CB*, *14*(5), 391-396. doi:10.1016/j.cub.2004.02.028

Sekunova, A., & Barton, J. J. S. (2008). The effects of face inversion on the perception of long-range and local spatial relations in eye and mouth configuration. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1129-1135. doi:10.1037/0096-1523.34.5.1129

Sergent, J., Ohta, S., Macdonald, B., & Zuck, E. (1994). Segregated processing of facial identity and emotion in the human brain: A PET study. *Visual Cognition, 1*(2-3), 349–369.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75*, 221-242.

Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology, 3*(1), 137-145.

Smith, S. M, & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*(2), 203–220.

Smith, S. M., Stinson, V., & Prosser, M. A. (2004). Do They All Look Alike? An Exploration of Decision- Making Strategies in Cross-Race Facial Identifications. *Canadian Journal of Behavioural Science, 36*(2), 146-154. doi:10.1037/h0087225

Smith, S. M., Lindsay, R. C. L., Pryke, S., & Dysart, J. E. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross-race situation? *Psychology, Public Policy, and Law, 7*(1), 153-169. doi:10.1037/1076-8971.7.1.153

Sporer, S. L., Trinkl, B., & Guberova, E. (2007). Matching faces: Differences in processing speed of out-group faces by different ethnic groups. *Journal of Cross-Cultural Psychology, 38*(4), 398-412. doi:10.1177/0022022107302310

Sporer, S. L. (1991). Deep-Deeper-Deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 323-333.

Stern, W. (1939). The psychology of testimony. *The Journal of Abnormal and Social Psychology*, *34*(1), 3.

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

Strathie, A. (2010). *Person Identification in a Legal Setting* (PhD). Glasgow Caledonian University.

Tanaka, J. W, & Sengco, J. A. (1997a). Features and their configuration in face recognition. *Memory & Cognition*, *25*(5), 583–592.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *46A*(2), 225-245.

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 122.

Tanaka, J. W., & Simon, V. B. (1996). Caricature recognition in a neural network. *Visual Cognition*, *3*(4), 305-324.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study. *Cognition*, *93*(1), B1-B9. doi:10.1016/j.cognition.2003.09.011

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*(2), 255.

Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*(5), 42-50.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press. CUP Archive.

Tranel, D., & Damasio, A. R. (1985). Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science*, *228*(4706), 1453-1454.

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*(5), 690-696.

Tulving, E. (1982). Synergistic ecphory in recall and recognition. *Canadian Journal of Psychology*, *36*(2), 130.

Tulving, E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences*, *7*(2), 223–68.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, *87*(1), 116.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124.

Ucros, C. G. (1989). Mood state-dependent memory: A meta-analysis. *Cognition and Emotion*, *3*(2), 139–169.

Valentine, T., Davis, J. P., Thorner, K., Solomon, C., & Gibson, S. (2010). Evolving and combining facial composites: Between-witness and within-witness morphs compared. *Journal of Experimental Psychology: Applied, 16*(1), 72.

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

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

Valentine, T., & Bruce, V. (1986a). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*(5), 525-535.

Valentine, T., & Bruce, V. (1986b). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40*(3), 300-305.

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

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefèvre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision*, *10*(5). doi:10.1167/10.5.10

Walker, P. M., & Hewstone, M. (2006). A Perceptual Discrimination Investigation of the Own-Race Effect and Intergroup Experience. *Applied Cognitive Psychology, 20*(4), 461-475.

Walker, P. M., & Hewstone, M. (2008). The influence of social factors and implicit racial bias on a generalized own-race effect. *Applied Cognitive Psychology*, *22*(4), 441-453. doi:10.1002/acp.1382

Weimann, G., Fishman, G., & Rattner, A. (1986). Social Distance and Misidentification. *International Journal of Comparative Sociology*, *27*, 217-225. doi:10.1163/156854286X00177

Wells, G. L, & Hasel, L. E. (2007). Facial composite production by eyewitnesses. *Current Directions in Psychological Science*, *16*(1), 6.

Wells, G. L. (1984). The Psychology of Lineup Identifications. *Journal of Applied Social Psychology, 14*(2), 89-103. doi:10.1111/j.1559-1816.1984.tb02223.x

Wilkinson, C., & Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science & Justice*, *49*(3), 191-196.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law*, *7*(1), 119-133. doi:10.1037/1076-8971.7.1.119

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology*, *17*(3), 365-373. doi:10.1002/acp.898

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141-145.

Young, A. W, Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14*(6), 737–746.

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

Young, A. W., Ellis, A. W., & Flude, B. M. (1988). Accessing stored information about familiar people. *Psychological Research*, *50*(2), 111-115.

Young, A. W., McWeeny, K. H., Ellis, A. W., & Hay, D. C. (1986). Naming and categorizing faces and written names. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *38A*(2), 297-318.

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Access to identity-specific semantic codes from familiar faces. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *38A*(2), 271-295.

Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology*, *76*(4), 495-523.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747-759.

Yovel, G., & Kanwisher, N. (2004). Face Perception:: Domain Specific, Not Process Specific. *Neuron*, *44*(5), 889–898.