# Learning Faces From Photographs

Christopher A. Longmore

Thesis submitted for the degree of Ph.D.

University of York

Department of Psychology

May 2007

# Abstract

This thesis examines how faces become familiar through the learning of photographs of faces. Chapter One reviews the literature on how unfamiliar and familiar faces are processed and recognised. Is face recognition achieved via a viewpoint-invariant structural model or by image-based codes, and are particular features of the face more useful for recognition than others? The roles of expertise, motion and the modelling of face recognition are discussed. Chapter Two presents five experiments which examine whether face recognition is achieved via a viewpoint-invariant structural model or by image-based codes such as viewpoint-dependent codes or pictorial-codes. It was found that recognition of faces learnt from photographs is likely to be mediated by image-based codes. Chapter Three investigates which of two types of image-based coding; viewpoint-dependent codes or pictorial codes, most adequately explain recognition performance after transformations of lighting and size. The results suggest that the recognition of faces learnt from photographs is highly image specific and is mediated by pictorial codes. Chapter Four examined whether the recognition process requires the pictorial codes, formed during learning and extracted from the test image, need to be identical for recognition. Colour changes between study and test reveal that a greyscale image intensity map is sufficient for recognition. Chapter Five studied whether particular features of the face are useful for recognition. It was found that faces learnt from a single photograph are recognised better form their internal features than from the external features. In addition, it appears that hair is a potentially misleading cue and removing it leads to improved recognition rates. Finally, Chapter Six summarises the main findings, discusses implications for theories of face recognition and suggests some future research directions.

# Contents

# List of Tables

# List of Figures

**Acknowledgements**

I would like to thank a number of people who have been extremely supportive during the period in which the work presented in this thesis was undertaken. In particular, special thanks to Isabel, Yang-Ming and Michael for providing stimulating critiques of the work and for providing new ideas and directions. Thanks also go to the large number of friends made in York (in no particular order); Michelle, Cindy, Naomi, Paul, Padraig, Uzma and Laura. Alex – I'll never forget Toffs!

Thank you to my family, for their constant support through the good times and the bad. Without them completing my studies at both degree and masters level would not have been possible. It is through them that I was able to achieve my goal of working towards a PhD.

I would also like to thank the members of my research committee; Alan Baddeley and Peter Thompson for their valuable comments and input into this work. Finally, a big word of thanks goes to my supervisors; Andy Young and Chang Hong Liu for their guidance and support throughout the preparation of this thesis! Without them, it simply just would not have been possible.

# Declaration

This thesis contains original work completed by myself under the supervision of Professor Andy Young and Dr Chang Hong Liu

The data presented in Chapter Two have been accepted for publication in the paper:

> Longmore, C.A., Liu, C.H. & Young, A.W. (in press) Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*.

and were presented at the British Psychological Society Cognitive Section Annual Conference in September 2005 as:

> Longmore, C.A., Liu, C.H. and Young, A.W. (2005) Learning faces from photographs. Paper presented at the BPS Cognitive Section Annual Conference, 31st August – 3rd September 2005

and in a poster presented to the Experimental Psychology Society – London Meeting in January 2006 as:

> Longmore, C.A., Liu, C.H. & Young, A.W. (2006) Learning new faces – recognition after pose and lighting transformations. Poster presented at the EPS London Meeting, 5th – 6th January 2006.

The work presented in this thesis was supported by research studentship grant G78/8256 from the Medical Research Council.

# CHAPTER ONE

Literature review

## 1.1 Introduction

For the majority of us, recognising the faces of people we know is an effortless task. We are able to recognise friends, family members and famous celebrities even if their appearance has altered since the last time we saw them (e.g. due to changes such as hairstyle and aging). How we go about recognising people has been the focus of much research. This is not surprising; face recognition is important for our day-to-day social interactions. Without it, our ability to communicate successfully with those around us would be impaired as we would not be able to discriminate family members, colleagues, and others known to us from the individuals that we encounter on a daily basis. As well as the interest in face recognition from a social point of view, experimental cognitive psychology has been interested in the topic as it represents what could be considered the ultimate within-category discrimination. After all, recognition of familiar faces is something that seems effortless and accurate despite the fact that all faces are constituted of the same parts (eyes, nose and mouth), located in approximately the same positions.

At the same time, there is a strong need to understand our ability to recognise less well known faces. For example, the eyewitness has long been seen as one of the key components of the prosecution case during a criminal trial. If our performance with unfamiliar faces is as good as it is with familiar faces (i.e. we can recognise unfamiliar faces as well as we recognise familiar faces) then if we could satisfy ourselves that the person providing the identification is reliable and has no ulterior motive for making a false accusation, then the evidence would be pretty damning for the defendant. But what if unfamiliar face recognition isn't as good as familiar face recognition? Would we then be happy convicting someone on this evidence?

Early studies from the late 1960s and early 1970s indicated that peoples' recognition memory for pictures (Shepard, 1967) and photographs of unfamiliar faces (Galper & Hochberg, 1971; Hochberg & Galper, 1967) was very good. These early studies employed the same picture of the face at study and test. The problem with this type of experimental design is that it does not distinguish between picture recognition and face recognition. Whilst picture recognition only requires the recognition of a previously seen picture, true face recognition would imply that the face can be recognised from a different image, for example after a transformation of lighting or pose.

Tests which do examine face recognition paint a very different picture. As will be discussed throughout this review, it appears that unfamiliar face recognition is poor in comparison to familiar face recognition when different images are used for learning and recognition. This chapter begins by reviewing previous work on the mental representation of faces. It then moves on to examine how we manage to finely discriminate between faces, the role of motion in learning new faces, whether faces can be recognised from any part of the face or if there are particular features of the face that are useful for recognition. Finally, the cognitive and computational modelling of faces is considered.

## 1.2 How are faces recognised?

We are able to identify friends and family after large changes in appearance that can result from a change in pose, expression or a change in lighting, as well as more subtle changes, for example after they have acquired a suntan. How do we recognise people we know despite these changes? There are two main possibilities. Firstly, we might construct a three-dimensional, viewpoint invariant structural model of a face. Alternatively, an image-based description, which may be viewpoint dependent (i.e. recognition is dependent upon seeing a face in a previously seen *view*) or pictorially-dependent (i.e. recognition is dependent upon seeing a face from a previously seen *image*) may be formed.

The distinction between viewpoint-invariant and viewpoint-dependent recognition has been examined extensively in the object recognition literature. Whether objects are recognised from viewpoint-invariant structural information (e.g. Marr, 1982; Marr & Nishihara, 1978) or from multiple instances of different views (e.g. Bulthoff & Edelman, 1992; Bulthoff, Edelman, & Tarr, 1995; Poggio & Edelman, 1990; Tarr & Pinker, 1989) has led to much debate (Biederman & Gerhardstein, 1993; Biederman & Bar, 1999; Hayward & Tarr, 1997; Hayward & Tarr, 2000; Hayward, 2003) without as yet clear consensus on the issue. Viewpoint-invariant and viewpoint-dependent models of recognition are discussed in the following sections in the context of face recognition research.

### 1.2.1 Recognition via edge-based descriptors and viewpoint-invariant structural models

One possible type of mental representation formed for faces is a three-dimensional viewpoint-invariant structural model. These models (e.g. Biederman, 1987; Marr, 1982) propose that the internal representation of an object consists of a three-dimensional structure which can be rotated to any view, allowing for the recognition of that object in a novel viewpoint. At an early stage in the construction of the model, the edges and contours of the object are identified. As such, structural models also encompass edge-based models which highlight the importance of the extraction of information about the edges and contours of an object. For example, in Marr's influential model of vision (Marr, 1982), an edge-based description will be created during the full-primal sketch stage which takes place before the final stage of object recognition, which is assumed to be viewpoint-invariant.

Such a theory sounds promising for the recognition of familiar faces as they can indeed be recognised in a variety of poses (Bruce, 1982) (both achieved via rotating the structural model), after changes in hue (Kemp, Pike, White, & Musselman, 1996) and lighting direction (both achieved via an edge-based description of the face). However, evidence exists that suggests that recognition via a structural model or edge-based description may not provide an adequate account of familiar face recognition.

#### 1.2.1.1 Recognition via edge-based descriptions

Edge-based models would suggest that the recognition of a face, familiar or unfamiliar, is possible under different lighting conditions as the edge description will not be affected by a change in lighting, despite the fact that a lighting change produces large changes in the facial image (indeed, Adini, Moses and Ullman (1997) showed that changing the lighting direction of a face between two images of the same person produces larger image differences than changing the identity). Moses, Ullman and Edelman (1996) presented participants with a series of unfamiliar faces together with a name which they were to learn. They found that participants were able to accurately name newly learned faces after a change in lighting direction, lending support to an edge-based account of face recognition.

There is contrasting evidence however, that indicates that when the lighting direction of an unfamiliar face is changed between study and test, recognition accuracy falls. Braje, Kersten, Tarr and Troje (1998) demonstrated that participants were poorer at matching two images of an unfamiliar face when the lighting direction changed between study and test than when the lighting direction remained consistent. Hill and Bruce (1996) report similar effects.

Lighting changes have also been demonstrated to significantly disrupt the recognition of familiar faces (Johnston, Hill, & Carman, 1992). Johnston et al. presented people with images of their friends' faces in which the lighting was from an unusual direction (from below) and found that recognition was significantly poorer in this lighting condition than when the faces were lit from a natural direction (above).

Another change to the facial image that, according to an edge-based account of face recognition should not have an impact on face recognition is reversing the contrast polarity of an image to produce a photographic negative image. This is because despite the change in pigmentation, the edge-based descriptors will remain unchanged. However, it seems that recognising familiar faces in photographic negative is hard (Galper & Hochberg, 1971; Hayes, Morrone, & Burr, 1986; Phillips, 1972). Bruce and Langton (1994) found that whilst familiar face are recognised well from normal photographs (with approximately 95% accuracy), the recognition of faces presented in photographic negative is significantly poorer (approximately 55% accuracy).

Figure 1-1 provides an illustration of the difficulty edge-based models have in explaining the poor recognition rates observed for faces presented in photographic negative. If face recognition functioned on the extraction of edges from a facial image then recognising a familiar face from edge maps should not be difficult. However, identifying the bottom-left image of Figure 1-1 as Tony Blair is very hard. Furthermore, recognition of the photographic negative image (the top right image of Figure 1-1) should be no more difficult than the photographic positive as edge-based descriptions of both photographic positive and negative images yields very similar results (the bottom row of Figure 1-1).

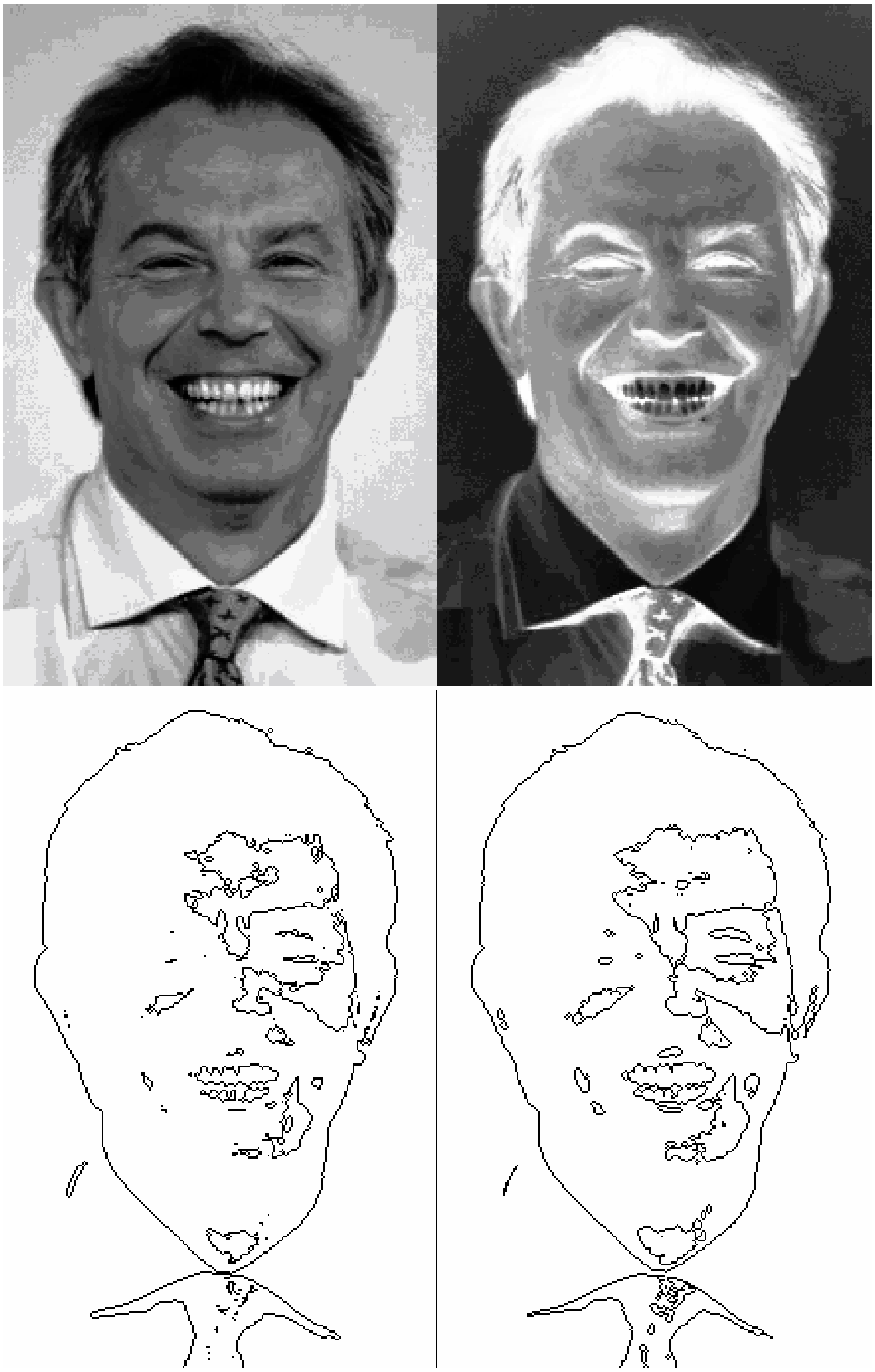

**Figure 1-1: Photographic positive (left) and negative (right) images presented with shading (top row) and as edge maps (bottom row) of British Prime Minister Tony Blair. Despite the high familiarity many people have with him, he is hard to recognise in photographic negative.**

#### 1.2.1.2 Recognition via a viewpoint-invariant structural model

Marr's (1982) and Biederman's (1987) models of object recognition claim that recognition of objects is viewpoint-invariant. That is, it does not matter from which viewpoint the object is seen in for successful recognition. If a viewpoint-invariant structural model could be constructed for faces then this would enable the recognition of a face after a variety of transformations such as pose and expression. These changes do not appear to affect the recognition of familiar faces but do affect the recognition of unfamiliar faces (Bruce, 1982). Indeed, in their influential model of face recognition, Bruce and Young (1986) propose that such a structural model is used for the recognition of familiar faces (see section 1.7.1) and there is some evidence of viewpoint-invariant recognition of unfamiliar faces (Davies, Ellis, & Shepherd, 1978; Patterson & Baddeley, 1977).

A structural model can explain why faces are hard to identify in photographic negative. In order for a face to be recognised via a structural model, the three-dimensional structure of the face must be extracted from the two-dimensional image. For faces presented in photograph positive this is not a problem as the face shape can readily be extracted from shape-from-shading cues (Bruce & Young, 1998). However, reversing the contrast polarity disrupts the shape-from-shading cues resulting in difficulties in recovering the three-dimensional structure of the face. This in turn makes matching the extracted information to the stored representation difficult and leads to poor performance on photographic negative images.

In order to test whether a three-dimensional structure of a face is sufficient for recognition, it is necessary to examine how well three-dimensional structural models of faces without information about skin texture (which could be used to aid the recognition process) can be recognised. Bruce et al. (1991) demonstrated that the recognition of familiar faces from their three-dimensional shape alone is difficult, suggesting that a structural model is unlikely to be used on its own and additional information (such as information about skin texture) is also employed for familiar face recognition.

Krouse (1981) tested recognition memory for faces presented in full-face view or three-quarter view and subsequently tested on either the same image or the same face after a pose transformation. She found that performance was significantly poorer when different images were presented at study and test than when the same images were given in both phases. Later studies have supported this finding, indicating that the recognition of unfamiliar faces is dependent on the viewpoint remaining unchanged between study and test (e.g. Baddeley & Woodhead, 1983; Hill & Bruce, 1996; O'Toole, Edelman, & Bulthoff, 1998) and not upon the formation of a three-dimensional model.

Bruce (1982) conducted a series of experiments using both the accuracy and response latencies of the recognition of familiar and unfamiliar faces. She found that unfamiliar faces presented in the same view at study and test were recognised significantly better than faces that were subject to a single transformation of pose or expression, which in turn were recognised significantly better than a face subjected to both pose and expression transformations. A similar pattern of results was found for the latency data, in which the originally studied image was faster to recognise at test than an image in which a transformation had been applied. Bruce claimed that is not the *type* of change that was important for recognition; it was the *amount* of change. Bruce (1982) also examined the effect of pose and expression transformations on the recognition of familiar faces. She found that changing the pose and expression of familiar faces had no effect on recognition accuracy compared to testing recognition of the same image, suggesting that a pre-existing representation of familiar faces mediates recognition. Participants clearly remembered something about the originally studied picture however as they could successfully identify that the image of a familiar face had changed between the study and test phases.

The drop in performance reported by Bruce for unfamiliar faces was considered to be due to the amount of change between the learned and test images. Hill, Schyns and Akamatsu (1997) investigated how well a facial structure (i.e. a face without any texture information) initially seen in one view could be recognised after different angles of rotation. They presented participants with a single image of a facial structure from one of five views. It was found that recognition memory performance was dependent upon the view seen during the initial presentation and that recognition accuracy was a function of the rotation away from the learnt view, with accuracy decreasing as the angular change between the study and test images increased. Hill et al. also investigated the effects of providing texture information about the face. They reported that the overall pattern of results was the same as when only the face shape was presented, although they did report viewpoint invariance for faces learned in the three-quarter view (see section 1.2.2.4, page 18 for a more detailed discussion of the three-quarter view advantage). O'Toole, Edelman and Bülthoff (1998) replicated these findings. Both studies produced results consistent with previous research on object recognition (e.g. Hayward & Williams, 2000) in which the recognition of objects, which differed in viewpoint between study and test was found to be a function of rotation between learnt and test views.

Overall, familiar faces appear to demonstrate some features consistent with a viewpoint-invariant structural model of recognition although it seems that such a model will not be used in isolation without the use of image-based cues (such as skin texture information). Unfamiliar faces on the other hand appear to be processed differently. The generalisation of an unfamiliar face to a novel view is difficult and error prone. It seems that unfamiliar face recognition is dependent not only upon the view that was originally seen, but also, in cases whether a face is learnt from a photograph, the particular picture that was studied.

## 1.2.2 Image-based models of recognition

Image-based models of recognition are relative newcomers to the scene of object recognition (Tarr, Kersten, & Bulthoff, 1998). These models claim that instead of creating and storing a single viewpoint-invariant structural model, recognition is tied much more closely to a previously seen image. Image-based models can be broken down into two distinct types; models of viewpoint-dependent recognition and recognition via pictorial codes.

### 1.2.2.1 Recognition via a viewpoint-dependent model

Viewpoint-dependent models (e.g. Bulthoff & Edelman, 1992; Gauthier & Tarr, 1997; Poggio & Edelman, 1990) propose that a number of discrete viewpoints of an object are stored for future use by the recognition system. Recognition is possible providing the object is presented in a view that has been, or is very close to, a view seen before.

Unfamiliar face recognition, on the surface, does appear to be viewpoint-dependent in nature as demonstrated by pose change studies. Familiar face recognition on the other hand, might be influenced by a viewpoint-invariant structural model of the face. Despite the implication of structural codes in the recognition of familiar faces, there is evidence of viewpoint-dependence in the recognition of familiar faces. Troje and Kersten (1999) reported viewpoint dependence in the recognition of familiar faces. In an intriguing study they presented participants with a series of faces in both full-face and profile views. The set of faces consisted of individuals known to the participant and also views of their own face.

Troje and Kersten hypothesised that if face recognition was viewpoint-dependent then, as people have a limited amount of experience with their own face in profile compared to their experience with the profile view of other familiar faces, performance will be poorer when viewing their own face in profile than either their own face in a full view or either view of the face of another, familiar person. This was indeed found, with participants being significantly slower at identifying their own face in profile compared to other familiar faces in profile. Troje and Kersten argue that this difference is due to the limited experience that people have with their own face when viewed in profile and suggest that familiar face recognition is viewpoint dependent.

It is important to note that whilst viewpoint-invariant and viewpoint-dependent theories of recognition have their roots in work conducted on object recognition, unfamiliar face recognition and object recognition may not employ the same processes. It is interesting to note that the viewpoint-dependency that characterises unfamiliar face recognition (Baddeley & Woodhead, 1983; Bruce, 1982; Jeffery, Rhodes, & Busey, 2006; Krouse, 1981; Hill et al., 1997; O'Toole et al., 1998; Troje & Bulthoff, 1996) has also been found for the recognition of familiar (Jolicoeur, 1985; Lawson & Humphreys, 1996) and novel objects (Bulthoff & Edelman, 1992; Tarr & Pinker, 1989; Tarr, Williams, Hayward, & Gauthier, 1998). Whilst Tarr and Cheung (2003) question whether object recognition is ever viewpoint-invariant, objects have also been shown to demonstrate viewpoint-invariance in some studies (Biederman & Bar, 1999). However this time, the same cannot be said for unfamiliar faces (which typically show a dependence on viewpoint).

#### 1.2.2.2 Recognition via pictorial codes

The preceding section reviews work which suggests that the recognition of unfamiliar faces is viewpoint-dependent. However, there is a growing body of evidence which suggests that unfamiliar face recognition in laboratory settings may be more specific to the image studied than simply the viewpoint studied.

The pose change, which has been the favoured transformation of those investigating unfamiliar face recognition, changes both the viewpoint of the face and the image of the face itself, and the drop in performance seen after a pose change may be due to either of these two changes. Hence, pose change studies, whilst providing valuable information about the nature of unfamiliar face recognition, cannot adequately separate viewpoint-dependent and pictorial coding. Instead, other transformations, such as lighting, size, and employing different images of the same person at study and test can help tease apart viewpoint and pictorial dependency as these transformations leave the viewpoint of a face unchanged but alter the image.

Studies which have investigated the recognition of unfamiliar faces after a change in lighting (Braje, Kersten, Tarr, & Troje, 1996; Braje et al., 1998; Braje, 2003; Hill & Bruce, 1996; Liu, Collin, Burton, & Chaudhuri, 1999) suggest that recognising or matching unfamiliar faces requires that the image of the face seen during learning and the image of the face seen during the test phase need to be very nearly identical. Indeed, in their study into unfamiliar face recognition after a lighting transformation, Braje et al. (1998) found that both matching and recognition performance of computer rendered faces, which can be rendered into the same viewpoint in both the probe and target images, is significantly decreased if the lighting direction is changed between the two images, suggesting that the recognition of unfamiliar faces is pictorially dependent.

Another transformation which can separate viewpoint-dependent coding and pictorial coding is that of size. As with the transformation of lighting, a transformation of size of a face between study and test maintains the same viewpoint of the face, but alters the image. Kolers, Duchnicky and Sundstroem (1985) reported that changing the size of an unfamiliar face between the learning and recognition phases of an experiment significantly decreased recognition accuracy, again suggesting that the recognition of unfamiliar faces is highly image specific.

Some of the most striking results indicating the image specificity of unfamiliar face recognition have come from studies on how well people can be recognised from closed-circuit television (CCTV). Typically, these experiments present participants with either video footage or a static frame from video footage during the study phase and then high quality photographs at test (e.g. Burton, Wilson, Cowan, & Bruce, 1999; Bruce, Henderson, Newman, & Burton, 2001; Henderson, Bruce, & Burton, 2001; Davies & Thasen, 2000; Roark, O'Toole, & Abdi, 2003). These studies indicate that when the face is familiar to the participant, recognition rates are good. When the observer is unfamiliar with the individual depicted in the video however, recognition accuracy is poor. This is despite the fact that the task of the participants is to match two pictures – there is no memory component in this task at all.

Bruce et al. (1999) demonstrated quite elegantly the image-specific nature of unfamiliar face recognition. They presented participants with a static image of an unfamiliar face taken from high-quality video. The participant was then asked to identify if this person was present in an array high of quality photographs (in trials in which the target was in the array of faces the two images would different) presented below the target face. Even with the viewpoint of the faces kept constant and all the faces displaying a neutral expression, people still performed quite poorly.

Familiar face recognition may also be reliant on a previously seen image of a face. For example, Davies, Ellis and Shepherd (1978) presented participants with a series of line drawings and photographs of famous faces. If familiar faces can be recognised from images quite different from the instances in which they have been seen before, the line drawings and photographs of famous faces should be recognised as equally well. Instead, they found that line drawings were recognised significantly poorer than photographs. Furthermore, Bruce, Hanna, Dench, Healy and Burton (1992) also demonstrated that line drawings of face were poorly recognised.

In conclusion, it appears that unfamiliar face recognition, as studied in the laboratory, is highly image-specific and recognition is often achieved by extracting pictorial codes from an image and comparing them against stored pictorial representations. In order to recognise faces from pictorial codes, the image presented for recognition must be sufficiently similar to the stored representation of the face. In addition, there is some suggestion that familiar face recognition might also be mediated by image-based cues as familiar faces are hard to recognise without skin texture information. Furthermore, familiar face recognition may be reliant upon seeing a face from a particular viewpoint which has been previously viewed. Familiar faces tend not to be learnt from photographs and therefore it is unlikely that pictorial codes are responsible for familiar face recognition. Instead, familiar face recognition might be instance-based in which snapshots, created during particular encounters with faces, are used for recognition. Therefore, it may be that the pictorial nature of unfamiliar face recognition mimics a real world face recognition mechanism which functions on particular instances of faces.

If face recognition is viewpoint or image dependent, generalisation from one view to another will require a number of different views to be stored to enable successful recognition across a wide range of viewpoints. However, not all views have been considered equally useful for face recognition. In fact, one particular view, the three-quarter view, has been considered especially useful for recognition. The advantage for the three-quarter view and reasons that may lie behind the advantage are discussed in sections 1.2.2.3 and 1.2.2.4 respectively.

#### 1.2.2.3 The three-quarter view advantage

Most studies on unfamiliar face recognition typically employ the use of the full face, the three-quarter or the profile view and a number of studies have found that of these views, the three-quarter view leads to better performance when used as the learnt view or test view. Woodhead, Baddeley and Simmons (1979) reported that when participants studied a face in full-face view, a three-quarter view at test was better recognised than a profile view. Baddeley and Woodhead (1983; also cited in Logie, Baddeley, & Woodhead, 1987) also report that when a single view is used for learning, the three-quarter view produces greater recognition performance than either the full-face or profile views.

Siéroff (2001) provided more specific evidence of a three-quarter view advantage. He presented participants with nine views of a face ranging from the left profile to right profile and found that during a recognition test, the left three-quarter view produced faster reaction times than any of the other views (including the right three-quarter view). A similar result was also reported by Laeng and Rouw (2001). They presented participants with full-face views and subsequently tested recognition accuracy for four views ranging from full-face to profile, and found that reaction times to faces presented in the three-quarter view were significantly shorter than for the other views.

All studies discussed so far on the three-quarter view advantage have used photographs as stimuli. Troje and Bülthoff (1996) employed the use of laser scanned three-dimensional computer generated models of human heads as their stimuli. They found that for both textured and untextured faces, the three-quarter view led to greater performance than either the full-face or profile views when it was used as a learning view. Conversely, Valentin, Abdi and Edelman (1997) also found a three-quarter view advantage when the three-quarter view was used at test.

Whilst most studies have employed the use of a recognition memory paradigm, the three-quarter view advantage has also been demonstrated in matching studies. Marotta, McKeeff and Behrmann (2002) presented participants with three computer-rendered faces comprising a probe face that they had to identify from two target faces. Both target faces underwent the same pose change from the probe image. They found that participants made the fewest errors when the target faces were presented in the three-quarter view compared to other views.

#### 1.2.2.4 Reasons for the three-quarter view advantage

Why should the three-quarter view be so useful for face recognition? Neurophysiological evidence is sparse but intriguing. Perrett et al. (1985) used single-cell recordings in the temporal cortex of the macaque monkey. They found that of all the cells that were responsive to faces, nearly two-thirds were sensitive to a particular orientation. Most of these cells responded strongly to either the full-face or profile view, with activation falling off the further the face was rotated away from the preferred face orientation of the cell. When a face was presented in a three-quarter view, both groups of cells were active, resulting in a larger overall level activity than either the full-face or profile views alone, increasing the chances of successful recognition.

Another argument for the three-quarter view advantage comes from studies of object recognition. Studies have revealed that some views of objects are easier to recognise than others. Palmer, Rosch and Chase (1981) demonstrated that certain views of objects are considered by participants to be "better" than other views (i.e. a more representative view of an object) and that these views lead to faster naming of objects. It has been suggested that the three-quarter view is the "best" view of a face as it provides the clearest information about the features in a face as firstly, most of the features are visible in the three-quarter view and secondly, some degree of information about the structure of the face is available.

Another reason for the emergence of a three-quarter view advantage is revealed in the studies of Hill et al. (1997) and O'Toole et al. (1998). In both of these studies recognition performance was greatest for the view that was learned and accuracy dropped off as the face was rotated away from the learned view. If recognition accuracy is a function of the angle of rotation away from a learned view, then learning or testing with a three-quarter view might be expected to lead to an *overall* higher level of performance. This is because the three-quarter view is closer to the full-face view and the profile views (the two other most commonly used views) than the full-face view and profile view are to each other.

However, not all studies have revealed an advantage for faces learned or tested in three-quarter view, whilst others find only limited support for the hypothesis. Bruce, Valentine and Baddeley (1987) reported that the three-quarter view advantage maybe limited to a particular task. In their first experiment, Bruce et al. found that when participants were required to recognise familiar faces, no three-quarter advantage was found. However, a second experiment using a matching paradigm revealed that pairs of unfamiliar faces presented in the three-quarter view were matched significantly faster (although not more accurately) than unfamiliar face pairs presented in full-face or profile views, suggesting that any three-quarter advantage is limited to unfamiliar faces.

Liu and Chaudhuri (2002) cast further doubt on the three-quarter view advantage using both recognition memory and matching paradigms. In their review of the three-quarter advantage they report some previously unpublished results of their earlier studies into the effects of photographic negation on face recognition (Liu & Chaudhuri, 1998; Liu et al., 1999). In a recognition memory task, Liu and Chaudhuri (1998) found that performance was very similar when the pose of a face was changed between study and test from full-face to three-quarter view or vice-versa. Similarly, using a sequential matching task, Liu et al. (1999) found that presenting either a full-face view followed by a three-quarter view or vice-versa led to equivalent levels of performance.

Considering the variability of the results obtained from studies looking at the three-quarter view advantage, it cannot be said that the three-quarter view holds any special properties that allow it to be recognised better than any other view. Bruce and Young (1998) suggest that the primary reason that some studies have found a benefit for learning or recognising the three-quarter view is simply because the three-quarter view allows for greater levels of generalisation than other views (such as the full-face and profile views).

#### 1.2.2.5 The number of views needed for invariant face recognition

If faces are recognised from a number of discrete views, then how many views are needed for successful recognition? Harries, Perrett and Lavender (1991) examined which views of a human head participants would choose to look at when trying to learn new faces. They presented participants with clay model heads on a turntable which the participant could use to turn the head around to any view they chose and there was no limit on the proportion of time that they could spend on one view. They found that during the learning phase, participants tended to focus upon two particular views; the full-face view and a view just short of full profile. This result suggests two things; first, the particular view(s) seen during learning is considered by the participants to be of importance for the effectiveness of the learning. Secondly, that the number of views may be important. Most experiments have presented the participant with a single view during the learning phase whereas the results of Harries et al. suggest that at least two views are preferred for proper learning of a face.

## 1.3 The number of stored images within a viewpoint for the recognition of faces

A key question is whether every instance of a known face is stored (exemplar-based approach), or whether various instances of an individual's face become merged so that only one representation needs to be formed (prototype-based approach) and stored.

### 1.3.1 Prototype and exemplar models of familiar face recognition

The exemplar-based approach (Knapp, Nosofsky, & Busey, 2006; e.g. Nosofsky, 1988; Nosofsky, 1991) proposes that the individual exemplars themselves are stored and recognition is achieved by the matching of the presented item with the stored exemplars. Employing such a system for face recognition would require storage of each known view of an individual's face. Prototype accounts of face recognition on the other hand emphasise the creation of a prototype or average face. Experimental evidence for the prototype effect in face recognition has mainly relied on the use of artificial stimuli such as identikit faces (Bruce et al., 1991; Inn, Walden, & Solso, 1993; Solso & Mccarthy, 1981; as cited in Cabeza, Bruce, Kato, & Oda, 1999). Bruce, Doyle, Dench and Burton (1991) asked participants to rate a series of computer-created faces for age and sex. For each face, two variations were created in which the face was extended (i.e. made "longer") relative to the average and another two in which the face was shortened relative to an average face. Crucially, this average face (the prototype) was never seen during the rating phase. During the test phase, participants were shown the previously unseen prototype and another variation of the face which was novel. Participants consistently selected the prototype as the face that had previously been seen. It seems that participants had taken the four variations which they saw during the rating part of the experiment and combined them to produce an overall average face which was then stored as the representation for that face.

Cabeza, Bruce, Kato and Oda (1999) successfully replicated the prototype effect using high quality photographs, suggesting that the creation of the prototype was not an artefact of the artificial stimuli used. The prototype effect also emerged when faces were presented in the three-quarter view, indicating that the effect is not limited to the full-face view. Cabeza et al. also addressed the issue of what limits, if any, are imposed upon prototype formation. They compared effects of angular variation with feature placement (i.e. the changing of the facial features within one view). They found that the prototype effect emerged even when there were large variations in the locations of the features of the face. When angular variation was examined, however, the prototype effect only emerged when small angles of rotation were introduced.

The process of the formation of a prototype formed from an average face is appealing as averaging a number of different images will remove most of the variations in lighting, yielding a representation of the face that is largely devoid of illumination information that would aid its recognition is new lighting conditions (Burton, Jenkins, Hancock, & White, 2005).

It would appear that whilst a prototype can be formed easily *within* views, it is not formed readily *between* views, suggesting viewpoint-dependent or pictorially-dependent face representations. Considering the results of Cabeza *et al*. (1999), it is quite possible that a number of prototypes are stored to enable recognition across viewpoints. Within a particular viewpoint however, prototypes may be formed to allow for moderate changes in appearance. Hence, whilst debates have ensued between proponents of exemplar and prototype theories, it may be that both types of representation are used in familiar face recognition (Bruce & Burton, 2002).

### 1.3.2 Distinctiveness

It is clear that some faces are easy to recognise whilst others are more difficult. For example, someone might have a large nose which easily differentiates them from other people whilst others seem to be "one of the crowd". There have been many investigations into the distinctiveness of faces (e.g. Hosie & Milne, 1996; Valentine & Bruce, 1986a; Valentine & Bruce, 1986b; Valentine, 1991). Valentine and Bruce (1986a) measured how quickly people could identify whether a face was familiar or not. They found that faces that were considered distinctive were identified as familiar faster than faces that were considered more typical or average. However, when a face decision task was used (i.e. is the stimulus a face or a not a face) the opposite result was found with typical faces exhibiting faster reaction times.

Bruce and Valentine (1986a) argue that this switch in performance occurs because of the way we encode faces and recognise them. Each face we encounter is compared against an average face created from all known faces and the difference between the average face and the presented face is calculated. The greater the similarities between a target face image and the average face then the harder the recognition task becomes, resulting in longer response times. Changing the task, however, to deciding whether a stimulus is a face or not results in faster reaction times for typical faces as they are closer to the stored representation of a face (the average) than distinctive faces are and are therefore more "face-like".

Valentine (1991) formalised the results of previous work on distinctiveness and proposed that all faces exist within a multi-dimensional space called face-space. This space codes the various dimensions on which faces differ, for example nose length or age (Bruce & Young, 1998). The theory of face space has gained support from the work on caricatures. Caricatures exaggerate features of the face to make them more distinctive. Conversely, anti-caricatures make the face a lot more "average looking" and therefore, more typical. It would be expected that caricatures would be easier and faster to identify (due to the large differences between the caricatured face and stored typical faces) than a normal image of the face. The anti-caricature should be the hardest image to process. Rhodes, Brennan and Carey (1987) examined performance on the recognition of famous faces from caricatures and anti-caricatures. As predicted by face space theory, they found that caricatures were recognised faster than the original image of the face. Furthermore, they also found that the original image was recognised faster than an anti-caricature image.

Whilst face space has been taken to provide an explanation of experimental findings, there have been criticisms regarding the idea that typical faces are located close together towards the centre of face space (Burton & Vokey, 1998). Burton and Vokey (1998) point out that in typicality measures, faces tend not to be clustered around the typical end of the scale but are instead located somewhere in the space between typical and distinctive. Very few faces are ever regarded as extremely typical or extremely distinctive.

## 1.4 Discriminating between individual faces

The review of how faces might be recognised (see sections 1.2 and 1.3) considered the nature of the stored representations of faces. A second issue that needs to be considered is what is it about a face that allows us to tell one person from another and enables us to individuate people when all faces share the same overall appearance (i.e. two eyes placed either side above a centrally placed nose and mouth)?

It has been suggested that the reason we are so good with recognising familiar faces is that we are experts at familiar face recognition (e.g. Diamond & Carey, 1986; Mondloch, Le Grand, & Maurer, 2002) and even that we have some level of innate knowledge about faces (Johnson, Dziurawiec, Ellis, & Morton, 1991; Mondloch et al., 1999) as it has been shown that that newborn infants look longer at a face like pattern of dots (i.e. two dots representing the eyes placed above a single dot representing the mouth) compared to a pattern of dots arranged in a random order (Johnson et al., 1991).

### 1.4.1 Face processing – configurations and features

The preference infants have for a face like pattern of dots over a random pattern of dots, as reported by Johnson et al. (1991) suggests that infants prefer to look at something that has the same configuration as a face. This is interesting because it suggests that there is something important about the arrangement or configuration of the features of a face and there is persuasive evidence that faces are not processed by their individual features alone. For example, Tanaka and Farah (1993) reported that people found it hard to identify faces on the basis of individual features in isolation compared to other objects such as houses. Instead, it would appear that the overall configuration of a face is important for successful recognition. Maurer et al. (2002) identify three ways in which we can process the configuration of a face; first-order relationships which define the overall layout of a face (as in the case of the face patterns), holistic processing in which the face is processed as a whole rather than as specific features, and finally second-order relationships which define the relationships between the individual features of the face.

First-order processing refers to recognising a stimulus as a face due to the location of two eyes spaced either side of and above of a nose which, in turn, is above a mouth. There is evidence from adults that indicates that people are very good at detecting faces (whether real or not) based upon this relationship alone (Mondloch et al., 2002). For example, The Vegetable Gardener, a painting by Archimbaldo, looks like a bowl of fruit when viewed in one orientation. Invert the image however, and the image of a face appears, based upon the apparent configuration of a face (see Figure 1-2).

**Figure 1-2: Arcimdolo's The Vegetable Gardener (c. 1590). The image on the right appears to be a bowl of vegetables. Once the image is inverted however (left), the image of a face becomes apparent.**

As well as processing features and the relationship between them, there is compelling evidence that faces are processed as a whole (or holistically), rather than by individual components. Young, Hellawell and Hay (1987) constructed composite faces, by combining the top half of one famous face (incorporating the hair, eyes and top of the nose) with the bottom half of another famous face (comprising the bottom of the nose, the mouth and jaw). When the two halves of the face were presented in alignment, people had great difficulty in identifying the two individuals. If the two halves were offset however, people found it much easier to identify the two celebrities. Young et al. argue that the reason people have difficulty in naming the identities of the two halves when they are in alignment is because the two halves combine to produce a new configuration which gives rise to the perception of a new unfamiliar face. Separating the two halves disrupts this holistic processing so that the identification task becomes much easier.

Whilst first order relations allow a face to be identified as a face, they do not allow for individuation between faces. Instead, second-order relationships, the relationships between the features of the face which will be subtly different for each individual, can be used to identify known faces (Diamond & Carey, 1986). Peoples' sensitivity to second-order relationships has been shown to be sufficient for even small changes in the spatial layout of a face to be detected (Haig, 1984). There appears to be a relationship between the ability to detect changes in the spacing of the features of the face and our expertise with faces (Bruce & Young, 1998; Diamond & Carey, 1986). Typically we see face in an upright orientation; we rarely have experienced faces in an inverted orientation. If expertise and the ability to detect second-order changes are related, then inversion might give a useful insight into how we differentiate one face from another.

### 1.4.2 The role of expertise in face processing

There are situations when the processing of faces is disrupted. Probably the best known of these is the inversion effect (Diamond & Carey, 1986; Valentine, 1988; Yin, 1969). Yin (1969) reported that when presented upside-down, faces were recognised significantly less well than other objects, such as houses or aeroplanes. It has been suggested that the difficulty observed in identifying faces after inversion is due to the disruption of second-order configural processing. For example, when Young et al. (1987) presented participants with inverted composite faces, recognition of the two halves of the face became much easier than when the composite was presented in an upright orientation. Young et al. argue that the face composite effect disappears after inversion because the new facial configuration, created by aligning two halves of different faces, can only be perceived when the face is presented in an upright orientation. Inversion of the face disrupts second order configural processing so that a new configuration cannot be created.

It has also been suggested that sensitivity to changes in second-order relationships is due to our expertise with faces (Diamond & Carey, 1986). Diamond and Carey (1986) tested this by hypothesising that if the inversion effect is the result of expertise, then people who are experts in a domain other than faces should also demonstrate an inversion effect for the subject of their expertise. Diamond and Carey tested dog experts on their recognition of particular breeds of dog for both upright and inverted dogs. They found that the experts' recognition of inverted dogs was significantly impaired, compared to non-experts and was comparable to their recognition performance on upright and inverted faces.

Further evidence for the inversion effect being the result of experience with processing upright configural information comes from the Thatcher illusion (Thompson, 1980). This illusion, originally depicting the former British Prime Minister, is created by taking the eyes and mouth and inverting them. The result is a face that appears grotesque in appearance when it is viewed in an upright orientation but looks normal when the face is inverted.

Why should faces be so hard to identify and exhibit impairments in processing as demonstrated by the Thatcher illusion when inverted, yet in the composite faces used by Young et al. (1987) become easier to identify the identities of the two individuals? Bartlett and Searcy (1993) suggest that in an inverted face, recognition of the identity carrying second-order relationships is impaired, and the individual features of the face are processed separately. In an upright face the face is processed as a whole and the relationships between the features is now of importance.

The point at which the disruption of second-order relationships takes place, such as the point at which the grotesqueness of the Thatcher illusion appears, has been debated recently. Stürzel and Spillmann (2000) reported that a face must be rotated through approximately 94° for the grotesque appearance of a Thacterised to appear. Lewis (2001) however reports that Stürzel and Spillmann's methodology overestimates the angle at which the change occurs and places the angle of rotation required at about 74°(Lewis, 2003). The results reported by Lewis (2003) indicate that a face can still be considered more inverted than upright, yet the grotesqueness of the Thatcher illusion is visible.

Another indication of our expertise with faces comes from the interesting finding that people are better at recognising faces of their own race (Meissner & Brigham, 2001) and making judgements such as age (Dehon & Bredart, 2001) than for other races, an effect which has become known as the other-race effect. Chiroro and Valentine (1995) found that African observers who had considerable experience with Caucasian faces recognised African and Caucasian faces equally well. In contrast, Caucasian observers, who did not have much experience with African faces, were able to recognise Caucasian faces, but had difficulty with African faces. In a similar vein, Dehon and Brédart (2001) found that Caucasians living in Belgium were able to make accurate age estimations about Caucasian faces but not for African faces. African Belgians, who had been living in predominantly Caucasian Belgium for a number of years, were able to make accurate age judgements for both Caucasian and African faces. In both these studies the general argument is that sufficient experience with a particular race of face allows faces of that race to be processed well. Insufficient experience of a particular type of face will, however, reduce accuracy on tasks such as recognition and age estimation.

As we gain expertise with human faces, it appears that ability to process other species faces which share the same first-order configural layout as a human face (e.g. a monkey face) becomes impaired. Pascalis, De Haan and Nelson (2002) have demonstrated that up to 6-months of age, infants are able to successfully discriminate between examples of monkey faces as effectively as they could for human faces. However, by the age of 9-months and through adulthood, the ability to process monkey faces and human faces equally diminishes, and an expertise for processing human faces emerges.

## 1.5 The role of movement in learning new faces

It can be claimed that attempting to familiarise people with faces from a static photograph lacks ecological validity as when we encounter people in the real world they are very rarely, if ever, totally static. How important is movement in the learning of new faces? For example, do our representations of faces incorporate movement information, or is movement of little or no importance?

Bruce and Valentine (1988) investigated how useful an aid movement could be in the learning of new faces. They presented participants with a series of faces that were learned under three possible conditions; a video clip showing the face in motion, five still images created from the video and a single still image from the video. Bruce and Valentine found that there was no benefit for recognition from learning the face through a video clip over a single image or multiple static images, suggesting that movement is not an important factor in the learning of new faces.

Pike, Kemp, Towell and Phillips (1997) argued that the images used by Bruce and Valentine (1988) during the test phase were too different to allow for good generalisation as the studied images were full-face and the test images were three-quarter views. Pike et al. claimed that if side-to-side motion (i.e. from one profile view to the opposite profile view) was studied during the learning phase, then this would provide the participant with more views of the face that could aid recognition. In addition, movement from side to side would also give more three-dimensional structural information about the face, which also could be useful for recognition.

Pike et al. presented participants with three types of stimuli; moving video clips of faces rotating through 360°, or five static images taken from the video clip providing five different views of the face, or a single full-face image. They found that presenting a face in a video clip led to greater recognition levels of static images than learning either a single image or multiple static images. Pike et al. also found that this benefit for a dynamic display was not simply due to the presentation of more images of the face. They found that presenting participants with ten still images of a face led to the same level of performance as showing five still images, suggesting it was the motion itself that was aiding the learning of the faces as oppose to merely seeing more image of the face.

A crucial difference between the studies of Bruce and Valentine (1988) and Pike et al. (1997) is the type of motion that was used in the learning phase. The video clips used in Bruce and Valentine's study showed a face in the full-face view exhibiting expression changes (a type of motion termed non-rigid motion). Pike et al. on the other hand used angular rotation of the face from side-to-side (rigid motion). Therefore, it may be that any benefit to learning obtained through motion may be in the type of motion that was used. To investigate this, Christie and Bruce (1998) presented participants with either animated faces performing a rigid action (nodding and shaking of the head) or a non-rigid motion (a change from smiling to a sad expression or vice-versa), or a series of static images comprising the same number of frames as shown in the video footage. At test, faces were presented as static images or a single video sequence. They found no overall benefit for learning faces in motion. Interestingly, it was revealed that when rigid motion had been studied, effects of pose change were far smaller than if non-rigid movement had been studied.

It would appear therefore that rigid movement may provide some useful information during face learning. However, comparing the results of Pike et al. (1997) and Christie and Bruce (1998) is difficult. Roark, Barrett, Spence, Abdi and O'Toole (2003) point out that the types of rigid motion employed in the two experiments were fundamentally different. The nodding and shaking of the head used by Christie and Bruce is more socially engaging, and, as a result, more likely to be interpreted as movement by the subject (i.e. perceived as an actor moving in front the participant, rather than the participant moving around the actor). In the study by Pike et al., the actor sat passively on a chair that was rotated through 360°. This, Roark et al. claim, will be seen as movement of the viewer (i.e. the participant in the experiment moving around the actor) which may provide a more perceptually rich experience for the viewer, enhancing learning.

In conclusion, the work conducted on the role of motion in the learning of new faces has provided conflicting results (see Roark et al., 2003 for an overview) with some studies providing support for the theory that motion plays an important role in the forming of new representations of faces whilst other studies find that movement provides little or no assistance in the formation of new facial representations. However, it does appear that it is not simply movement *par se* that may be important in the learning of new faces, rather the type of movement.

## 1.6 Differences in facial cues used in the processing of unfamiliar and familiar faces

Do different parts of the face provide different levels of information for the recognition process? The central region of the face (i.e. the inner features), contains the eyes and mouth. These areas may be considered important for social interaction. For example, the eyes, and particularly eye-gaze, give an indication as to where another persons attention is directed whilst the lip-reading of mouth has a profound effect on our perception of sounds (McGurk & Macdonald, 1976). As a result, greater attention is paid to the internal features than to the external features during interaction with others. Indeed, studies examining eye saccades indicate that the internal features receive most attention when viewing faces (Althoff & Cohen, 1999; Luria & Strauss, 1978; Stacey, Walker, & Underwood, 2005). Hence it is plausible to suggest that faces that we are familiar with (and have therefore paid a lot of attention to) will be recognised better from the internal features than the external features.

The external features do however contain one particularly distinctive and therefore memorable feature – the hair. Hairstyle can change regularly though and it would make little sense to recognise people well known to us via this cue. As such, it might be expected that external features would be of more use when learning and recognising unfamiliar faces (where we need an instant, distinctive feature) than for familiar faces.

Ellis et al. (1979) presented participants with a series of unfamiliar faces and then gave them a recognition task in which either the whole face, the internal features only or external features only of the face were presented. In another experiment, participants were asked to identify famous faces from either the internal features of the face only or the external features of the face. They found that not only were familiar faces recognised better than unfamiliar faces, but also that the familiar faces were recognised significantly better from the internal features than they were from the external features. For unfamiliar faces on the other hand, there was no benefit observed for viewing the internal features. This study was carried out using Caucasian faces and observers and has been replicated using Japanese faces and observers (Endo, Takahashi, & Maruyama, 1984). Whilst Ellis et al. found an advantage for the recognition of the internal features over the external features of a familiar face, they did not find that the reverse was true; that is, for an unfamiliar face there was no advantage for the external features. Bruce et al. (1999) however did find such an effect. They presented participants with a probe face and an array of target faces, all of which were unfamiliar. They were told that the probe was present in the target array and asked to identify which face they though it was. Two different pictures of the person to be identified were used (one for the probe image and one in the target array of faces), thus participants could not rely upon simple picture matching. The faces were either presented as whole faces, the internal features only or the external features only. It was found that presenting the external features led to significantly better recognition accuracy than the internal features did and to a level of performance similar to when the whole face was presented.

Young, Hay, McWeeny, Flude and Ellis (1985) performed a similar experiment using a different method. They were concerned that the recognition task used by Ellis et al. (1979) employed the same photographs at study and test and such a method cannot distinguish between face learning and picture learning. Young et al. presented participants with a series of slides in which the internal or external features of a person familiar or unfamiliar to them was shown alongside a whole image of a face (which was either the same or a different person) and asked if the two images depicted the same person or two different people. It was found that reaction times for matches based upon internal features were significantly faster for familiar faces than for unfamiliar faces. For the external features, no difference was observed between familiar and unfamiliar faces.

Young et al. also examined whether this effect is present when the faces presented to the participant can be processed as pictures as opposed to faces by using the same image at study and test. The advantage in speed for the matching of familiar faces on internal features disappeared. Young et al. (1985) noted that overall latencies were shorter when the participant was required to match the pictures than when they were required to match different images, suggesting that different processes were at work when the task is to match faces as opposed to match pictures.

Finally, in a study conducted by Kemp et al. (1997), investigating the usefulness of photographs on credit cards to detect fraud reveals potential importance for one particular external feature of unfamiliar faces – the hair. Kemp et al. found that it was harder for their cashiers to detect fraud in female shoppers. They suggest that this may be because cashiers were paying more attention to the hairstyle in female shoppers and accepting cards that depicted another individual who had a similar hairstyle. Nearly all the male shoppers had short hair and so cashiers might not have found this cue to be so useful and hence, focussed on other features which may be more indicative of identity. In support of this, Shepherd, Davies and Ellis (1981) suggest that the hair is one of the most important cues in recognition.

It would appear then that the internal features of a familiar face are recognised better and faster than the external features and for an unfamiliar face there is no speed advantage for matching external features over the internal features but there is some evidence for matching accuracy. This situation would seem to make logical sense. Attempting to recognise those known to use merely by the use of the external features could lead to some serious problems if one of these people decides to get a haircut!

The internal features consist of the eyebrows, eyes, nose and mouth and these features themselves might carry different weights in the recognition of familiar faces compared to unfamiliar faces. This possibility was investigated by O'Donnell and Bruce (2001). They found that in faces in which the eyes had been altered, this change was detected more accurately for familiarised faces than for the unfamiliar faces, suggesting that as a face becomes more familiar, the eyes play a greater part in the recognition process.

O'Donnell and Bruce (2001) do not explicitly state whether when making changes to the eyes it was indeed just the eyes that was altered or whether the eyebrows was included as part of the eyes. Whether the eyes and eyebrows contribute equally to usefulness of the eyes for recognition was investigated by Sadr, Jarudi and Sinha (2003). They studied recognition performance for famous faces after the face had either the eyes or eyebrows removed. They found that not only removing the eyes led to a decrease in recognition accuracy but that removing the eyebrows also led to a significant decrease in performance. Additionally, it was found that performance after the removal of the eyebrows was significantly worse than after the removal of the eyes.

### 1.6.1 Tracking the shift towards an internal feature advantage

It would appear that overall, if a face is familiar then the internal features provide a more accurate cue for identification than the external features. Angeli, Bruce and Ellis (1998) examined the development of this advantage. On nine consecutive days, participants viewed video clips of faces and were given other information about the people depicted in the clips. Participants then completed a face matching task using static images. The results over the nine days indicated that recognition performance for the internal features alone and the external features alone was initially similar.

From the fourth day onwards, however, the internal features did seem to be recognised better than the external features.

Bonner, Burton and Bruce (2003) conducted a similar study to track the apparent shift towards internal features for recognition by training participants to recognise a series of faces from video clips or still images over a period of three consecutive days. On each day, the participant viewed video clips of faces, some of which were moving whilst others were static images, after which they performed a face matching task using static faces. In this task, participants were presented with pairs of faces which could be either one of the learnt faces or a novel face. The participants' task was to indicate whether they thought the two people were the same. They found that over the duration of the experiment, matching accuracy on the internal features improved for the familiarised faces compared to the novel faces.

Clutterbuck and Johnston (2002) examined whether the internal feature advantage for familiar faces was an all-or-nothing effect or whether the advantage gradually increases as a face becomes more familiar. Using three groups of faces (highly familiar, moderately familiar and unfamiliar), participants were presented with a whole face and either the internal or external features of the same person or a different person. They were then asked to state whether the two people were the same or different. For the whole face-internal feature pairs, they found that highly familiar faces were matched faster than moderately familiar faces, which in turn were matched faster than unfamiliar faces. For the whole face-external feature pairs, no effect of familiarity was found. This result suggests that the internal feature advantage is not an all-or-nothing effect and may develop over time, as a face becomes more familiar.

Clutterbuck and Johnston extended this work to examine the usefulness of the matching task as a way of assessing familiarity of new faces learned over a period of two days (Clutterbuck & Johnston, 2004). In a result similar to their previous study, they found that when participants were presented with the internal features alone, famous faces were matched faster than either faces with which the participant was familiarised with over two days or novel faces. The familiarised faces were, in turn, matched faster than the novel faces.

In summary, as faces become more familiar there appears to be a gradual shift towards recognition being mediated by the internal features, with the eyes playing a dominant role in the identification process. This shift does not appear to be an "all-or-nothing" effect but rather a gradual change over time.

## 1.7 Modelling face recognition

For those who are familiar to us, what is the process for recalling their names, occupations and so on? Throughout the 1980s a number of models were proposed (Bruce, 1983; Bruce & Young, 1986; Hay & Young, 1982). The most influential model to date was proposed by Bruce and Young (1986) (see Figure 1-3).

### 1.7.1 The Bruce and Young model of face recognition

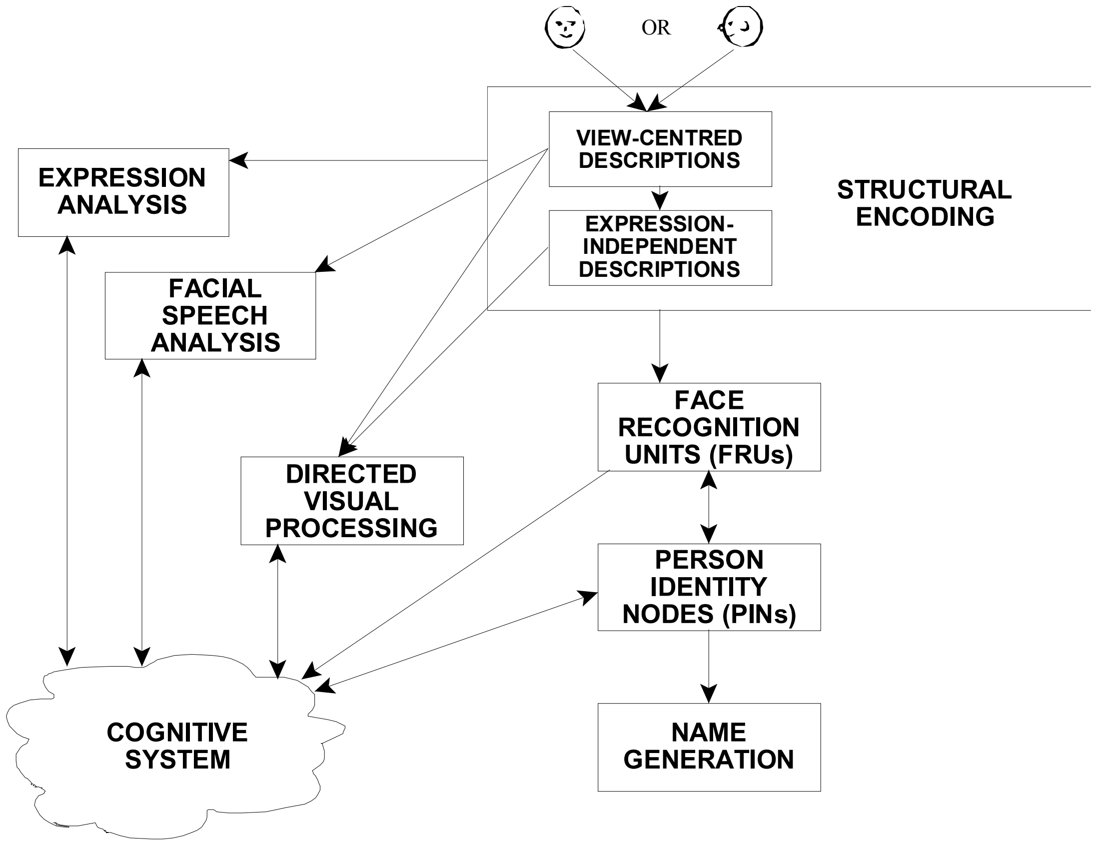


**Figure 1-3: Bruce and Young's model of face recognition (reproduced from Bruce and Young, 1986).**

Bruce and Young's model can be split into two distinct paths which both starting with a common initial process, *structural encoding*, which produces *view centred descriptions* and *expression independent descriptions*. The first pathway incorporates the processes of *expression analysis*, *facial speech analysis* and *directed visual processing* which are applicable to both familiar and unfamiliar faces as the identity of the face is not important for these tasks. The process of expression analysis allows for the recognition of facial expressions via the configuration of various facial features whilst facial speech analysis uses the movements of the mouth and tongue to achieve its goal. The third of these processes, directed visual processing, is used to provide selective attention to the visual form of the face and thus has an important role to play in comparing or remembering unfamiliar faces.

The second pathway, from the *face recognition units (FRU)*, through to the *person identity nodes (PIN)* and finally *name generation* stage does require a stored identity of the person and hence is applicable to familiar faces. Bruce and Young (1986) claim that abstract *expression independent descriptions*, provide structural information about the face, which are fed into the FRUs. For each known individual known there is a single FRU which is activated upon the presentation of that face. The strength of this activation is "dependent on the degree of resemblance between its stored description and the input provided by structural encoding" (Bruce & Young, 1986 p. 311 - 312). Hence, according to the model, successful recognition will rely upon how similar a stored three-dimensional representation is to the three-dimensional information extracted from the image. The activation of the FRU then allows for the retrieval of semantic information about the individual, such as their occupation (from the PINs), which in turn, allows for the retrieval of the persons name. The PINs do not require a facial image input however, and can also be activated via the *cognitive system* (through cues such as gate and voice) as well as the FRUs.

The model has been used to explain many failures of face recognition. For example, everyday failures to recall any information about an individual despite knowing that the face is familiar are explained by the FRU not activating the corresponding PIN for that person. Another difficulty that many people will probably remember happening to them at least once is the "tip-of-the-tongue" phenomenon, whereby a face is correctly recognised and identity information is retrieved (e.g. saying "I know he's a politician") but being unable to recall the person's name (e.g. being unable to say "It's Tony Blair"). This type of error indicates that an FRU has been activated (e.g. for Tony Blair), which activates his PIN to enable recall of the fact that he is a politician (as well as he is the Prime Minister etc.). However, it has not been possible to access the name generation stage and thus his name cannot be recalled. In a study by Young, Hay and Ellis (1985) in which people kept diaries of their face recognition failures, whilst cases of people being unable to name someone but recalling other information about them did occur, the reverse never happened. That is, people were never able to produce the name of someone without providing some other semantic information about them (in a current example people would not say "that's Tony Blair" without being able to recall he was a politician).

Bruce and Young (1986) also cite neuropsychological evidence to support the model, for example, double-disassociations in patients suffering from prosopagnosia, a neurological condition in which people are unable to identify familiar faces (Ellis & Florence, 1990). Malone, Morris, Kay and Levin (1982) reported two patients who exhibited two contrasting impairments of face recognition. The first could recognise familiar faces well but could not match unfamiliar faces, whilst the second patient could match unfamiliar faces but could not recognise familiar people. Such a disassociation would be predicted by the model as familiar and unfamiliar face recognition is undertaken via separate pathways. Similar dissociations have been observed for the processing of facial expressions and facial identity providing support for the notion that identity and expression are processed separately (but see Calder & Young, 2005).

### 1.7.2 Computational modelling of face recognition and learning

Burton, Bruce and Johnston (1990) have successfully implemented a section of the Bruce and Young (1986) model in a connectionist network. Their Interactive Activation and Competition (IAC) model comprises a number of pools of units which correspond to components of the Bruce and Young model (Figure 1-4). The connections in the model are bi-directional, and are inhibitory between the units within each pool, and either inhibitory or excitatory between pools. The pools consist of groups of units that correspond to certain pieces of information about people. The pool of FRUs contains units that respond to the visual image of a known face. Each FRU is connected to a unit in the PIN pool that relates to a particular individual. The units in the PIN pool are further connected to two pools of units; Name Input Units (NIUs) which are dedicated to written or spoken names, and Semantic Information Units (SIUs), which specify occupations, interests and so on.

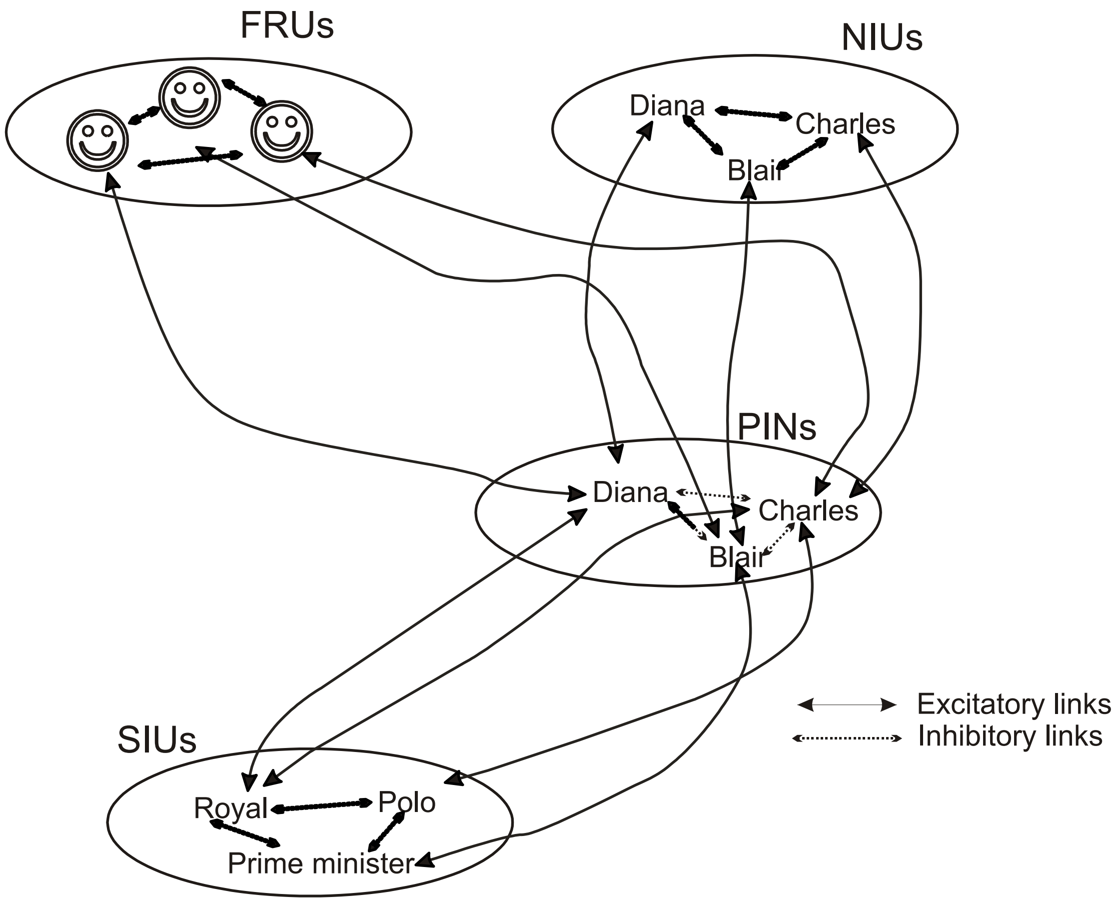


**Figure 1-4: The central architecture of Burton et al.'s (1990) IAC model of face recognition (reproduced from (Bruce, 1995)).**

Presentation of a known face results in the FRU most associated with that face becoming active. The inhibitory link between that FRU and all the other FRUs results in the other FRUs reducing in activity, eventually leading to a situation where only one FRU is active. An excitatory link from the selected FRU excites a particular PIN, which, in turn, actives associated SIUs and NIUs that enables recollection of names, occupations and so on.

The nature of the links within and between the pools of units can explain two phenomena found in human participants; the semantic and repetition priming of faces. In semantic priming, prior presentation of the face of someone closely related to the target face speeds response times (Bruce & Valentine, 1986). The IAC models this elegantly as demonstrated by this example involving Charles and Diana. If Charles's face is presented, the FRU for his face becomes active and excites the PIN associated with him. The PIN, in turn, activates the SIUs "polo" and "royal". The royal SIU is also connected to Diana's PIN and thus receives activation. This level of activation will be higher than all of the other PINs (bar Charles') so that if Diana's face is later presented, the PIN associated with her will already be somewhat active, leading to a shorter amount of time required for her to be identified as familiar.

Repetition priming, in which a face is recognised faster, or more accurately, on a second exposure to the face than on the first, is also simulated by the IAC model. After a known face has been presented, the link between the FRU for that face and its corresponding PIN becomes strengthened. After the Charles PIN returns to its resting level of activation, the strengthened link between the appropriate FRU and PIN units remains, meaning that re-presenting Charles' face will result in Charles' PIN reaching the required level for activation faster than it would have done had the link not been strengthened.

The IAC model has been extended to encompass the learning of new faces in the form of the IACL (IAC with Learning) (Burton, 1994). In his demonstration of how the IAC model could learn new faces, Burton created a pool of 100 FRUs. Half of these units were strongly associated with particular patterns drawn from 12 input feature pools which were used to simulate the factors that might be used to discriminate between faces. The connections from the feature pool units to the other half of the FRUs were set to random values and these FRUs were effectively set aside for faces that had not been learnt yet. Burton (1994) demonstrated that the model could learn to recognise new input patterns which became associated with one of the un-used FRUs. Importantly, the FRUs representing known faces were left unchanged by the learning process.

In addition, it was also shown that other known effects of face processing could be simulated, for example, the composite face effect reported by Young et al. (see section 1.4.1). In his simulations, Burton found that presenting half of an input pattern in isolation enabled the FRU associated with that part of the pattern to become active. Combining two halves of different known input patterns put together resulted in the network learning the input as a new face and becoming associated with an unused FRU.

### 1.7.3 Computer recognition of faces from images

Both the Bruce and Young model and the IAC model require an input to the recognition process for known faces through the FRUs. However, little is known about how an FRU becomes active. If faces are indeed recognised via images as suggested by work on unfamiliar face recognition, image processing and statistical analysis techniques may be able to uncover the mystery as to how an FRU becomes active from an image.

The primary method of investigation into computer recognition of faces has focused upon principle components analysis (PCA). The general principle of PCA is to analyse data, which may vary on a large number of dimensions, and reduce them down to a smaller number of dimensions which can be used to describe the data set. Applying this to faces, all faces can be considered to exist in multidimensional space. For example, of these dimensions, one may be for the gross variations in hairstyle, another might be for face size (Bruce & Young, 1998).

Turk and Pentland (1991) described how PCA could be used for face recognition using computer images as their input. The pixels of a facial image, $n$ x $n$ pixels in size, are reshaped to form a single vector of size $n^2$. This vector refers to the face's position in $n^2$ dimensional space. The goal of PCA is to reduce this very large dimensional space (with a small image size of 128x128 pixels, each face occupies one point in 16,384-dimensional space!) to a much smaller dimensional space, with each dimension coding something specific about the faces. These dimensions are created by analysing a number of facial images (the training set). Each face can then be placed into this smaller dimension face space. In the recognition process, a new test facial image is projected into this face space and if the face is located sufficiently close to one of the training set images then the face is considered to have been identified as that person.

PCA is a particularly appealing technique for modelling face recognition for two reasons. Firstly, whilst cognitive and computational models have made overall suggestions as to how the face recognition process operates, neither has provided a detailed explanation of how an FRU becomes active. PCA provides an explicit description of how a particular FRU can become active from a stimulus. Secondly, unfamiliar face recognition, and to a certain extent familiar face recognition, appear to be image-based. PCA mimics this process by also operating on low-level image properties. However, proponents of PCA stress that it is not intended as an implementation of human face recognition (Burton, 1994; Burton et al., 2005; Calder & Young, 2005). Instead, PCA seems to extract components of the image that humans might use to discriminate faces. For example, O'Toole, Abdi, Deffenbacher and Valentin (1993) have demonstrated that early components appear to code the sex of the face (Bruce & Young, 1998).

Typically, the training set of faces used for PCA contains only one image per person. Burton et al. (2005) report an interesting study in which multiple images (1, 3, 6 or 9 images), or an average image created from these images, of the same person were used during training. Burton et al. compared the results of PCA simulations in which either separate images (to model an exemplar system of recognition) or the average of the input images (to model a prototype system) were used. They found that using the average of the input faces as the training faces produced better performance than using separate images (with the exception of when only one image was used, as in this case the average image is identical to the single exemplar). Additionally, they also found that the PCA system performed better the more images were used. Whilst both exemplar and average systems benefited from the use of multiple images, the average model benefited more from these extra images. However, even the best performance was limited at around 40% accuracy.

Burton et al. repeated the simulations with a greater number of training images (3, 6, 9 or 19 images) and found that hit rates of the PCA system rose to 75% when averages made from 19 images were used. Whilst these data look convincing to support a prototype model of face recognition, Burton et al. stress that the data do not imply that a prototype system is superior to an exemplar system and they doubt whether any such assertion is possible. However, it is clear that averaging removes some of the pictorial dependency whilst retaining face identity information. Therefore, averaging appears to be a promising technique for future research into both familiar and unfamiliar face recognition.

## 1.8 Overview of the current work

All the models developed thus far for face recognition have been useful in understanding and even predicting performance on face recognition tasks. However, whilst models such as the IACL demonstrate how FRUs can become associated with semantic information about a face, none of the models explicitly state what form the input pattern to an FRU takes. Is FRU activation achieved via a viewpoint-invariant structural model, or are there many different FRUs for the same person, each one tuned to a particular view or image? As seen in the Bruce and Young (1986) model, the FRU is a crucial component in face recognition, yet little is known about it.

A key property of familiar faces is that they are recognised well despite changes in pose, illumination and expression (Bruce, 1982). Bruce and Young (1986) propose that this recognition is due to the formation of a three-dimensional structural code of the face which can be rotated to allow recognition in novel views and it is this structural model that is associated with a FRU. However, previous work on object recognition has revealed that this may not be the case and instead, recognition may be mediated by a limited number of stored views. In this case, a FRU will not be linked to a single entity but rather to a number of instances of the face.

This thesis examines the nature of face representations. The experimental work is split into three chapters. Chapter Two examines what can be learned about a face from photographs. Typically, research into face learning has provided participants with only a single exposure to a face. However, such a learning technique does give sufficient experience of a face to allow for the formation of a robust representation of the face. The first experiment of Chapter Two investigates whether face recognition after a change in pose or lighting is improved once participants are given ample opportunity to learn face images presented to them. The second experiment examines in greater detail recognition performance when the pose of a face is changed between study and test. The final three experiments of Chapter Two look at whether two separate images of a face can be successfully combined to produce a more robust representation of a face than only a single view allows. It addition, it examined whether providing two images during learning promotes the learning of a single individual or two separate pictures.

Chapter Three reports two experiments aimed at separating two possible image-based methods of face recognition; viewpoint-dependent coding and pictorial coding. This was done by investigating recognition accuracy after transformations which alter the image properties (and therefore disrupt any pictorial codes that may be used) but leave the viewpoint unchanged. Chapter Four investigates whether the recognition process requires the previously seen image and the test image to be identical in every way. This was examined using one transformation that does not appear to have any effect on the recognition of familiar faces; the transformation of colour.

Chapter Five shifts the focus to examine the features of the face that are useful for recognition. The first experiment of Chapter Five examines whether newly learnt faces demonstrate the internal feature advantage, typically seen for familiar faces. The remaining experiments in Chapter Five investigate one particular feature of the face that appears to draw extra attention – the hair, and explores how useful a cue the hairstyle is for recognition and whether (paradoxically) enhanced levels of recognition can be obtained if the hairstyle is removed. Finally, Chapter Six summarises the main findings of the current work and concludes by discussing possible future directions stemming from the research presented in this thesis.

# CHAPTER TWO

## Learning faces from photographs

## 2.1 Introduction

The predominant paradigm in face learning experiments remains that of testing recognition memory for an unfamiliar face after a single exposure to a single photograph of a face. This has obvious advantages for the control of experimental factors (for example, control over the number of exposures and views of the face seen by the participant), but faces are not learnt in this way. Firstly, familiar (and therefore learnt) faces have been seen many times as opposed to only once. Secondly, familiar faces have been seen in many different views. For example, in the case of people we are interacting with there is constant change in the appearance of their face due to movement of the head or a change in their expression, resulting in many different images of the face being seen. Even in the case of celebrities, which are usually the faces of people we have never personally met, we are likely to be exposed to a large number of images of their face.

The fact that many experiments have only used a single exposure to a single view of a face during learning limits the generalisation of the results of such studies to the process of face familiarisation in the real world. Five experiments were conducted to examine how well faces from photographs can be recognised after they had been learnt to a high level. In Experiment 1, recognition memory for faces after the lighting direction or pose of a face was changed between study and test was compared to recognition memory for the studied image. Experiment 2 examined whether the previously reported decrement in recognition accuracy after a pose change was consistent regardless of the angle of rotation, or whether accuracy became a function of the extent of rotation away from the learned image. Experiments 3, 4 and 5 compared the effects of providing multiple exposures to more than one image in contrast to a single image of the face to be learnt.

## 2.2 Experiment 1 – The effect of multiple exposures to a facial photograph during the learning of novel faces

Experiment 1 examined whether training people to recognise previously unknown faces to a high level through the use of repeated exposure to the faces helps the extraction of invariant information about the face so that it can be successfully recognised from a novel image.

Previous research has shown that after learning an unfamiliar face from a single exposure, a transformation of pose (Bruce, 1982; Krouse, 1981) or of lighting (Braje, Kersten, Tarr & Troje, 1996; Braje, Kersten, Tarr & Troje, 1998; Braje, 2003; Hill & Bruce, 1996) results in a significant drop in the recognition accuracy for that face. It is therefore predicted that recognition accuracy will be lower after the lighting and pose transformations than it will be for the same image following single exposure learning. However, if increased learning of the photograph used enables participants to encode invariant information about the face, these differences should not occur when multiple exposures are given.

## 2.2.1 Method

### 2.2.1.1 Design

The experiment consisted of three phases; a first presentation phase, a training phase and a testing phase. Participants were assigned to one of two groups. One group received only a single exposure to each face and completed the first presentation and testing phases only (no-training group) whilst the other group also received multiple exposures to the faces via an interactive training technique which involved the participant naming the faces (training group). During the test phase, all participants were tested on their recognition accuracy for each face in its original form (i.e. same image), after a lighting change, and after a pose change. The experiment therefore had a 2x3 mixed factorial design with the training type (no training vs. training) and image transformation type (same image, lighting change and pose change) as independent variables. The dependent variables were the number of faces recognised at test (hits) and the number of distractor faces correctly rejected (correct rejections) which were used to calculate an *A*' statistic, a non-parametric signal detection measure (Rae, 1976).

#### 2.2.1.2 Participants

Twenty-four undergraduate students (5 males and 19 females) aged between 18 and 21 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision.

#### 2.2.1.3 Materials

Images of 48 faces (36 male, 12 female) from the Pose, Illumination and Expression (PIE) face database (Sim, Baker, & Bsat, 2002) were used. None of the images used depicted the individual with facial hair or wearing glasses. Each face was used in two lighting conditions and two poses, resulting in a total of 192 images. In the two lighting conditions, the light source was located either directly in front of the model (0°) or to their right (30°). The images for the two poses were also taken from directly in front of the model (0°) or to their right (31°). The 48 faces were placed into four sets of 12 faces each. Faces were allocated to each set on the basis of how hard the face was to learn, such that all four sets of faces had approximately the same level of difficultly based upon pilot data. Each set contained nine males and three females. Half the faces in each set were Caucasian and the other half Asian. Figure 2-1 shows an example face in these pose and lighting conditions.

**Figure 2-1: Examples of the different poses and lighting conditions used in Experiment 1; from left to right - 0°/0°, 0°/30°, 31°/30°, 31°/0° (camera angle/lighting angle).**

Each image was manipulated to remove all irrelevant background information, leaving only the head visible. The background was replaced with a homogenous grey. The original colour images were converted to greyscale. Each image was resized so that it was 384 pixels high in order to normalize face height and the background was expanded horizontally to create a final image of 384x384 pixels subtending a visual angle of 4.87° when viewed from a distance of 60cm.

#### 2.2.1.4 Apparatus

The faces were presented on a 17" LCD flat screen monitor, set to a resolution of 1280 x 1024 pixels and a colour depth of 32 bits per pixel using a custom written computer program created in the Microsoft Visual Basic programming language. Participants made their responses through the use of a standard mouse.

#### 2.2.1.5 Procedure

Participants were randomly assigned to either the single (no training) or multiple (training) exposure conditions. For half the participants, the first set of faces was allocated as the target set and sets 2, 3 and 4 were allocated as distractor sets. For the other half, the second set was allocated to be the target set and sets 1, 3 and 4 acted as distractors.

All participants completed a first presentation phase and a test phase of the experiment. Those in the multiple exposures condition also received training trials between the first phase and the test phase. The participant sat in front of the computer screen at a distance of approximately 60cm and was given written instructions before the experiment began.

*2.2.1.5.1 First presentation phase*

During the first presentation phase, participants saw 12 faces for a duration of 5 seconds each, with 0.5 seconds between each face. The faces were evenly distributed across the two views (full-face and three-quarter view) and two lighting conditions (0° and 31°) so that three faces in each pose/lighting combination were presented. Each individual face was presented to the participant once and was accompanied by a first name, presented below the image of the face. These name/face pairings were randomly generated for each participant from a set of 12 fixed names (e.g. John, David). The only limiting factor on the pairings was that male names were only assigned to male faces and likewise for females.

*2.2.1.5.2 Training phase*

The training task was completed by the multiple exposures group only and was divided into two parts. In the first part, the 12 face photographs shown during the first presentation phase were divided into three blocks containing four faces each. In the second part, all 12 face photographs were presented in a single block. All participants completed the training in the same order – the three blocks first containing 4 faces followed by the single block with all 12 faces. The task was the same for both parts.

The training task required participants to match a name to a presented face. Name options were given in the form of on-screen buttons located below the image of the face and participants were required to click on the name that they believed belonged to the face. Only male names were presented for a male face and only female names for females. After a response was made, immediate feedback that took the form "Yes, this is David" (correct answer) or "No, this is Robert" (for an incorrect answer) was given as appropriate. To move from one block of trials to the next, participants were required to correctly name all the faces in the presented block, without making an error, on 3 separate occasions. In the event of the participant making an error, the faces that were correctly named were removed from the set and the remainder re-presented. This process was repeated until all faces in the block of trials had been correctly identified, upon which all the faces were re-entered into the set to begin the next block of training trials. To complete the training, the entire set of 12 faces had to be identified without error three times.

This name-face pairing task was used merely to ensure that the participants had successfully individuated the different facial images. Participants were not tested on their knowledge of the name/face pairs after the training phase.

*2.2.1.5.3 Test phase*

The testing phase was divided into three blocks of 24 individually presented faces, for which the participant made a "yes/no" decision as to whether they had seen the presented face during familiarisation. For each part, 12 of the individuals were those in the familiarisation set, whilst the other 12 were taken from one of the distractor sets. In two of the blocks a transformation was applied to the target images. This was either a change in lighting or a change in pose from the face presented in the learning phase (for example, if a participant learned a face with full-face lighting (0°) and in full-face view (0°), changing the lighting resulted in an image with three-quarter lighting (30°) and full-face view (0°) whereas changing the pose would result in an image with full-face lighting (0°) and three-quarter view (31°) ). In the other block the same image that the participant learned was presented. Blocks were rotated for each participant in a Latin square design. Faces were presented one at a time and two buttons (labeled "Yes" and "No") were used for responses. Participants were required to click on "Yes" if they thought they recognised the individual as a member of the training set and "No" if they did not. They were not required to recall or recognise the name of the individual in any of the test trials.

### 2.2.2 Results

The mean *A*' and hit rates obtained for single and multiple exposure conditions are shown in Figure 2-2 and Figure 2-3 respectively.

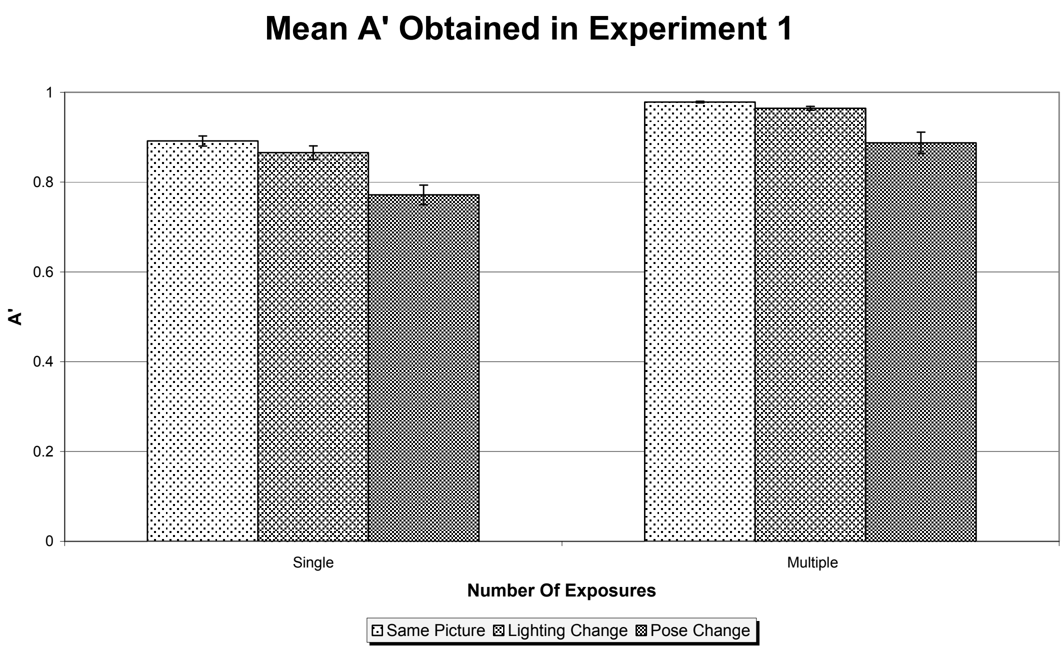


**Figure 2-2: Mean *A*' obtained in Experiment 1 for the three testing conditions (same picture, lighting change and pose change) after a single exposure or multiple exposures. Error bars represent standard error.**

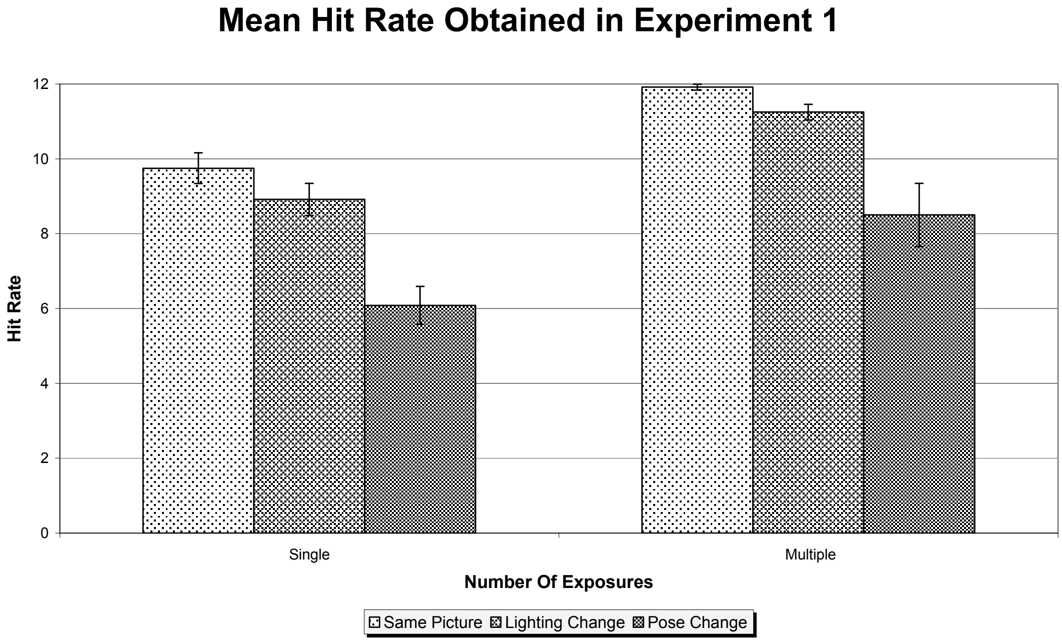


**Figure 2-3: Mean hit rate obtained in Experiment 1 for the three testing conditions (same picture, lighting change and pose change) after a single exposure or multiple exposures. Error bars represent standard error.**

### 2.2.2.1 Hit rate analysis

The hit rates were entered into a mixed design two-way ANOVA with training condition (single or multiple presentation, between-subjects) and transformation type (pose or lighting, within-subjects) as independent variables and number of hits as the dependent variable. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There were significant main effects of training condition; $F(1,22) = 23.66$, $MSE = 12.13$, $p < .001$, $\eta_G^2 = 0.431$ (observed power = 1.00) and transformation type; $F(1.42,31.27) = 33.06$, $MSE = 3.56$, $p < .001$, $\eta_G^2 = 0.307$. The interaction did not reach significance ($F < 1$). Planned contrasts revealed that performance after the pose change was significantly worse than with either the same image or after the lighting transformation; $F(1,22) = 37.11$, $MSE = 25.94$, $p < .001$. Performance after the change in lighting was found to be significantly poorer than on the same image; $F(1,22) = 9.19$, $MSE = 1.47$, $p < .01$.

#### 2.2.2.2 *A'* analysis

The calculated *A'* scores were entered into the same ANOVA design as the hit rate data. There were significant main effects of training condition; $F(1,22) = 39.33$, $MSE < 0.01$, $p < .001$, $\eta_G^2 = 0.571$ (observed power = 1.00) and transformation type; $F(2,44) = 31.41$, $MSE = 0.01$, $p < .001$, $\eta_G^2 = 0.268$ (observed power = 1.00). The interaction did not reach significance ($F < 1$). Planned contrasts revealed that performance after the pose change was significantly worse than performance with either the same image or after the lighting transformation, $F(1,22) = 37.77$, $MSE = 0.02$, $p < .001$. Performance after the change in lighting was found to be significantly worse than performance on the same image; $F(1,22) = 5.16$, $MSE < 0.01$, $p < .05$.

#### 2.2.2.3 Analysis by race of face

Many studies have shown that participants are more accurate at recognizing faces of their own race than of other races, producing a benefit in recognition accuracy at test for participants' own race faces. It is therefore possible that the Caucasian and Asian faces used in this experiment might have produced a different pattern of results in terms of recognition accuracy. The mean hit rates obtained for both Caucasian and Asian faces are shown in Figure 2-4 and Figure 2-5 respectively.

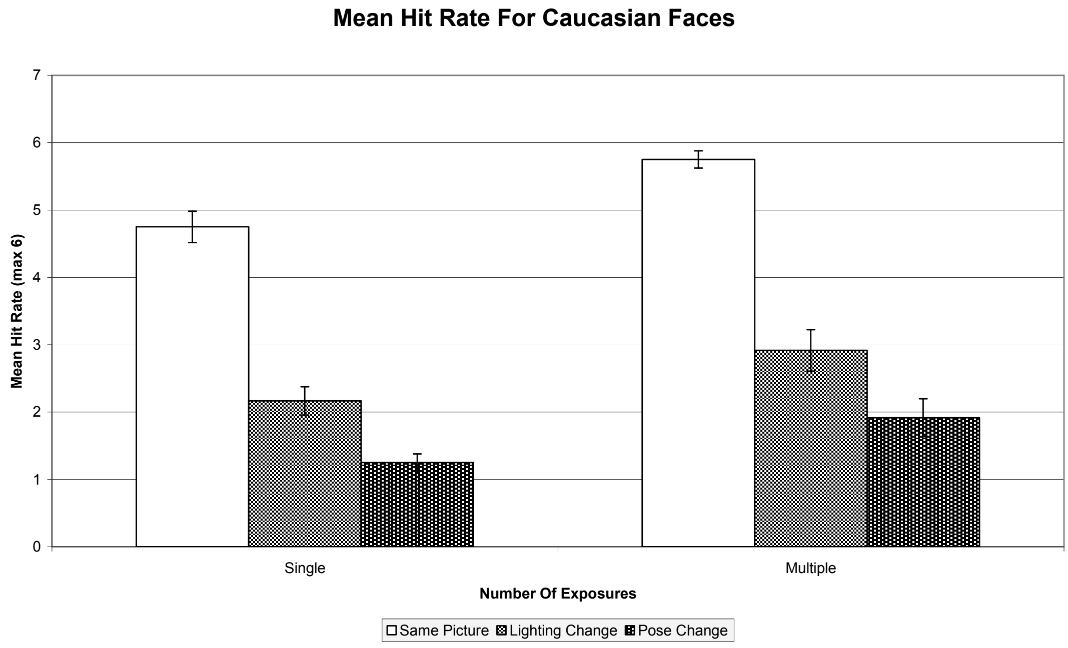


**Figure 2-4: Mean hit rate obtained in Experiment 1 for Caucasian faces in the three testing conditions (same picture, lighting change and pose change) after a single exposure or multiple exposures. Error bars represent standard error.**

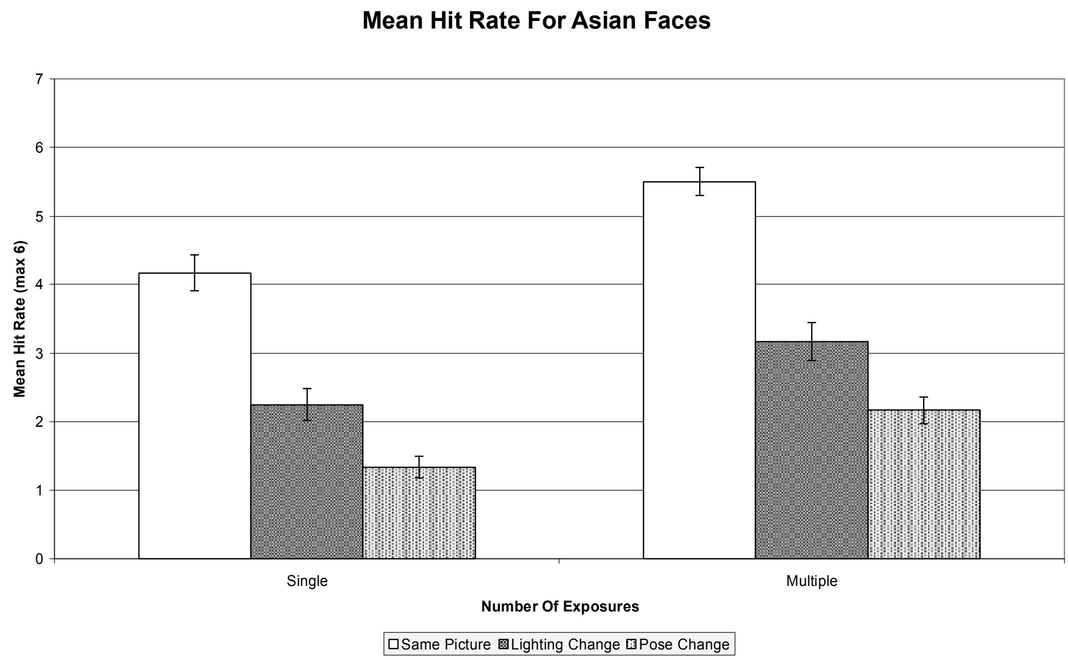


**Figure 2-5: Mean hit rate obtained in Experiment 1 for Asian faces in the three testing conditions (same picture, lighting change and pose change) after a single exposure or multiple exposures. Error bars represent standard error.**

The hit rate data were subject to a 2 (race of face – Caucasian or Asian) x 2 (single or multiple exposure) x 3 (type of transformation – same image, lighting change or pose change) mixed design ANOVA. As in the subjects analysis, there was a significant main effect of exposure type; $F(1,44) = 21.88$, $MSE = 1.38$, $p < .001$, $\eta_G^2 = 0.160$ (observed power = 1.00) and a significant main effect of transformation type; $F(2,88) = 129.94$, $MSE = 1.12$, $p < .001$, $\eta_G^2 = 0.646$ (observed power = 1.00). No reliable difference was found for face race type ($F < 1$, *ns*). None of the interactions were significant (all $F$ ratios < 1 except the interaction between race of face and type of transformation; $F(2,88) = 1.22$, $p > .05$).

### 2.2.3 Discussion

The results of Experiment 1 indicate that recognition performance after a pose change or after a lighting change is significantly worse than recognition of the same image used during training, with this factor yielding a large effect size in both the hit rate analysis and $A'$ analysis. The lack of any interaction reveals that this fall in accuracy was comparable across both single and multiple exposure conditions. Hence, despite receiving a large number of exposures to the face photographs and being very good at remembering them, participants in the multiple exposure condition were unable to generalise any better from the learned image to a novel view. This shows that an important influence on participants' performance involved something they were learning about the picture rather than a more generalised representation of the face.

There are two possibilities as to what is being learnt from the picture. Firstly, participants might be learning about the face in a specific viewpoint so that successful recognition of the face is possible as long as the viewpoint of the face remains unchanged between study and test. Alternatively, participants might be learning pictorial codes that are highly specific to the photograph studied (e.g. a mark on the image or a particular spot of illumination on the face). In this scenario, any change to the image would result in a drop in recognition accuracy. To separate these two possibilities, it is necessary to investigate recognition accuracy after a transformation which leaves the viewpoint unchanged between study and test, but alters the image properties. One transformation that does this is the transformation of lighting. For example, changing the lighting direction from full-face to another direction leaves the viewpoint of the face unchanged (i.e. both images depict a face in full-face view) but alters the image properties (e.g. one part of the face that was illuminated prior to the transformation may now be in darkness).

If performance is not affected by a lighting change (whilst keeping the viewpoint constant), recognition is likely to be more viewpoint-dependent than pictorially based as the viewpoint in both the study and test images is the same. On the other hand, if recognition is adversely affected by such a transformation then recognition is more likely to be pictorially based as, despite congruent viewpoints at study and test, the properties of the image have changed. In Experiment 1, recognition accuracy after a change in lighting fell significantly. Therefore, it appears that participants are learning something that is specific to the picture studied rather than viewpoint-specific information. In effect, despite being seen many times, the face has not become truly familiar in the sense that it can support pose or lighting-invariant recognition.

Interestingly, no evidence of the other-race effect was found; performance was equivalent to own race and other race faces. If participants were indeed learning about the picture rather than the structure of the face then this result might be expected. However, due to our interest in exploring what can optimally be learnt, it was decided to ensure that only faces from the same race (Caucasian) would be used in the remaining experiments.

## 2.3 Experiment 2 – Recognition of newly learnt faces across multiple pose changes

Experiment 1 successfully replicated previous work (Braje et al., 1996; Braje et al., 1998; Braje, 2003; Bruce, 1982; Krouse, 1981; Hill & Bruce, 1996) in demonstrating that after a single exposure to a picture of a face, recognition accuracy of that individual after a change in viewpoint was lower than recognition of the original image. These results suggest that recognition of unfamiliar faces learnt from photographs is pictorially dependent. However, an alternative explanation for the drop in recognition accuracy is that a viewpoint-invariant structural model is used for recognizing the face in novel views but the limited amount of experience with the face (e.g from a small number of exposures and only a single viewpoint) means that this structural model would be impoverished, preventing totally invariant recognition across all viewpoints.

On this account, despite the inaccuracies of an impoverished structural model, some level of recognition across pose changes would be possible. Any errors in the structural model would be consistent and present in all views, regardless of the angle through which the model was rotated. This would give rise to levels of recognition accuracy across different poses being similar regardless of the angle of rotation (e.g. it would not matter if the face was rotated 15°, 30°, 45° and so on, recognition accuracy would be approximately equal across all these changes). However, in this case, recognition accuracy of the originally studied picture will be better than that after a change in viewpoint because, as well as using this newly formed structural code, pictorial codes created from the picture used during training can also be employed for recognition, having the effect of boosting performance.

Recognition exclusively via the use of pictorial codes would yield a different result. In this case, recognition accuracy would decrease proportional to the angle of rotation through which the face is rotated. This is because the further the face is rotated, the greater the difference between the two images (study and test) becomes.

Therefore, the crucial difference between a structural account of face recognition and a pictorial account is the nature of recognition accuracy across differing degrees of change in pose. As Experiment 1 only employed a single pose change it does not provide an unequivocal indication of whether participants are learning a limited structural representation of the face as well as learning something about the photographs studied. The drop in accuracy could be due to either pictorial-dependency of recognition or to the creation of an impoverished three-dimensional structural representation, limited in terms of its generalization due to the limited amount of information from which the representation was formed.

Experiment 2 therefore examined further the roles played by structural and pictorial codes. Participants were again trained on faces presented from a single view. At test, five possible images were shown (one was the image originally studied, the other four were the same face shown in different angles of rotation). The pattern of results for the various degrees of rotation was analysed.

## 2.3.1 Method

### 2.3.1.1 Design

The experiment employed a within-subjects two-factor design with learning view and testing view as factors. Each factor contained five levels (0°, 17°, 31°, 44° and 62° of rotation from full-face). The dependent variable was the number of learned faces correctly identified for each of the five test images.

### 2.3.1.2 Participants

Fourteen participants (5 males and 9 females) from the University of York aged between 18 and 47 years were paid to take part in the study. All participants had normal or corrected to normal vision and none had participated in the previous experiment.

### 2.3.1.3 Materials

Images of 20 individuals from the PIE database, representing 0°, 17°, 31°, 44° and 62° angles of rotation away from the full-face view were used (see Figure 2-6). In all the images the lighting direction was kept constant (0°, full-face). All faces were male and Caucasian. The 20 individuals were split into two sets of 10 faces each.

**Figure 2-6: Examples of the stimuli used in Experiment 2 (from left to right, rotations from full-face: 0°, 17°, 31°, 44° and 62°).**

#### 2.3.1.4 Apparatus

The same apparatus as employed in Experiment 1 was used.

#### 2.3.1.5 Procedure

Participants were allocated one of the sets of faces to act as targets and the other as distractors, with this allocation counterbalanced across participants. Within the target set, faces were randomly assigned to one of the five views with the proviso that two faces represented each of the five views (to ensure participants learned two faces in each view). All participants completed three phases of the experiment; a first presentation phase, a training phase and a test phase. These phases were the same as in Experiment 1 with the following exceptions:

##### *2.3.1.5.1 Training phase*

During the first part of training, the 10 faces were broken down into two blocks of 5 faces each and the second part of training presented participants with all 10 faces.

##### *2.3.1.5.2 Test phase*

During the test phase, participants were presented with five blocks of 20 faces (the 10 targets and the 10 distractors). Participants were required to click "Yes" if they thought that the person presented was in the set of 10 that they learned or "No", if they thought they were not. Within each block, the five different views were depicted by four faces (two targets and two distractors) and across the five blocks participants saw each face in all five possible views.

###### *2.3.1.5.3 "Top-up" training*

Between each block, participants had to repeat the naming procedure presented during the final training phase for the 10 target faces. In this "top-up" training phase, the images were the same 10 as they had learned during the training phases. They were required to name all 10 without making an error, on a single occasion before moving onto the next test block. This "top-up" training was done because the distractor faces were presented multiple times. As all targets and distractors were shown an equal number of times during the test phase, participants may otherwise have become confused over which faces were members of the target set and which were members of the repeated distractors.

### 2.3.2 Results

#### 2.3.2.1 Effects of learning view and testing view on recognition

The mean percentages of correct responses for each of the five learned views when tested for recognition in each of the five possible testing views are presented in Figure 2-7.

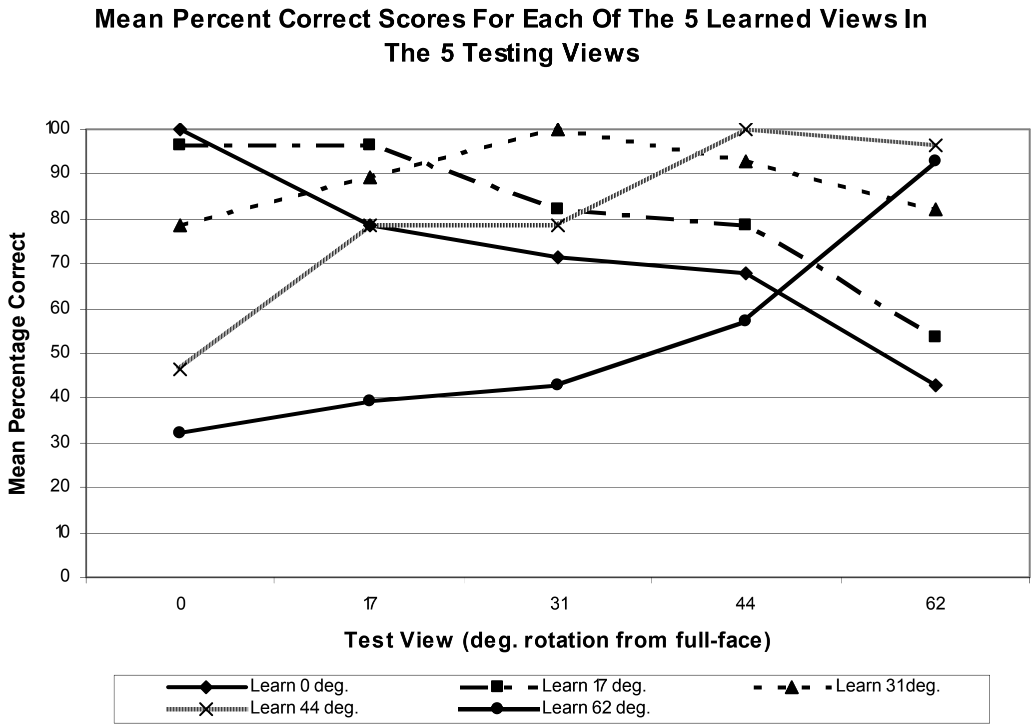


**Figure 2-7: Mean percentage correct scores in Experiment 2 for the five learned views tested in each of the five possible views.**

Since the learned faces could vary on two factors (learned view and test view) and the distractors on only one factor (test view), a signal detection measure such as *A*' was not appropriate. As participants learned two faces in each view during the training procedure, then at test for each learned view/test view cell they could score 0 (i.e. neither of the learned faces correctly identified), 1 (i.e. one face identified) or 2 (i.e. both the faces correctly identified). The ANOVA technique has been found to be robust despite departures from normality of the data (Norton, as cited in; Vician & Desanctis, 2000), even with very discrete data as was obtained in this experiment (Hsu & Feldt, 1969), and thus was used for analysis with the hit rates as the dependent variable.

The hit rate data were entered in a 5x5 within-subject ANOVA with learnt view (0°, 17°, 31°, 44° and 62°) and test view (0°, 17°, 31°, 44° and 62°) as factors. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There was a significant main effect of learned view; $F(4,52) = 10.06$, $MSE = 0.52$, $p < .001$, $\eta_G^2 = 0.088$ (observed power = 0.99), whilst the main effect of test view did not reach significance; $F(4,52) = 2.14$, $MSE = 0.13$, $p = .089$. The interaction between learned view and test view was significant; $F(16,208) = 11.44$, $MSE = 0.22$, $p < .01$, $\eta_G^2 = 0.155$ (observed power = 1.00).

The significant interaction was examined using a simple main effects analysis. An effect of test view was found for the 0°, 17°, 44° and 62° learning conditions; $F(4,52) = 10.53$, $MSE = 0.23$, $p < .001$, $F(4,52) = 7.56$, $MSE = 0.23$, $p < .001$, $F(4,52) = 11.39$, $MSE = 0.22$, $p < .001$, $F(4,52) = 18.98$, $MSE = 0.17$, $p < .001$ respectively. The effect of test view in the 31° learning condition only just failed to reach significance; $F(4,52) = 2.53$, $MSE = 0.16$, $p = .051$. Each of the learning conditions yielded a significant effect on the 0°, 17°, 31°, 44° and 62° test view conditions; $F(4,52) = 17.62$, $MSE = 0.29$, $p < .001$, $F(4,52) = 8.34$, $MSE = 0.33$, $p < .001$, $F(4,52) = 9.09$, $MSE = 0.27$, $p < .001$, $F(4,52) = 7.83$, $MSE = 0.22$, $p < .001$, $F(4,52) = 11.17$, $MSE = 0.29$, $p < .001$ respectively.

#### 2.3.2.2 Recognition as a function of rotation

From the results of the previous analysis, it appears that performance drops as a face is rotated further away from the view in which it was learned. To study this in more detail, it is necessary to examine recognition accuracy after increasing angles of rotation away from the learnt view. To do this, the difference between the angle of rotation in which a face was learnt and later tested was calculated for all faces. For example, consider a face that was learnt in the full-face view. Testing with the face presented in full-face has an angular change between the learnt and test view of 0°. When the face is tested in 17° rotation, the angular change is 17°. For a face learnt in 31° rotation and tested in 44° rotation there is an angular change between the learnt and test views of 13°. Table 2-1 details all the possible angular changes between study and test for the five learning views.

**Table 2-1: Matrix of the possible angular changes between study and test image in degrees. Positive angles indicate rotation towards to the profile. Negative angles indicate rotation toward the full-face.**

| Learn View (deg from full-face) | Test View (deg from full-face) | | | | |
|---|---|---|---|---|---|
| | **0** | **17** | **31** | **44** | **62** |
| **0** | 0 | 17 | 31 | 44 | 62 |
| **17** | -17 | 0 | 14 | 27 | 45 |
| **31** | -31 | -13 | 0 | 13 | 31 |
| **44** | -44 | -27 | -13 | 0 | 18 |
| **62** | -62 | -45 | -31 | -18 | 0 |

To simplify the analysis, similar angular changes were combined (e.g. the angular change of 17° is similar to the change of 18° and these two changes were collapsed to produce a single level of overall change). The angular changes that were combined were (negative angles indicate rotation towards full-face): -45°/-44°, -31°/-27°, -18°/-17°, -14°/-13°, 13°/14°, 17°/18°, 27°/31° and 44°/45°. The mean percentage correct score for these groups is given in Figure 2-8.

**Error! Objects cannot be created from editing field codes.**

**Figure 2-8: Mean percentage correct scores in Experiment 2 for differing angles of rotation away from the learned view. Where two angles were very close to each other (e.g. 17° and 18°) they have been collapsed together. Negative angles of rotation indicate rotation towards full-face. Positive angles indicate rotation towards the profile.**

To compare the effect of the direction of rotation (rotation from full-face to profile or vice-versa), the percentage data were entered into a two-way repeated-measures ANOVA with angle of rotation and direction of rotation as independent variables. There was a significant main effect of angle of rotation; $F(4,52) = 38.62$, $MSE = 0.18$, $p < .01$, $\eta_G^2 = 0.330$ (observed power = 1.00), but no main effect of direction of rotation; $F(1,13) = 3.51$, $MSE = 0.46$, $p = 0.84$ and no interaction ($F < 1$, *ns*). A linear trend was observed for the main effect of rotation; $F(1,13) = 71.00$, $p < .01$. This result indicates that recognition accuracy falls as the angle of rotation away from the learned view increases, and this holds for both rotation directions.

### 2.3.3 Discussion

The aim of this experiment was to extend the findings of Experiment 1 by investigating whether the fall in recognition accuracy observed after a pose change of a face learned from a single image was equal across different rotations of the face or was instead a function of the degree of rotation. The former would suggest that a structural model was being formed for the face that could then be used as the face was rotated through differing angles to new orientations. Imperfections in the model would be present in all new orientations resulting in similar levels of performance across all views. However the data clearly demonstrate that recognition is a function of the degree of rotation, supporting previous results with untextured faces (Hill et al., 1997).

An advantage for the three-quarter view was also found for both learning and testing views. The three-quarter view has received a lot of attention in the literature as it has been suggested that there is an advantage for this view over other views at both learning (Troje & Bulthoff, 1996) and at test (Baddeley & Woodhead, 1983; Krouse, 1981; Patterson & Baddeley, 1977; Siéroff, 2001; Woodhead et al., 1979). However, this advantage is not always found (e.g. Bruce et al., 1987) and in a recent review Liu and Chaudhuri (2002) suggest that the three-quarter view advantage may not even exist at all. Considering Figure 2-7, it would appear that in the present experiment this advantage is due to the three-quarter view having smaller possible angles through which it can be rotated, rather than anything unique about this view, as the generalization slopes follow similar decrements in accuracy for all learnt views across the same changes in angle at test.

## 2.4 Experiment 3 – The impact of learning two views of a face

Experiments 1 and 2 demonstrated that learning a face from a single photograph leads to learning something about the picture of the face over and above any face learning. The lack of invariant recognition across all views, and the fact that recognition is directly related to the degree of change between study and test images, suggests that not much three-dimensional structural information is being extracted during learning to be used at test.

It may be the case that providing only a single image of a face makes it difficult to create structural codes due to the limited amount of three-dimensional information that can be obtained from shape-from-shading extracted from one image. If this is so then learning two different images of the same face would provide more information about the face to aid the formation of such codes. If structural codes are indeed created for familiar faces, then learning two images would enable a more robust three-dimensional model to be formed in comparison to a learning single image. In consequence, we might expect to see invariant recognition from learning two views, at least across the difference between the two learnt views. If, on the other hand, recognition is mediated solely by the views learnt, with each view having its own generalization gradient, having access to two images would not boost recognition beyond that possible from the best of the two learnt views.

Experiment 3 examined this by training participants with faces presented in a single view (either 0° or 62°) or in two views (0° and 62°). Recognition accuracy was tested for three views; full-face (0°), near profile (62°) and a three-quarter view falling half way between the two (31°). If a three-dimensional structural model is created then both views will contribute to this model which will enable the model to be more accurate than if it is created from a single view. Testing on a previously unseen view will thus yield a higher level of accuracy than if only a single view had been learnt. In addition, a viewpoint-invariant three-dimensional structural model created from two views would lead to similar levels of accuracy across all rotations. In contrast, if all that is learnt are the different views themselves (i.e. if no structural code is formed) then there would be no benefit for learning two views of a face compared to the best of the single views when testing with a novel view of the face.

## 2.4.1 Method

### 2.4.1.1 Design

A 3x3 within-subjects factorial design was used with learning view (full-face, profile or both) and test view (full-face, three-quarter or profile) as factors. The dependent variable was the number of faces correctly recognised.

### 2.4.1.2 Participants

Twelve participants (8 females and 4 males) aged between 19 and 28 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected-to-normal vision and none had participated in the previous experiments.

### 2.4.1.3 Materials

Faces of 24 people from the PIE database were used. All were male and Caucasian with no additional distinguishing features such as beards or glasses. For these 24 individuals, three types of image were selected; full-face (0°), three-quarter (31°) and profile (62°) resulting in a total of 72 images. Lighting for all images was located at 0°. The images were prepared as in Experiment 1. The faces were split into two groups of 12 people.

### 2.4.1.4 Apparatus

The same apparatus as used in the previous experiments was employed.

#### 2.4.1.5 Procedure

The experiment consisted of the same three phases as the previous experiments. Each participant was allocated one set of 12 faces to act as targets and the other as distractors. The allocation of the two sets was counter-balanced across participants. Within the learning set, for each participant, four faces were randomly designated to be learnt from the full-face view only (yielding four images), four were randomly assigned to be learnt in profile only (yielding four images) and four randomly assigned to be learnt in both views (yielding 8 images) resulting in 12 *individuals* to be learned with a total of 16 *images*.

##### *2.4.1.5.1 First presentation phase*

Twelve images were used during the first presentation phase. Eight of these 12 images consisted of the four faces to be learned in full-face view and the four to be learned in profile. Of the final four images, depicting individuals to be learned in both views, two were selected to be presented in full-face and the other two were randomly selected to be presented in profile. Thus, only one view of each person was given during the first presentation. Each face was presented with its name underneath it for 5 seconds with ½ second between each face.

##### *2.4.1.5.2 Training phase*

In the training phase, all 16 images were used. For the first part of training, the 16 images were split into four blocks of four images each. Within each block the faces were shown in either full-face or profile, with eight faces in each view. Participants learned either the full-face views followed by profile views or vice-versa (counter-balanced across participants). As in the previous experiments, to progress through the blocks participants were required to name all four people in the block without making an error, on 3 separate occasions.

The second phase of training presented the participant with all 16 images. During this phase, only the 12 possible names were given as selection choices. Participants were required to name all 16 images without making an error, on three separate occasions, to complete the training task.

##### *2.4.1.5.3 Test phase*

The test phase involved three blocks consisting of faces presented in full-face view, three-quarter view or profile view. These blocks were counter-balanced across participants. Each block comprised 24 images, 12 of which were photographs of faces learned during the training phase and 12 distractors. Within each block, all faces were presented in the same view. Participants were instructed to press "Yes" if they thought the person presented was a member of the set of 12 they originally learned or "No" if they were not. There was no "top-up" stage in the testing phase as each distractor was only displayed 3 times which was not considered to be sufficiently high enough to warrant a top-up stage.

### 2.4.2 Results

#### 2.4.2.1 Recognition results

The mean percent correct scores for each of the three learning views in each of the three testing views are given in Figure 2-9. The critical comparison between the learning conditions is at the three-quarter test view, which was not seen at any time during training.

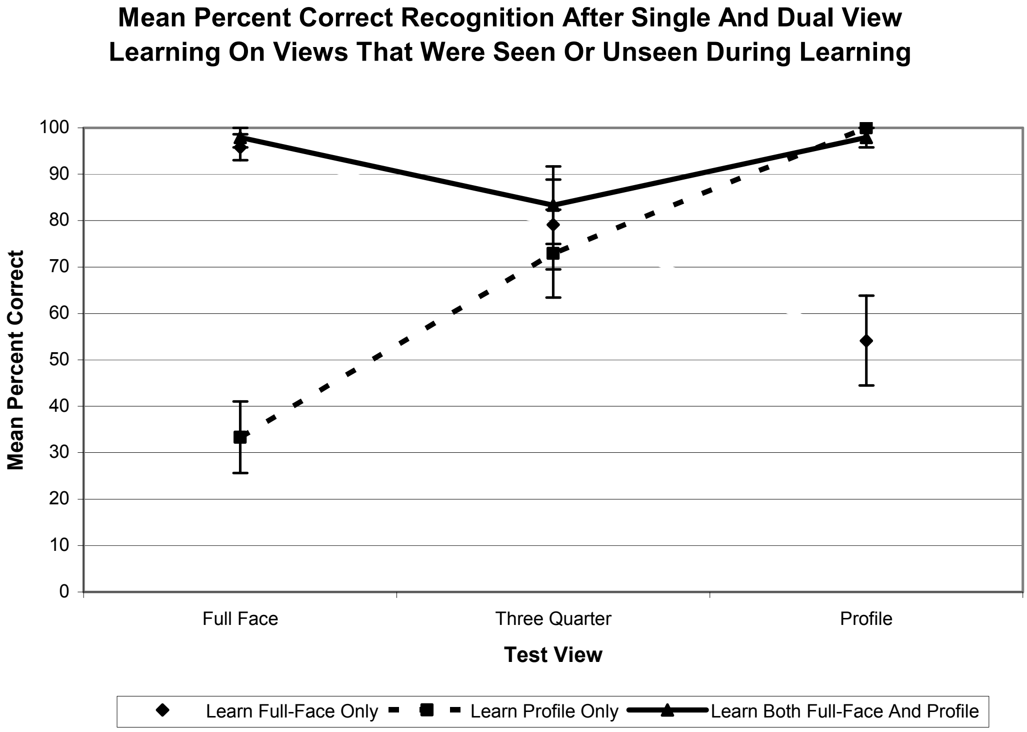


**Figure 2-9: Overall effect of learning two views or one view of a face on recognition for the three test views; full-face, three-quarter and profile in Experiment 3. Lines represent mean percent correct. The three-quarter test view was not seen during either the first presentation or training phases. Error bars represent standard error.**

A 3x3 repeated-measures ANOVA was conducted with learned view and test view as independent variables and number of hits as the dependent variable. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). A significant main effect of learned view was found; $F(2,22) = 23.86$, $MSE = 0.37$, $p < .01$, $\eta_G^2 = 0.041$ (observed power = 1.00). The main effect of test view was not significant ($F(1.61,17.65) = 1.01$, $p = .368$). There was a significant interaction between the two factors; $F(4,44) = 37.17$, $MSE = 0.41$, $p < .01$, $\eta_G^2 = 0.129$ (observed power = 1.00)

It was expected that the faces trained in two views would necessarily show better overall performance since they would achieve high recognition rates for both the full-face and profile views (i.e. the two trained views) whereas the only high score for the faces learned in full-face or profile would be for the view in which the face was learnt. Planned contrasts revealed this to be the case, with higher scores achieved when faces were learned in two views; $F(1,11) = 47.69$, $MSE = 2.03$, $p < .01$. Full-face only and profile only learning did not differ significantly from each other but approached significance; $F(1,11) = 4.12$, $p = .067$, indicating that the full-face only learning produced higher recognition accuracy than faces learnt from the profile view only. This result is to be expected as it is known that generalization from the profile to other views is poor compared to the generalization obtained with other views.

To analyse the interaction further, a simple main effects analysis was conducted. It was expected that the full-face and profile test views would yield significant higher performance due to good accuracy when learning and testing views matched, but poor accuracy when the opposite views were used. This was indeed found for both the full-face view test and the profile view test; $F(2,22) = 74.75$, $MSE = 0.35$, $p < .01$ and $F(2,22) = 21.86$, $MSE = 0.59$, $p < .01$ respectively.

There was however, no significant effect of learned view on testing in the important three-quarter view comparison; $F(2,22) = 2.07$, $p = .15$. This test view is critical because it was never seen during training and thus was a novel view during the test phase. It therefore allows for a direct comparison between the learning conditions.

As there was no statistical difference between the learning conditions at this point, this indicates that performance on the unseen three-quarter view is not significantly better whether a single image of a face is learned or two images are learned.

#### 2.4.2.2 Learning data analysis

The recognition test results indicate that participants received no benefit from learning two images of the same person when asked to recognise them from a novel viewpoint. To examine whether this result might be due to the number of times a face was displayed during the training sessions, the number of exposures received for single and multiple images was analysed. The mean number of exposures to each view for both single and dual view learning are shown in Figure 2-10 below.

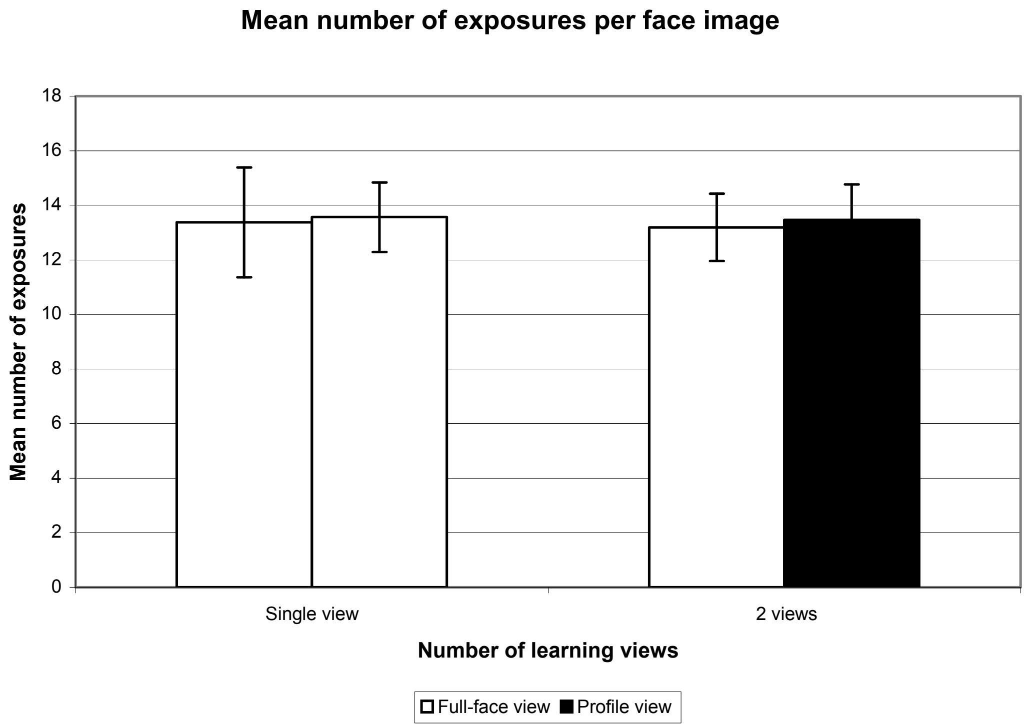


**Figure 2-10: Mean number of exposures given per image in single and dual view learning conditions for both full-face and profile views in Experiment 3. In the single-view learning condition, the individual depicted in the photograph was seen in that view only. In the two-view learning condition, two different photographs depicted the same individual. Therefore, whilst**

**the number of exposures is comparable for full-face and profile learning, the individuals in the two-view learning condition will have been seen approximately twice as many times the individuals in the single-view learning condition. Error bars represent the standard errors.**

A two-way ANOVA with number of views (one or two views) and viewpoint (full-face or profile) as factors was conducted. This analysis revealed no main effect of number of views learned, no main effect of viewpoint learned and no interaction between the two main effects (all $F$ ratios < 1, *ns*). This result indicates that participants needed just as many exposures to the faces presented in two views as they did for faces presented in one view. Thus, the result obtained from the recognition data for the novel view is not due to the faces being learned from two views being learned with fewer exposures than faces learned from a single view. The data presented in Figure 2-10 suggest that learning a face from two views requires as many exposures to *each* viewpoint during learning as does learning the view from that view only. In essence, the faces of the *individuals* presented via two-view learning had to be shown approximately twice as many times in training as those learned from a single image.

### 2.4.3 Discussion

The results of this experiment indicate that performance on recognizing a face from a previously unseen three-quarter view was not significantly improved by learning two images of the face over learning a single image. In this experiment, learning two images of a face did not help participants to generalise to a novel view any more effectively than they could generalise after learning a single image. This result suggests that, during the learning of a face, invariant structural codes from the two views are not being combined to produce a three-dimensional structural model of the face. Instead, it appears that participants were learning properties of the images.

The experiment also allowed for an examination of the number of exposures required to learn a face. If information from the two images presented in the two-view learning condition was being successfully combined, then it would be expected that fewer exposures would be required to learn the face. The results indicate that, in contrast, just as many exposures were needed for each of the views in the two-view learning condition as were required to learn that view as a single view. Effectively, participants were again learning individual pictures, rather than individual faces. Overall, this experiment provides further evidence that in face learning experiments using pictures of faces as stimuli, participants readily learn properties of the photographs used rather than invariant features of the face.

## 2.5 Experiment 4 – Learning different names to different views of a face (I)

The results from Experiment 3 indicate that when people learn a face from two photographs, generalization to a novel viewpoint is no better than that obtained from a single view and that learning two images of a face did not lead to an overall reduction in the number of exposures required during the learning process. This suggests that learning faces from photographs is view-specific and that the two images are not being combined to form a single structural model of the face.

If, as it seems, participants are learning about a photograph over and above any invariant structural information, then training with two images (e.g. a full-face view and a profile view) of the same person with different names assigned to the two views (e.g. the full-face view is a person called "Robert" and the profile view of the same face is called “David”) should prove to be no more difficult to learn than learning the two views with the same name given to both images (e.g. both the full-face view and profile view are given the name “Robert”; see Figure 2-11). This is because the two views can be learnt as individual images with no single structural model (and by implication, no single identity) being created from the two views. As such, the two mental representations share no information and no confusion will arise as to the identity of the individuals.

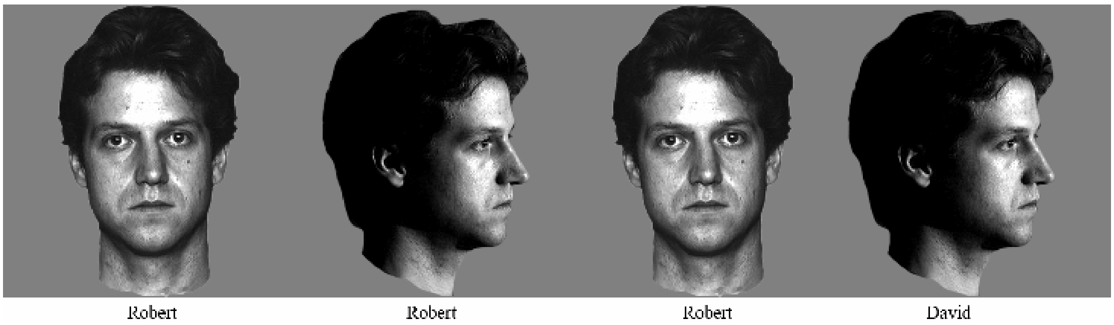


**Figure 2-11: Examples of the name and face learning conditions used in Experiment 4. In the two view same name learning condition, both the full-face and profile views are assigned the same name (in the two faces on the left, both views are assigned the name Robert). In the two view different name learning condition, the two views are assigned different names (depicted in the two faces on the right in which the full-face view is assigned the name Robert and the profile view the name David).**

On the other hand, successful generation of a three-dimensional model combining structural information from the two views would lead to confusion as to the name of the individual because, under this hypothesis, the two views both belong to the same structural model. Whilst learning different names to different views would still be possible, it would be hindered by this confusion (as the two images apparently depict different people yet the structural model is a representation of a single person).

Experiment 4 therefore extends Experiment 3 by adding a further condition to examine the effect of learning different names to different views of the same face. It involved training participants with four types of learning; full-face view only, profile view only, full-face and profile views with both views assigned the same name (as in Experiment 3), and full-face and profile view with the two views assigned different names.

## 2.5.1 Method

### 2.5.1.1 Design

A 4x3 within subjects factorial design was used with learning type (full-face view only, profile view only, full-face and profile views both with the same name, and full-face and profile views with different names for each view) and test view (full-face, three-quarter view, and profile) as factors. The data collected consisted of the number of correct identifications made during the test phase and the number of exposures required to learn the faces during the training procedure.

### 2.5.1.2 Participants

Twelve participants (3 males and 9 females) aged between 18 and 21 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected-to-normal vision and none had participated in any of the previous experiments.

### 2.5.1.3 Materials

The same 24 faces in the same two groups of 12 as used in Experiment 3 were utilized.

#### 2.5.1.4 Apparatus

The same apparatus as used in the previous experiments was employed.

#### 2.5.1.5 Procedure

Each participant was allocated one of the sets of 12 faces to act as targets, and the other set acted as distractors. The allocation of the two sets was counter-balanced across participants. Within the learning set faces were randomly assigned to one of the four conditions with an equal number of faces in each condition; three faces to be learned from the full-face view only (yielding 3 images), three faces to be learned from the profile view only (yielding 3 images), three to be learned from the full-face and profile views with the same name assigned to each image (yielding 6 images) and three to be learned from the full-face and profile views with different names assigned to each image (yielding 6 images). This resulted in the 12 people being learned from a total of 18 images. The participants were not told, at any stage of the experiment, that the same individual might have two different names assigned to the full-face and profile views.

##### *2.5.1.5.1 First presentation phase*

Twelve of the 18 images were shown during the first presentation phase. All 12 images depicted a different person. These 12 images comprised the three faces to be learned from the full-face view only and the three to be learned from the profile only. The remaining six faces were drawn from the set of faces to be learned from two views (three from the set to be learned with the same name and three to be learned with different names) such that three faces were shown full-face and three were shown in profile. As three faces were to be learned with two views with a different name assigned to each view, a total of 15 names were required. As participants only saw 12 images during the first presentation phase, they only saw 12 of the 15 names during this part.

*2.5.1.5.2 Training phase*

In the training phase, all 18 images were used. For the first part of training, the 18 images were split into six blocks of three faces each. Three of the blocks consisted of exclusively full-face views and the other three exclusively of profile views. Participants either learned full-face views followed by profile views or vice-versa (counter-balanced across participants). As in the previous experiments, to progress through the blocks participants were required to name all people in the block without making an error, on three separate occasions.

The second phase of training presented participants with all 18 images. During this phase, all 15 possible names were presented as selection choices. Participants were required to name all 18 images without making an error on three separate occasions to complete the training task.

*2.5.1.5.3 Testing phase*

The testing phase was identical to the testing phase used in Experiment 3.

## 2.5.2 Results

### 2.5.2.1 Recognition test results

The mean percentage correct scores for each of the learning types on the recognition test across the three test views are given in Figure 2-12.

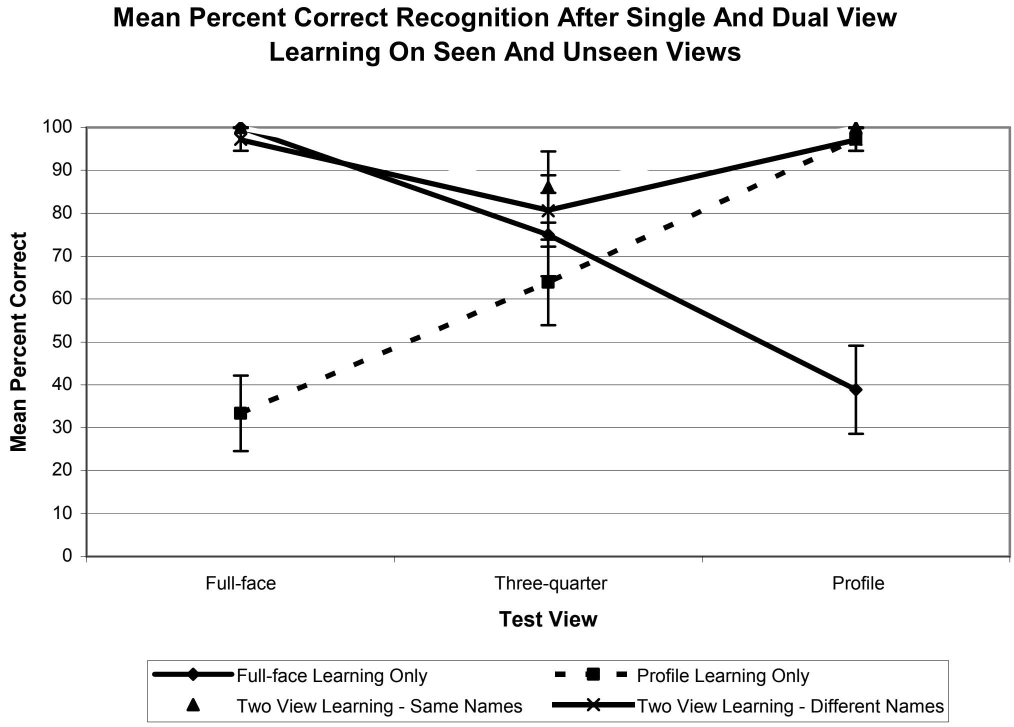


**Figure 2-12: Mean percentage correct scores in Experiment 4 for four types of learning (full-face only, profile only, two views with the same name and two views with different names) for each of the three test views (full-face, three-quarter and profile). Lines represent mean percent correct and error bars represent standard error. Note that the three-quarter test view was not shown during the training procedure.**

The raw scores were entered into a 4x3 repeated measures ANOVA with learning type (full-face only, profile only, two views with the same name, and two views with different names) and test view (full-face, three-quarter, and profile) as factors. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There was a significant main effect of learning type; $F(2.95,32.43) = 19.02$, $MSE = 0.39$, $p < .01$, $\eta_G^2 = 0.065$ (observed power = 1.00) but no main effect of test view ($F$

< 1, *ns*). There was a significant interaction between the two factors; $F(4.34,47.23) = 27.86$, $MSE = 0.37$, $p < .01$, $\eta_G^2 = 0.123$ (observed power = 1.00).

Just as in Experiment 3, it was expected that the faces trained in two views would show better overall performance since they would achieve high recognition rates for both the full-face and profile views (i.e. the two trained views) whereas the only high score for the faces learned in full-face or profile would be for the view in which the face was learnt. This was tested using planned contrasts. It was found that learning two views with each view having the same name or different names led to significantly greater accuracy than learning only a single view; $F(1,11) = 40.24$, $MSE = 2.09$, $p < .01$. Separating out the two view learning conditions, learning two views with the same name did not lead to better performance than learning two views with different names; $F(1,11) = 1.29$, $p = .281$. When learning from a single view, learning the full-face did not produce better recognition than learning the profile view; $F(1,11) = 2.20$, $p = .166$.

To analyse the strong interaction, a simple main effects analysis was conducted. There was a significant effect of learning type on the full-face test; $F(3,33) = 50.17$, $p < .01$ and on the profile test; $F(3,33) = 29.86$, $p < .01$. These results again are to be expected, as performance when a single view was learned in full-face and tested in profile or vice-versa was poor. In the other conditions (single view learning and testing with the same view or in either of the two view learning condition) where the test picture has been seen before, accuracy was good.

As in Experiment 3, the critical test concerns the data for the three-quarter view condition. Again there was no significant effect of learning type on testing with a three-quarter view; $F(3,33) = 2.66$, $p = .065$ indicating that learning two views does not produce better recognition results than learning a single view does. Importantly, the result of this experiment reveals that performance levels obtained when faces were learned in two views with the same name given to each view was no different from the faces learned in two views with different names, suggesting that participants were learning the faces as individual pictures.

#### 2.5.2.2 Number of exposures needed to learn the faces

Unlike Experiment 3, it is difficult to examine the learning data for Experiment 4 in detail. This is because in Experiment 3 all the names were presented to the participant together with an image of the face during the first presentation phase. As such, the participant was not exposed to a new name during the training session. In Experiment 4, however, participants did not see three of the names during the first presentation phase and these names had to be learnt during the training session, potentially artificially boosting the number of exposures required to learn the faces presented in two views with different names. The mean number of exposures required to learn the faces is shown in Table 2-2.

**Table 2-2: Mean number of exposures required to learn the faces in the three types of learning condition for both full-face and profile views in Experiment 4.**

| Learning Condition | Single View | | Dual View/ Same Name | | Dual View/ Different Names | |
|---|---|---|---|---|---|---|
| View | Full-face | Profile | Full-face | Profile | Full-face | Profile |
| Mean (*SD*) | 13.86 (*2.94*) | 14.25 (*2.8*) | 14.61 (*4.04*) | 14.61 (*3.51*) | 15.86 (*3.86*) | 15.14 (*3.59*) |

As can be seen in Table 2-2, participants needed a similar number of exposures to learn the faces in all three conditions, with the most exposures required for the dual view/different name learning condition. Indeed, participants require about 1 extra exposure to learn the faces in this condition. This is hardly surprising, as unlike the other two conditions, there were some names that they had never seen before. It would appear then that there is little cost involved in learning two faces from different views in which both views have different names assigned to them.

The results of Experiment 4 confirmed the result from Experiment 3 indicating that learning two views of the same face does not require significantly fewer exposures for learning the face in comparison to learning a single view of a face. For the reasons given above, though, this only held true when the two views were given the same name. In this case, despite the fact that the participant is being told that these two images depict the same person, they still in effect learn the images separately.

### 2.5.3 Discussion

The results of the recognition test used in Experiment 4 confirm the result from Experiment 3. Both experiments show that learning two images of faces did not lead to better recognition performance on a novel view than learning a single image. Importantly, whilst recognizing a previously unseen three-quarter view was more error prone than either of the learnt views, the lack a significant difference between learning conditions on the novel ¾ test view suggests that that this drop off in performance at the three-quarter view is consistent whether a single view, two views with both views having the same name assigned to them, or two views with different names assigned to each view is learnt. Therefore, overall, recognition is not statistically better after learning a single view than after learning two views.

The pattern of results regarding the number of exposures to learn the faces from a single image or from two images with the same name was repeated with a similar number of exposures required in both cases. Additionally, it appears that learning two views of a face with different names does not have a large effect on the number of exposures required to learn the images, suggesting that participants were not confusing two different names between two images of the same person. Experiment 4 therefore provides further evidence that what is being learnt from photographs of faces is properties of the image, not the face per se.

## 2.6 Experiment 5 – Learning different names to different views of a face (II)

Experiment 4 indicates that learning two views of a face does not lead to better recognition on a previously unseen view than does learning a single view. However, it is possible that the small number of participants used in Experiment 4 ($N$ = 12) might have reduced the power available to this experiment. To examine this, Experiment 5 repeated Experiment 4 with two changes to address this problem. First, the number of participants was increased. Secondly, response latencies were also recorded and analysed as latencies are a more sensitive measure than accuracy when performance is close to ceiling.

### 2.6.1 Method

#### 2.6.1.1 Design

A 4x3 within subjects factorial design was used with learning type (full-face view only, profile view only, full-face and profile views both with the same name, and full-face and profile views with different names for each view) and test view (full-face, three-quarter view, and profile) as factors. The data collected consisted of the number of correct identifications made during the test phase, the latencies to make these identifications, and the number of exposures required to learn the faces during the training procedure.

#### 2.6.1.2 Participants

Twenty-four participants (4 males and 20 females) aged between 18 and 33 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected-to-normal vision and none had participated in any of the previous experiments.

#### 2.6.1.3 Materials

The same 24 faces in the same two groups of 12 as used in Experiments 3 and 4 were utilized.

#### 2.6.1.4 Apparatus

The same apparatus as used in the previous experiments was employed.

#### 2.6.1.5 Procedure

The procedure was identical to that used in Experiment 4 except that in the final testing phase participants were instructed to use the keyboard to make their responses instead of the mouse. Participants had to press the left control (CTRL) key if they thought they had learnt the name of the individual presented or the right control key if they had not. They were also instructed, as is the norm in reaction time experiments, to answer as quickly and accurately as they could. Response latencies were recorded, as well as correct and incorrect choices.

## 2.6.2 Results

### 2.6.2.1 Accuracy data

The mean percentage correct scores for each of the learning types on the recognition test across the three test views are given in Figure 2-13.

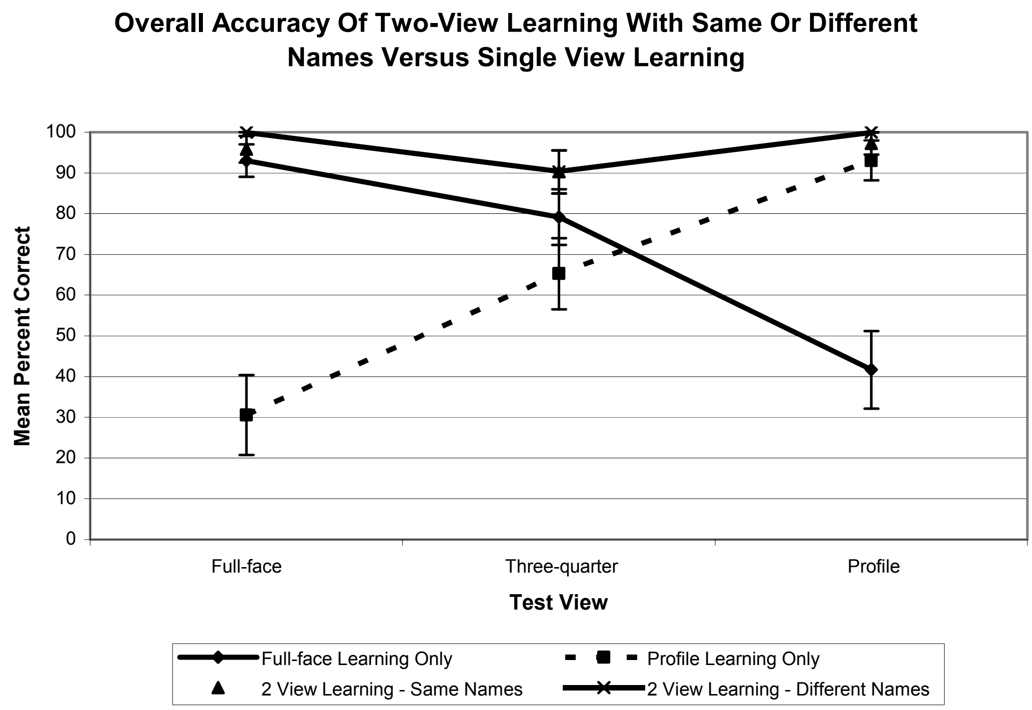


**Figure 2-13: Mean percentage correct in Experiment 5 scores for four types of learning (full-face only, profile only, two views with the same name and two views with different names) for each of the three test views (full-face, three-quarter and profile). Lines represent mean percent correct and error bars represent standard error. Note that the three-quarter test view was not shown during the training procedure.**

The raw scores were entered into a 4x3 repeated measures ANOVA with learning type (full-face only, profile only, two views with the same name, and two views with different names) and test view (full-face, three-quarter and profile) as factors. The

Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There was a significant main effect of learning type; $F(2.65,60.96) = 36.68$, $MSE = 0.57$, $p < .01$, $\eta_G^2 = 0.018$ (observed power = 1.00) but no main effect of test view ($F < 1$, *ns*). There was a significant interaction between the two factors; $F(4.38,100.69) = 42.34$, $MSE = 0.40$, $p < .01$, $\eta_G^2 = 0.226$ (observed power = 1.00).

The significant effect of training type was examined by planned contrasts. It was found that learning two views with each view having the same name or different names led to significantly greater accuracy than learning only a single view; $F(1,23) = 90.26$, $MSE = 6.98$, $p < .01$. Learning two views with the same name did not lead to better performance than learning two views with different names; $F(1,23) = 3.09$, $p = .092$. When learning from a single view, learning the full-face did not produce better recognition than learning the profile view ($F < 1$, *ns*).

The significant interaction was analysed using simple main effects. As was expected, the poor performance on faces learnt in profile and tested in full-face led to a significant effect of learning type within the full-face test view; $F(3,69) = 69.16$, $MSE = 0.34$, $p < .01$. A similar result was found for faces learnt in full-face view and tested in profile; $F(3,69) = 58.46$, $MSE = 0.28$, $p < .01$. A difference was also found for the four learning conditions at the three-quarter test view; $F(3,69) = 6.58$, $MSE = 0.46$, $p < .01$. To analyse this unexpected result, unplanned contrasts were performed on these data points. Significance was determined using the Scheffé criterion. It was found that faces learnt from the profile view only were recognised significantly less well than faces learnt from the other conditions; $F(1,23) = 10.36$, $MSE = 8.51$, $p < .05$. However, faces learnt from the full-face view only were recognised as well as faces learnt from two views; $F(1,23) = 4.80$, $MSE = 2.23$, $p = .093$. Faces learnt from two views with the same name and those learnt from two views with different names did not differ ($F < 1$, *ns*). Within the learning conditions, an expected significant effect of test view was found for face learnt from the full-face view only; $F(2,46) = 37.34$, $MSE = 0.41$, $p < .01$ and the profile view only; $F(2,46) = 35.25$, $MSE = 0.60$, $p < .01$. A significant effect of test view was also found for faces learnt with two views/different names; $F(2,46) = 6.75$, $MSE = 0.10$, $p < 0.005$. No effect of test view was found for faces learnt from two views with the same name ($F(2,46) = 2.21$, $MSE = 0.13$, $p = .357$).

#### 2.6.2.2 Latency data

Recognition accuracy after learning the full-face view only and testing on the profile view and vice-versa is poor. Only four of the 24 participants made a correct response to these pose changes, making an analysis of the response latencies to correct responses difficult. To analyse the latency data, it was therefore necessary to combine the full-face only and profile only learning conditions into one, single view learning condition, in which response times for the full-face view testing condition were taken from when the full-face view was learnt and response times for the profile view testing condition were taken from when the profile view was learnt (i.e. the views that were learnt during training). To determine the response times to be used for the unseen three-quarter view testing condition, the data from the learning condition which yielded the highest accuracy on this test view were used. From

Figure 2-13, it can be seen that the highest recognition accuracy for the three-quarter test view after learning either the full-face view only or profile view only is obtained from learning the full-face view. Therefore, response times in this condition were used as the single view learning, three-quarter test view condition. The mean reaction times to correct responses in each of the three test views for two view learning conditions plus the new single view learning condition are shown in Figure 2-14.

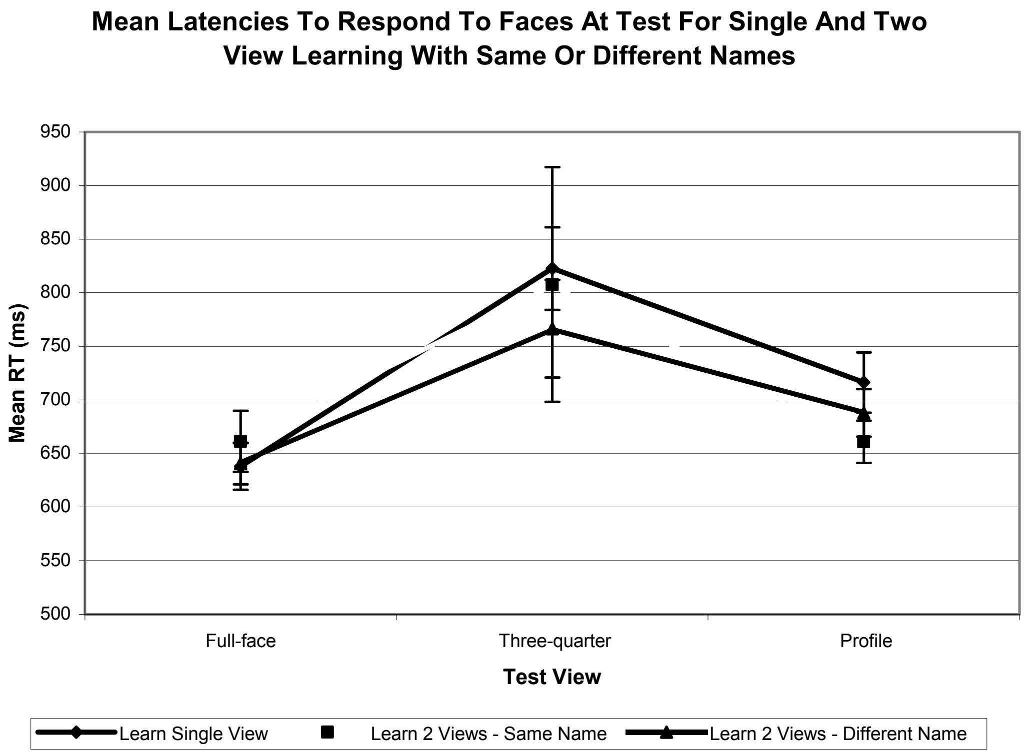


**Figure 2-14: Median response times in Experiment 4 for the best performance to the three test views (full-face, three-quarter and profile) after learning faces from a single view (learn full-face/test profile, learn full-face. test three-quarter, learn profile/test profile) or two views (for both same name and different name conditions). Lines represent median latency and error bars represent standard error. Note that the three-quarter test view was not shown during the training procedure.**

The data were entered in a 3x3 repeated measures ANOVA with learning condition (single view only, two view/same name and two view/different names) and test view (full-face, three-quarter and profile) as independent variables. The dependent variable was the mean reaction time. There was a significant main effect of test view; $F(1.25,28.64) = 7.50$, $MSE = 97951.23$, $p < .01$, $\eta_G^2 = 0.026$ (observed power = 0.81). Neither the main effect of learning condition nor the interaction were significant (both $F$ ratios < 1, *ns*). Planned contrasts revealed that latencies to the three-quarter view were significantly longer than to either the full-face or profile views; $F(1,23) = 7.79$, $MSE = 1916270.63$, $p < .05$. It was also found that reaction times to the full-face view were faster than those to the profile view; $F(1,23) = 5.03$, $MSE = 74549.98$, $p < .05$.

#### 2.6.2.3 Number of exposures needed to learn the faces

The mean number of exposures required to learn each picture of a face was calculated and is shown in Table 2-3

**Table 2-3: Mean number of exposures required to learn the faces in the three types of learning condition for both full-face and profile views in Experiment 5.**

| Learning Condition | Single View | | Dual View/ Same Name | | Dual View/ Different Names | |
|---|---|---|---|---|---|---|
| View | Full-face | Profile | Full-face | Profile | Full-face | Profile |
| Mean (*SD*) | 13.85 (3.40) | 14.04 (3.2) | 14.26 (3.73) | 13.99 (3.22) | 15.07 (*3.27*) | 15.51 (*3.64*) |

From Table 2-3 it can be seen that participants need about 1 extra exposure to learn the faces from two views when each view was given a different name. Again, as in Experiment 4, it would appear then that there is little cost involved in learning two faces from different views in which both views have different names assigned to them.

The results of this experiment confirmed the results from Experiment 4 indicating that learning two views of the same face does not require significantly fewer exposures for learning the face in comparison to learning a single view of a face.

### 2.6.3 Discussion

Experiment 5 was designed to build upon Experiment 4 to increase sensitivity by increasing the power of the experiment and by examining response latencies. With increased power, the results of Experiment 5 are similar to those of Experiment 4. Recognition accuracy on the previously unseen three-quarter view was no better if two views are learnt in comparison to learning the full-face view only. It is interesting to note however that on this test, recognition accuracy after learning the profile only was significantly poorer than when either two views or the full-face view only had been learnt, consistent with previous research which has demonstrated that generalization from the profile is poor compared to other views (e.g. Hill et al., 1997). This suggests that even after learning two views, participants may be predominantly relying on the full-face view for the recognition process on the unseen three-quarter view. However, in addition, the lower recognition accuracy on the three-quarter test view after learning the full-face only compared to learning two views of the face approached significance, suggesting that the profile view may have a contribution to make to the recognition process when the full-face view is learnt.

The response latency data revealed that the mean reaction time to faces presented in the previously unseen three-quarter test view were significantly longer than to either the full-face or profile test views (i.e. the studied views). The result was consistent regardless of whether a single view or two views were learnt indicating that learning two views of a face does not make the task of identifying the face in a novel view any faster than when a single view is learnt.

## 2.7 General discussion of Experiments 1 to 5

The experiments presented examined the effect of multiple exposures and multiple views on face learning. Experiment 1 tested the effects of providing multiple exposures to a face in comparison to a single exposure. Recognition accuracy of unfamiliar faces from a single exposure produced a similar result to that found by previous research (Bruce, 1982; Krouse, 1981) in that a change in the pose of the face led to a significant fall in recognition accuracy. In addition, performance was significantly affected by a change in the direction of illumination, yielding a similar result to that found in matching studies in which the illumination direction of the face is changed between the two images (Braje et al., 1996; Braje et al., 1998; Braje, 2003; Hill & Bruce, 1996; Liu et al., 1999).

Bruce (1982) suggested that recognition in such situations is mediated by a pictorial code in which the face is represented by the image presented rather than any invariant structural information. Recognition will depend upon how well the image at test can be matched onto the image presented during the study phase and, in essence, unfamiliar face recognition is therefore image-based.

Exactly the same pattern of lack of invariance across changes in pose or lighting was found at test when the studied picture had become highly familiar. The lack of an interaction between number of exposures and transformation type (same image, viewpoint or lighting change) indicates that the effect of applying a transformation to the face between study and test is the same whether it is learnt from a single or from multiple exposures. Hence, even when multiple exposures of an image have been provided, invariant information about the face has still not been extracted and recognition is based upon image-specific or view-specific codes.

Examining the effect of changing the lighting direction from which the face is illuminated between study and test provides an indication as to whether an image-specific or a view-specific code is the primary code for recognition. Changing the lighting direction leaves the view of the face itself unchanged. If view-specific codes were used for recognition then changing the lighting would have no effect on recognition. On the other hand, if image-specific codes were used then changing the lighting direction between study and test would be expected to have a detrimental effect on recognition. This is because changing the lighting direction changes the image, resulting in different images being used at study and at test. Experiment 1 found that changing the lighting direction of a face significantly lower recognition accuracy suggesting that recognition is more likely to me mediated by image-specific pictorial codes than by viewpoint codes.

Previous research has suggested that recognition accuracy after a single exposure to a single image of a face is a function of angular rotation. Experiment 2 extended this to examine whether all pose transformations of a face learned in a single view from multiple exposures lead to the same level of recognition, or if recognition accuracy remains a function of the rotation of the face. Results showed that recognition accuracy worsened the further the face was rotated away from the learned view. This decrease was evenly graded rather than leading to approximately the same level of performance across the differing views, suggesting that recognition is not mediated by a structural model that would yield similar recognition rates across the majority of views. Again, it would appear that what is remembered from learning a picture is the original image, and attempts to generalise to novel views are based upon how well the presented image can be matched to the stored one. As a face rotates away from the learned view, less information is shared between the two images, resulting in the pattern observed. These results should not be taken as an indication that generalization from a single image is not possible. Clearly some level of generalization is possible, but this generalization is graded and is far from perfect.

If face learning from pictures is indeed largely picture-specific, then learning two images of a face would lead to high levels of recognition accuracy for the two learned images but accuracy for a third novel view would still yield lower recognition rates. Additionally, if this third view fell a sufficient distance away from the learned views, such a hypothesis predicts that recognition accuracy on the third view will be no better than when only one of the views is learned. Experiment 3 examined this by training participants with either a single image or two images of a face. Learning two images of a face did not produce higher levels of recognition of a novel test view than learning only one of the images. This result suggests that the two views were not being combined effectively to aid recognition of a novel view. Importantly, this lower level of accuracy on the third view was not simply because participants were able to learn the faces presented in two views any more easily (leading to less exposures) than they were able to learn faces from a single view. On the contrary, faces learned in two views required just as many exposures to <u>each</u> of the views as when those views were learned alone as a single image. Hence, learning two views of the faces led to the individual depicted having to be presented approximately twice as many times as individuals learned from a single view. Despite this larger number of exposures to the face, generalization to a novel view was no better than learning a single view.

As well as looking for effects of generalization at test, Experiments 4 and 5 extended Experiment 3 by examining generalization during the learning phase as well as the test phase. At test, recognition accuracy was no better for faces learnt from two views in comparison to those learnt from a single view. Indeed, learning two views of a face with the same name given to both views (to encourage generalization) proved to be no more useful for recognition than learning two views of a face with different names assigned to each view (to interfere with generalization). Further to this, during learning itself faces that were learnt from two views *with different names assigned to each view* were learnt from approximately the same number of exposures as faces learnt from two views *with the same name given to each view*, indicating no confusion arising from the two views having different names. The results of Experiments 3, 4 and 5 thus suggest that participants are learning each view of the face as a separate image and are not generalizing between two views (when provided) to aid their learning of the faces.

These results suggest that recognition of unfamiliar and familiarised (during the experiment) faces is image-based (i.e. based upon pictorial codes) or viewpoint-dependent (based upon viewer-centered codes). Of these alternatives, pictorial codes are the preferred explanation because of the effects of changes in lighting. Even the matching of two pictures of the same person taken from the same viewpoint is difficult when the lighting differs (Bruce et al., 1999). The images used by Bruce et al. were created on the same day, from the same viewpoint. The primary differences were that the images were created at different times of day and also taken with a different camera. Similarly, the work of Braje and colleagues (Braje et al., 1998) together with the results of Experiment 1 suggests that the face recognition system encodes illumination effects within specific viewpoints, further suggesting that the pictorial code proposed for recognition is more likely to be image-specific as opposed to view-specific. Finally, textural information has been shown to be important for recognition when objects are structurally similar (Price & Humphreys, 1989). Conversely, Bruce and Langton (1994) found that participants found it very hard to recognise laser-scanned heads when no texture information was applied. These cases again highlight the importance of pictorial information for recognition.

It appears that recognition of both unfamiliar and thoroughly learnt photographs of faces is heavily influenced by image-based codes. The experiments presented here do not support the hypothesis that the learning of faces from photographs takes place via the construction of a three-dimensional structural model of the face that allows for invariant recognition across multiple poses. If such a model was created, recognition accuracy across various pose transformations would be similar yet, as seen in Experiments 2, 3, 4 and 5, recognition accuracy shows a graded slope across such transformations. Instead, in order for a face to become truly familiar it seems that exposure to a sufficient number of different views is required. Totally invariant recognition across pose is not possible after exposure to a single view. The accuracy of recognition of a face in a novel view depends upon how similar the novel image is to one of the previously stored images. If the match is a close one, then accurate recognition is possible. Figure 2-7 shows that, generally, the best performance on a novel view occurs when the test view is near to the learnt view. As the difference between the novel image and the nearest stored image increases, accuracy drops in a similar fashion to that found in previous research on object recognition (Bulthoff & Edelman, 1992; Tarr & Pinker, 1989). The results of the experiments presented here are consistent with theories of object recognition that propose multiple two-dimensional instances as the mechanism for recognition. In particular, in Bülthoff and Edelman's (1992) feature-based, multiple instance model, features within the view are extracted and allow for small variations in viewpoint changes (Wallis & Bulthoff, 2002).

In conclusion, the experiments reported in this Chapter suggest that learning faces from photographs of faces leads to the learning of the properties of the photograph over and above any invariant face information. In order for a face to be recognised from a novel image (created by a transformation of pose) the recognition system has to generalise between the stored and presented images. This generalization is however, open to error, especially if the two images are sufficiently different. From this perspective, successful face recognition across changes in the image created by transformations such as lighting and pose would be best achieved by making the differences between the stored and presented image as small as possible. One way to achieve this would be to store many different views of the face as separate images so that, upon presentation of a novel image, one of the stored images is a close match. It seems that faces become familiar not only because we have seen them a number of times, but also because we have seem them from a number of views.

# CHAPTER THREE

## Viewpoint-dependent and pictorial coding in face recognition

## 3.1 Introduction

Experiments 1 to 5 presented in Chapter 2 demonstrated that a transformation of lighting or pose applied between study and test, significantly decreases recognition accuracy of faces learnt from photographs compared to the performance obtained when the same image is used at study and test. Three possibilities as to why this drop in performance occurs was discussed. Firstly, recognition of faces could be mediated by a single viewpoint-invariant three-dimensional structural model. Secondly it is possible that recognition employs the use of viewpoint-dependent codes and thirdly, pictorial codes, specific to the image studied during learning might be used.

Experiments 1 to 5 did not find any support for the use of a viewpoint-invariant three-dimensional structural model in face recognition. Instead, image-based codes (either viewpoint-dependent or pictorially-dependent) seem to be heavily involved in the recognition of faces learnt from photographs.

The results of Experiment 1 indicated that changing the lighting direction of a face between study and test significantly decreases recognition accuracy, suggesting that recognition is mediated by pictorial, rather than viewpoint dependent codes. However, applying only one lighting change (a change of 30°) and comparing recognition accuracy to that obtained with testing with the same image does not allow for a thorough investigation of the roles played by pictorial and viewpoint-dependent codes. This is because when testing takes place with the same image that was used during learning, both pictorial and viewpoint-dependent codes may be used for recognition, leading to good levels of performance. The performance obtained after a lighting change (when only viewpoint-dependent codes would be available) might be inferior simply due to the fact that only one type of code is available to the recognition process.

Testing with different amounts of lighting change will allow for a closer inspection of the roles played by pictorial and viewpoint-dependent codes in face recognition, as performance after varying amounts of lighting change can be compared. If accuracy decreases as the amount of lighting change between the study and test images increases, then it is likely that pictorial codes are primarily used in the recognition process. This is because as the angular change in lighting between the study and test images increases, so the difference between the two images increases, reducing pictorial similarity between them. On the other hand, similar levels of performance across different amounts of lighting change would suggest that viewpoint-dependent codes are being used for recognition as, despite the differences in image properties, the viewpoint is consistent at study and test.

It is possible that changing the lighting direction of a face between study and test may hide features of the face that were used during learning, which in turn leads to poorer recognition accuracy at test. It can be argued then that recognition after a lighting change might still use viewpoint-dependent codes, as it may be that features contained within the view during learning, are no longer available at test for the recognition process, giving rise to what is essentially a new viewpoint. Therefore, a clearer demonstration of the relative roles of viewpoint-dependent and pictorial codes would be via the use of a transformation that alters the image, whilst leaving the viewpoint constant, and also keeps the same features visible in both the study and test images.

One transformation that fulfils this requirement is the transformation of size. Altering the size of a face between study and test changes the image but leaves the viewpoint unchanged and the features of a face used during learning visible. Previous research has found that unfamiliar faces do not exhibit size invariance (Kolers et al., 1985) but that familiar faces do (Brooks, Rosielle, & Cooper, 2002). This result suggests that faces are recognised via pictorial codes, as unfamiliar faces have usually been seen from a limited number of views (and in many cases only a single view) and a change in size leads to sufficient changes in the image that the study and test images are sufficiently different enough to make the recognition task difficult. Familiar faces on the other hand, have been seen from many different views (including different sizes). Therefore, the difference between the (previously created) stored facial representation and the face to be identified is smaller. However, whilst familiar and unfamiliar face recognition after the transformation of size has been studied, little is known about the recognition of newly learnt faces after a change in size. Accordingly, Experiment 7 was designed to investigate the recognition of newly learnt faces after a size transformation.

## 3.2 Experiment 6 – Face recognition after transformations of lighting and pose

Experiment 6 further investigated the role played by pictorial and viewpoint-dependent codes in face recognition by testing recognition accuracy with different amounts of lighting change. In addition, if recognition is mediated by pictorial codes, applying two transformations together (which will increase the difference between the study image and the test image to a greater extent than if either of the transformations were applied separately), will lead to poorer performance than if either of the two transformations were applied separately. To investigate this, recognition was also tested after the transformations of pose and lighting were applied separately, or together.

### 3.2.1 Method

#### 3.2.1.1 Design

The experiment took the same form as Experiments 1 – 5. During the test phase, all participants were tested on their recognition accuracy for each face in its original form (i.e. same image), after a change in lighting, a change in pose, or a combination of both transformations. There were three possible levels of each transformation, therefore the experiment had a 3x3 within subjects design with pose change (0° change, 31° change and 62° change) and lighting change (0° change, 30° change and

62° change) as independent variables. The dependent variable was the number of faces that were correctly identified during the test phase.

### 3.2.1.2 Participants

Sixteen undergraduate students (8 male and 8 female) aged between 18 and 34 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision.

### 3.2.1.3 Materials

Images of 18 faces (all Caucasian and male) from the PIE face database were used. None of the images used depicted the individual with facial hair or wearing glasses. Each face was used in three lighting conditions and three poses, resulting in a total of 162 images. In the three lighting conditions, the light source was located either directly in front of the model (0°), 30° to their right or 62° to their right. The images for the three poses were also taken from directly in front of the model (0°) or to their right (31° and 62°). Figure 3-1 shows an example of a set of images for one face.

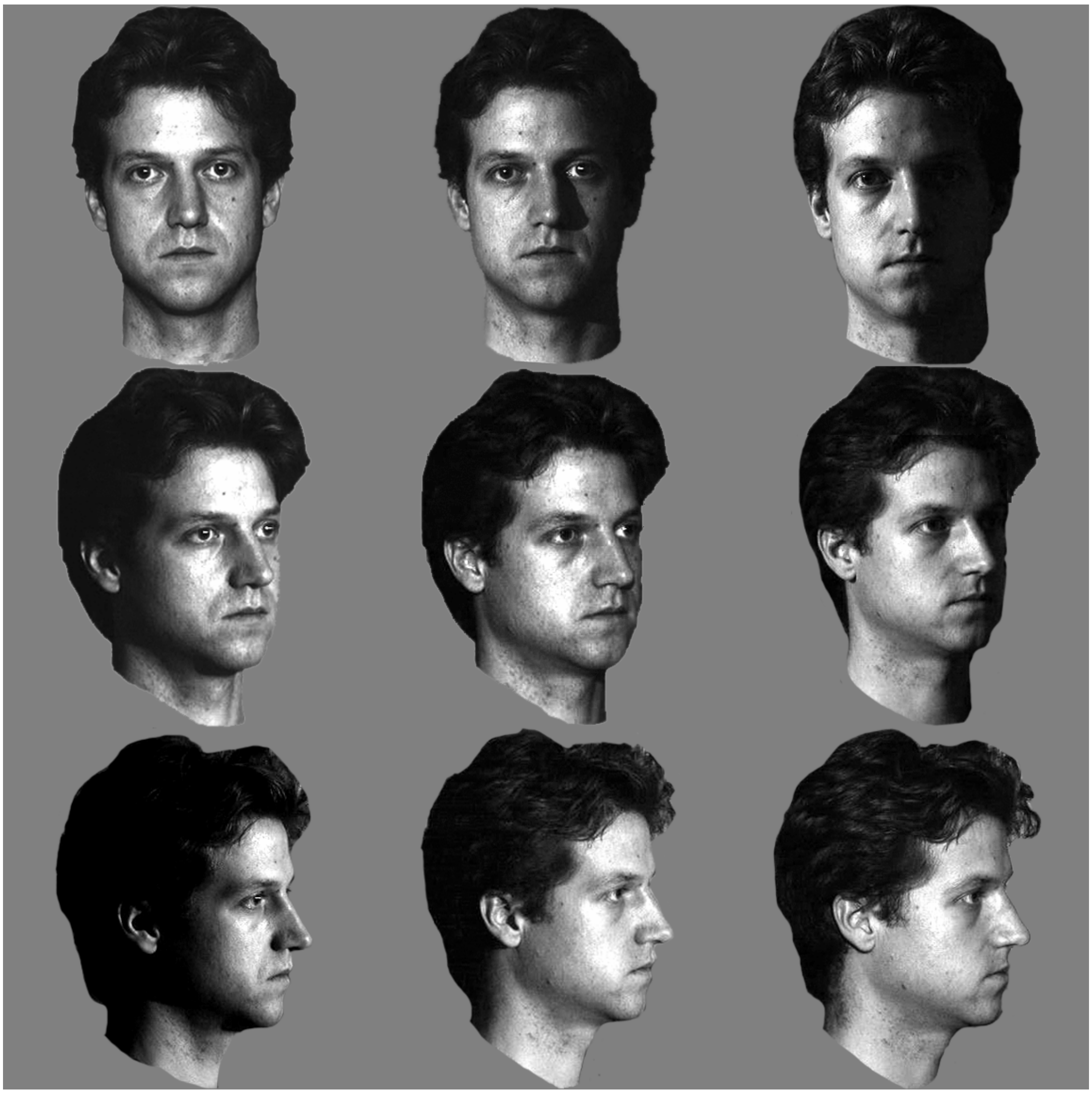

**Figure 3-1: Examples of the nine types of images used in Experiment 6. The top row depicts faces in full-face view (0°), the middle row shows faces in three-quarter view (31°) and the bottom row shows faces in near-profile view (62°). The leftmost column shows faces illuminated from the full-face (0°), the centre column shows faces illuminated from 30° to the model's right and third column shows faces illuminate from the right of the model (62°).**

Each image was manipulated to remove all irrelevant background information, leaving only the head visible. The background was replaced with a homogenous grey. The original colour images were converted to greyscale. Each image was resized so that it was 384 pixels high in order to normalize face height and the background expanded to create a final image of 384x384 pixels, yielding a visual angle of 4.87° when viewed from a distance of 60cm. The 18 faces were placed at random into two sets of nine faces before the experiment began and the faces remained in these groups for all participants.

#### 3.2.1.4 Apparatus

The apparatus was the same as in previous experiments.

#### 3.2.1.5 Procedure

For half the participants, the first set of faces was allocated as the target set and set two as the distractor set. For the other half, this order was reversed. All participants completed three phases; a first presentation phase, a training phase and a test phase. The participant sat in front of the computer screen at a distance of approximately 60cms and was given written instructions before the experiment began.

##### *3.2.1.5.1 First presentation phase*

During the first presentation phase, participants saw nine faces for a duration of 5 seconds with 0.5 seconds between each face. The faces were evenly distributed across the three views (0°, 31° and 62°) and three lighting conditions (0°, 30° and 62°) so that two faces in each pose/lighting combination were presented. Each individual face was presented to the participant once and was accompanied by a name, presented below the image of the face. These name/face pairings were randomly generated for each participant from a set of nine fixed names.

##### *3.2.1.5.2 Training phase*

The training phase took the same form as in the previous experiments. In the first part of training, the nine face photographs were divided into three blocks containing three faces each. In the second part, all nine photographs were presented in a single block. The criterion for completing the training phase was the same as in the previous experiments.

*3.2.1.5.3 Test phase*

During the testing phase participants were presented with 162 faces and they were required to make a "yes/no" decision as to whether they had seen the presented face during the training phase. Each target face was presented in all nine possible lighting and pose combinations (in terms of pose change/lighting change; 0°/0°, 31°/0°, 62°/0°, 0°/30°, 0°/62°, 31°/30°, 31°/62°, 62°/30° and 62°/62°) thus yielding recognition accuracy for the same picture (the 0°/0° condition) and eight types of change. To offset this large number of presentations of the target faces, each distractor was also presented nine times in the same pose and lighting combinations. Participants were warned before the experiment began that the both the target faces and the distractor faces would appear a number of times during the test phase

### 3.2.2 Results

The mean percentage of faces correctly identified in the test phase for the nine possible lighting change and pose change combinations are shown in Figure 3-2.

Mean Percent Recognition Accuracy For Each Of The 9 Test Images In Experiment 6

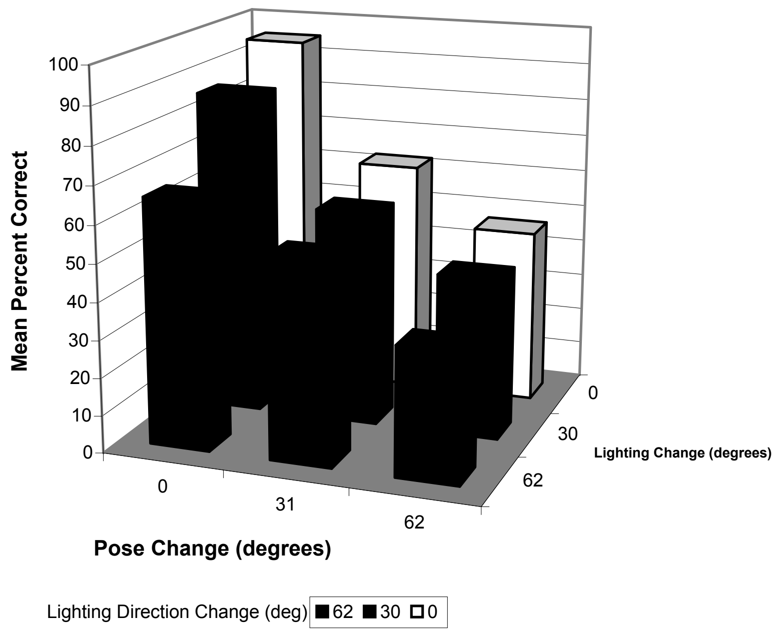


**Figure 3-2: Mean percentage recognition accuracy obtained in Experiment 6 for the nine possible transformations. The 0° pose change/0° lighting change condition represents accuracy when the same image was used at study and at test. The lighitng and pose changes refer to the *amount* of change between the learnt view and the test view.**

Each pose/lighting combination had a different total number of faces to be identified (for example, the 0° pose change and 0° lighting change combination yielded a hit rate out of 9 whilst 0° pose change and 30° lighting change yields a hit rate out of 12). Therefore, the mean proportion correct for each pose change and lighting change combination was used for analysis. Since mean proportion correct data are bounded between the values of 0 and 1, the data were transformed using the arcsine transformation prior to analysis.

The arcsine-transformed data were entered into a 3x3 within subject ANOVA with pose change (0°, 31° and 62°) and lighting change (0°, 30° and 62°) as factors. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There was a significant main effects of pose change; $F(1.68,25.19) = 59.1$, $MSE = 0.07$, $p < .001$, $\eta_G^2 = 0.389$ (observed power = 1.00) and lighting change; $F(1.63,24.45) = 14.08$, $MSE = 0.06$, $p < .001$, $\eta_G^2 = 0.107$ (observed power = 0.99). There was also a significant interaction between the pose and lighting changes; $F(4,60) = 4.41$, $MSE = 0.03$, $p < .005$, $\eta_G^2 = 0.046$ (observed power = 0.92).

It was expected that as the angular change in pose between the study and test images increased, recognition accuracy would decrease. Planned contrasts revealed this to be so with a pose change of 31° or 62° significantly reducing recognition over a 0° change; $F(1,15) = 112.64$, $MSE = 1.02$, $p < .001$ and a 62° change in pose had a greater decrement on recognition than a 31° change did; $F(1,15) = 13.23$, $MSE = 5.26$, $p < .005$. The same pattern of results was found for the lighting change. Changing the lighting (by either 30° or 62°) yielded significantly lower accuracy than keeping the lighting constant; $F(1,15) = 26.95$, $MSE = 0.58$, $p < .001$. Furthermore, a 62° change in lighting had a greater effect than a 30° change; $F(1,15) = 7.64$, $MSE = 0.39$, $p < .025$.

### 3.2.3 Discussion

The results of Experiment 6 confirmed the those of Experiment 1 in showing that changing either the pose (producing a large effect size) or lighting (producing a medium effect size) of a face between study and test significantly decreased recognition accuracy. Furthermore, a linear trend was also found for both pose and lighting transformations such that the further a face or it direction of lighting was rotated away from the view in which it was learnt, the lower recognition accuracy became. It was also found that changing both the lighting and the pose of the face together led to poorer performance than applying only one of the transformations on its own.

The results indicate that recognition of faces learnt from a single view appears to be tied closely to the image learnt, such that recognition is greatest for the image learnt and falls the more that the image differs from the learnt view. As in Experiments 2 to 5, Experiment 6 demonstrated that the further a face is rotated away from the learnt view, the lower recognition accuracy becomes, suggesting that recognition is mediated via viewpoint-dependent or pictorial codes. Additionally, the transformation of lighting helps determine whether viewpoint-dependent or pictorial codes are used for recognition as a change in lighting leaves the viewpoint unchanged whilst altering the image. Performance after a change in lighting was significantly lower than that obtained by testing with the learnt image, suggesting that participants were learning something specific about the picture, rather than more general viewpoint-specific information.

In Experiment 1, it was possible that the superior performance found when testing recognition accuracy of the studied picture in comparison to the same face after a transformation of lighting was due to the fact that when testing takes place with the originally studied image, both viewpoint-dependent and pictorial codes can be used for the recognition process. In contrast after a change in lighting, viewpoint-dependent codes will be predominately used as the image has changed, reducing the use of any pictorial codes. If this was indeed the case, different lighting changes should lead to similar levels of recognition. However, Experiment 6 revealed that performance after a 62° lighting change was significantly poorer than after a 30° lighting change. If viewpoint-dependent codes were used then recognition levels would be similar for both changes. Instead, it was found that the greater the change in lighting, the poorer performance becomes. This suggests that the recognition of faces is mediated by pictorial codes as the viewpoint of the face after both 30° and 62° lighting change is the same, yet the difference between the study and test images increases the further the greater the lighting change is.

Furthermore, accuracy was found to be poorer after both lighting and pose transformations were applied to the face than when a single transformation was made. This is consistent with the proposal put forward by Bruce (1982) that it is the *amount* of change, not the *type* of change that is important for recognition. It therefore appears that recognition is being mediated by pictorial codes over and above viewpoint-dependent or viewpoint-invariant codes.

## 3.3 Experiment 7 – Recognition of faces after a transformation of size

It has been shown that unfamiliar faces do not show size invariance. A change in image size between study and test significantly decreases recognition accuracy (Kolers et al., 1985), suggesting that unfamiliar face recognition is pictorially dependent. It is possible that information in the face required for size invariant recognition cannot be extracted from the single exposure used by Kolers et al.. Therefore, Experiment 7 examined recognition accuracy for faces learnt very well from a single photograph after a transformation of size. In addition, performance after a change in pose was also examined. This was done to provide a point of contrast for recognition accuracy obtained after the size change.

## 3.3.1 Method

### 3.3.1.1 Design

The experiment took the exactly the same form as Experiment 6, except the lighting change of Experiment 6 was replaced by a size change. The size change was measured in terms of a scaling factor which was calculated by comparing the study and test images. If both images were the same size, this resulted in a scaling factor of 0. If the test image was twice or four times the size of the learnt image, this yielded scaling factors of +2 and +4 respectively. If the test image was smaller than the learnt image (either a half or a quarter size), this yielded a scaling factor of –2 or –4. The experiment therefore had a 3x5 within subjects design with pose change (0°, 31° and 62° change) and size change (a scale factor of –4, –2, 0, +2 and +4). The dependent variable was how many faces were correctly identified during the test phase.

### 3.3.1.2 Participants

Sixteen undergraduate students (4 male and 12 female) aged between 18 and 20 from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision. None had participated in any of the previous experiments.

#### 3.3.1.3 Materials

Images of 18 faces (all Caucasian and male) from the PIE face database were used. None of the images depicted the individual with facial hair or wearing glasses. Each face was used in three poses and three sizes, resulting in a total of 162 images. The images for the three poses were taken from directly in front of the model (0°) or to their right (31° and 62°). Each image was manipulated to remove all irrelevant background information, to leave only the head visible. The background was replaced with a homogenous grey. The original colour images were converted to greyscale. Each image was resized so that it was 384 pixels high in order to normalize face height and the background expanded to create a final image of 384x384 pixels, yielding a visual angle of 4.87° when viewed from a distance of 60cm. Two more variants of each image were made. The linear dimensions of these images were 50% and 25% of the original image respectively (192x192 pixels, 1.95° visual angle and 96x96 pixels, 0.97° visual angle). Figure 3-3 shows an example of these images for one face.

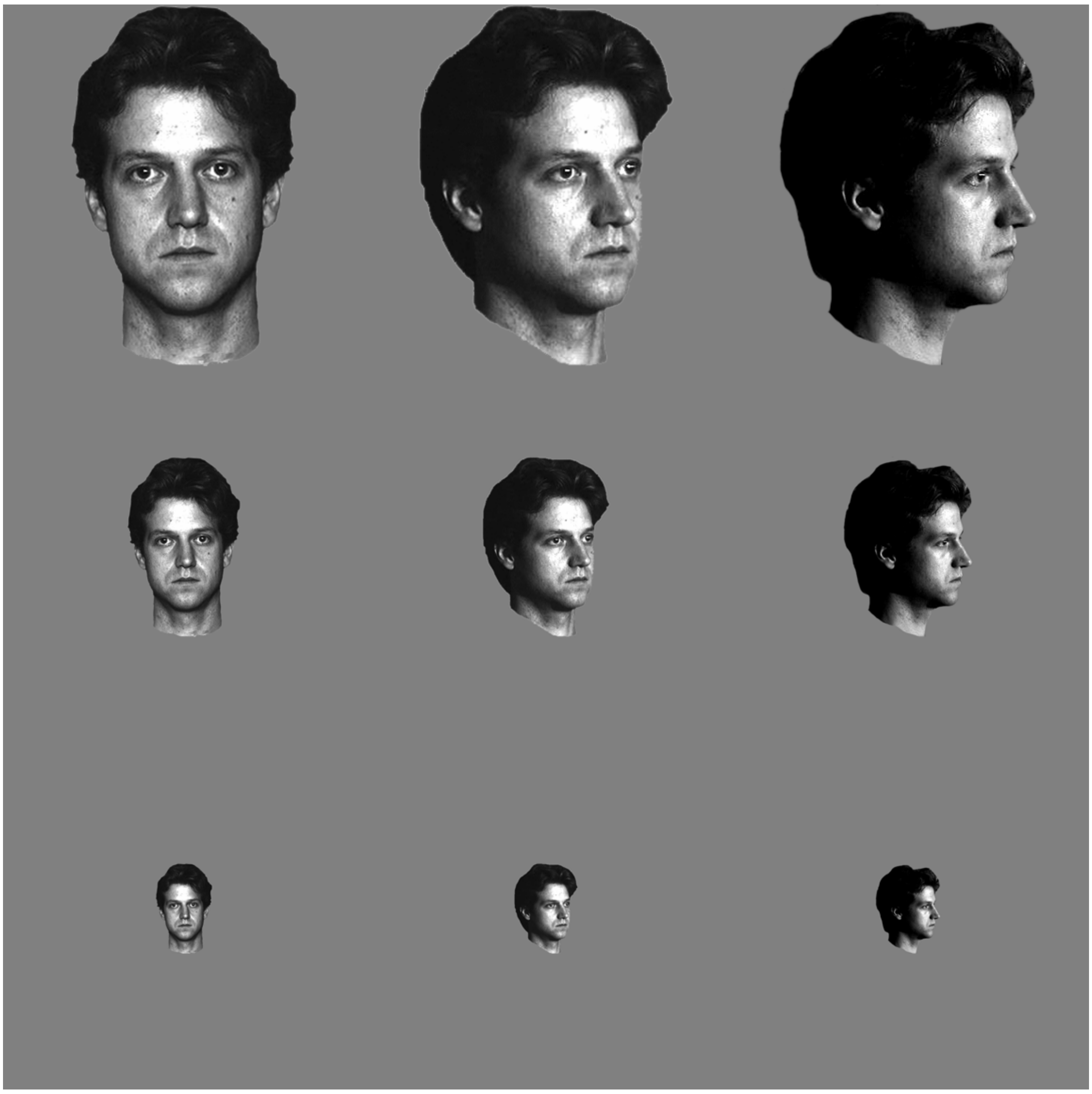

**Figure 3-3: Examples of the face stimuli used in Experiment 7. The faces in the top row were 384x384 pixels and subtended a viewing angle of 4.87°. The faces in the middle row were 192x192 (1.95°) pixels and the bottom row faces were 96x96 pixels (0.97°).**

The 18 faces were randomly placed into two sets of nine faces each before the experiment began and the faces remained in these groups for the all participants.

#### 3.3.1.4 Apparatus

The same apparatus as used in the previous experiments was employed.

#### 3.3.1.5 Procedure

For half the participants, the first set of faces was allocated as the target set and set two as the distractor set. For the other half, this order was reversed. All participants completed three phases; a first presentation phase, a training phase and a test phase. The participant sat in front of the computer screen at a distance of approximately 60cms and was given written instructions before the experiment began.

##### *3.3.1.5.1 First presentation phase*

During the first presentation phase, participants saw 12 faces for a duration of 5 seconds each with 0.5 seconds between each face. The faces were evenly distributed across the three views (0°, 31° and 62°) and three sizes so that one face in each pose/size combination was presented. Each individual face was presented to the participant once and was accompanied by a name, presented below the image of the face. These name/face pairings were randomly generated for each participant from a set of nine fixed names.

##### *3.3.1.5.2 Training phase and test phase*

These phases were the same as in Experiment 6.

### 3.3.2 Results

The mean recognition accuracy in percent for each of the pose and size changes is shown in Figure 3-4. Size change is represented by a scaling factor. A factor of –4 means the test image was ¼ the size of the learnt image and a factor of –2 means the test image was ½ the size of the learnt image. Positive scaling factors indicate the face had increased in size between study and test.

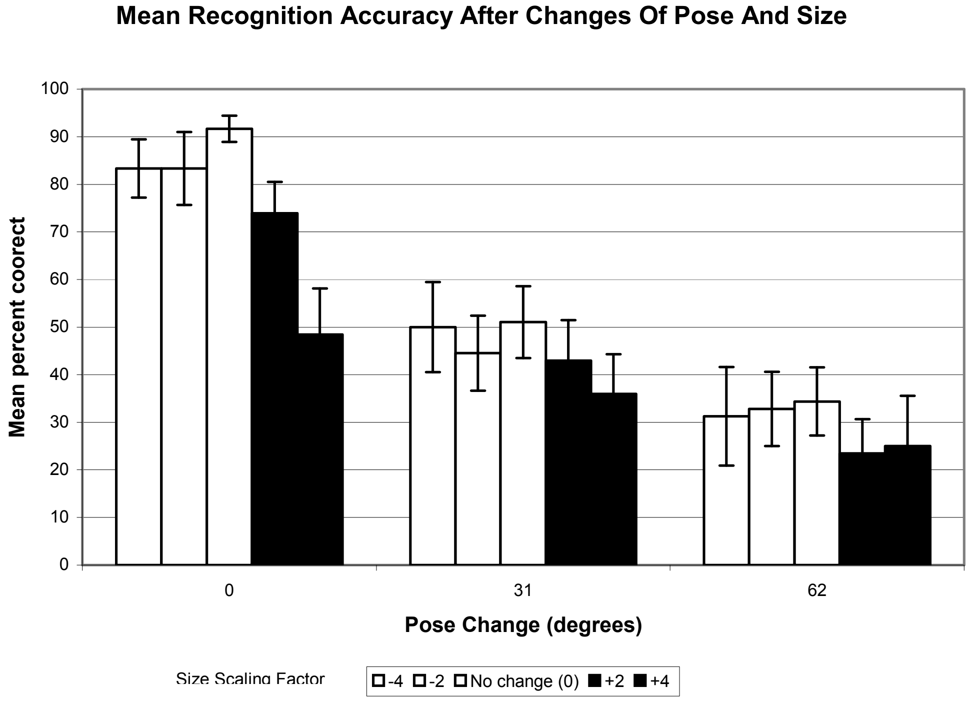


**Figure 3-4: Mean percentage correct recognition scores obtained in Experiment 7 after changes in pose and size. The pose change refers to rotation away from the learnt view. Therefore, a change of 0° means the face was tested in the same viewpoint as it was originally learnt. The size change is indicated by a scaling factor. A size scaling factor of "no change" indicates the face was presented in the same size as it was originally learnt. Scaling factors of +2 and +4 indicate the face was twice and four times larger than the learnt image respectively. Negative factors indicate the test image was smaller than the learnt image. Bars represent mean percent correct and error bars represent standard error.**

As in Experiment 6, the number of data points making up a particular cell varied. Therefore, analysis was conducted on the percentage correct data. Prior to analysis, the data were transformed using the arcsine transformation. The transformed data were entered into a 3x3 within subjects ANOVA with pose change and scaling factor as independent variables. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There was a significant main effect of pose change; $F(2,30) = 99.11$, $MSE = 0.01$, $p < .001$, $\eta_G^2 = 0.059$ (observed power = 1.00). There was also a significant main effect of size change; $F(2.096,31.44) = 5.93$, $MSE = 0.26$, $p < .01$, $\eta_G^2 = 0.011$ (observed power = 0.86) and a significant interaction between pose and size change; $F(7.73,115.92) = 2.44$, $MSE = 0.01$, $p < .025$, $\eta_G^2 = 0.004$ (observed power = 0.88).

All the previous experiments revealed that the greater the angular change in pose between the study and test images, the poorer recognition accuracy became. Planned contrasts on the data of Experiment 7 revealed that, as expected, a pose change of 0° was recognised significantly more accurately than a change of 31° or 62°; $F(1,15) = 158.89$, $MSE = 3.15$, $p < .001$. In turn, a 31° change was recognised significantly better than a change of 62°; $F(1,15) = 26.77$, $MSE = 0.87$, $p < .001$.

The significant interaction was further analysed using simple main effects analysis. An overall effect of pose change was found for the 5 size changes across the 3 pose changes (presented here in order from scaling factor –4 through to +4); $F(2,30) = 18.32$, $MSE = 0.14$, $p < .001$; $F(2,30) = 55.97$, $MSE = 0.04$, $p < .001$; $F(2,30) = 70.47$, $MSE = 0.04$, $p < .001$; $F(2,30) = 37.76$, $MSE = 0.06$, $p < .001$; $F(2,30) = 5.57$, $MSE = 0.09$, $p < .01$. There was also an effect of size change when no change was made in pose; $F(4,60) = 11.32$, $MSE = 0.08$, $p < .001$. However, there was no effect of size change after a rotation of 31°; ($F(4,60) = 1.44$, $p = .232$) or 62° ($F < 1$, *ns*).

### 3.3.3 Discussion

Previous research has suggested that a change in size affects recognition of an unfamiliar face (Kolers et al., 1985) but not of familiar faces (Brooks et al., 2002). If viewpoint-dependent codes are used for the recognition of faces then changing the size of the face should not affect recognition. However, Experiment 7 revealed that, even after a face had been learnt sufficiently well from a single image that the originally studied image is recognised at levels close to ceiling, a change in size between study and test, whilst keeping the pose constant, significantly decreases recognition accuracy. This result suggests that participants are learning about the particular picture studied and are relying primarily on pictorial codes for recognition. In addition, changing the size of a face keeps information about the features of the face highly similar. The errors seen after a change in size are therefore unlikely to be due to difficulties is extracting features that are only available in one particular view.

Interestingly, the significant interaction found between pose change and size change, and the subsequent simple main effects analysis indicates that after a face has undergone a pose transformation there is no effect of changing the size of a face. It is possible that changing the pose of a face between study and test has a sufficiently large enough effect on recognition accuracy that an effect of size change becomes difficult to detect as the pose transformation yielded a large effect size and the size transformation yielded a small effect size. In this case, an effect of size change might still exist (Figure 3-4 suggests that an effect of size change may still be present after a pose transformation of 31°) but it is not possible to detect it with the current experimental design.

In conclusion, changing the size of the face does not alter the viewpoint of the face, therefore if the recognition of face employed the use of viewpoint-dependent codes, changing the size of a face between study and test would not affect recognition accuracy of that face. The lack of size invariance demonstrated in Experiment 7 suggests again that the recognition of faces learnt from a singe photograph is mediated by pictorial codes over and above viewpoint-dependent codes.

## 3.4 General discussion of Experiments 6 and 7

The experiments presented in Chapter Two suggested that the recognition of faces learnt from photographs is mediated via image-based codes such as viewpoint-dependent codes and pictorial codes over and above a three-dimensional viewpoint-invariant structural model, but did not allow for a clear separation of the roles of viewpoint-dependent and pictorial coding. Experiments 6 and 7, presented in this Chapter, were designed to tease apart viewpoint-dependent and pictorial coding.

Experiment 6 was an extension of Experiment 1 to look at recognition after multiple amounts of lighting and pose change. An examination of performance after the transformations of lighting and pose were applied together was also conducted. Considering the lighting change alone, it was found that the greater the lighting change became, the worse was recognition accuracy. Regardless of the lighting change, the viewpoint of the face remained the same suggesting that participants were relying on pictorial codes for recognition. If viewpoint-dependent codes were used for recognition then similar levels of performance regardless of the lighting change would be expected.

In addition, Experiment 6 examined recognition accuracy after changes of lighting and pose were made together. If pictorial codes are used for recognition then the greater the difference between the study and test images, the poorer performance would be expected to be. It was found that performance was significantly poorer after the two transformations were made together compared to when only one of the transformations was made, suggesting again the recognition is mediated by pictorial codes.

In conclusion, transformations of lighting and size were used to investigate the roles of viewpoint-dependent codes and pictorial codes in face recognition. These transformations were chosen as, after the change is made, the face retains the same viewpoint but the image itself has changed. Experiment 6 found that changing the lighting direction of a face between study and test significantly reduced performance, implicating the use of pictorial codes in the recognition process. Experiment 7 examined performance after a transformation of size. The results of Experiment 7 indicated that recognition was significantly poorer after a transformation of size, lending further support to the idea that face recognition is achieved via pictorial codes.

# CHAPTER FOUR

The role of colour information in the recognition of faces

## 4.1 Introduction

The experiments presented in Chapters 2 and 3 suggest that unfamiliar face recognition is mediated by a pictorial code, but from Experiments 1 to 7, little has been established about how this code is used other than it is highly image-specific and any change made to the image can have a detrimental effect on recognition accuracy. However, not all image changes may affect performance. Previous research has indicated that people are able to recognise familiar faces well after hue reversal (Kemp et al., 1996), a transformation in which the hue values of an image are changed to complimentary hue values (i.e. red becomes cyan, green becomes magenta and yellow becomes blue) but the brightness values remain unchanged. Furthermore, there is evidence to suggest that the matching of unfamiliar faces is also unaffected by hue reversal (Bruce & Young, 1998).

Despite the fact that familiar faces have been seen from a number of different images, familiar faces have rarely, and most probably never, been seen after hue reversal. If pictorial codes encode colour information about a face image along with brightness information, then this suggests that colour is not required for the successful identification of faces (although it is clearly stored as even subtle differences in colour, such as when someone familiar to us gets a suntan, are relatively easy to detect). Such a result would place a boundary upon how tightly matched the stored pictorial representation of a learnt face and a probe face must be to enable successful recognition. Experiment 8 therefore investigated the recognition accuracy of faces learnt in one of three colour conditions (either greyscale, coloured green or coloured red) after the transformation of colour.

## 4.2 Experiment 8 – Face recognition after the transformation of colour

Previous research (Kemp et al., 1996) has suggested that colour is not required for successful recognition of familiar faces. Experiment 8 explored the transformation of colour further to examine the effect a colour change has on the recognition of newly learnt faces. In addition, a pose change has been shown to have a significant detrimental effect on recognition accuracy. Therefore, a pose change was also used to provide a point of contrast for the effect of the colour transformation.

### 4.2.1 Method

#### 4.2.1.1 Design

The experiment took the exactly the same form as Experiment 6, except the lighting change of Experiment 6 was replaced by a colour change. Faces were learnt and tested in three different colours (greyscale, green and red). The experiment therefore had a 3x3x3 within-subjects factorial design with pose change (0°, 31° and 62°), learnt image colour (greyscale, green and red) and test image colour (greyscale, green and red) as factors. The dependent variable was the number of faces correctly recognised during the test phase.

#### 4.2.1.2 Participants

Sixteen undergraduate students (3 male and 13 female) aged between 18 and 23 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision and none had participated in any of the previous experiments.

#### 4.2.1.3 Materials

Images of 18 faces (all Caucasian and male) from the PIE face database were used. None of the images depicted an individual with facial hair or wearing glasses. Each face was used in three poses (0°, 31° and 62°) with three versions of each pose created, one for each of three colours (original greyscale, red and green), resulting in a total of 162 images. All images had full-face lighting. Each image was manipulated to remove all irrelevant background information, to leave only the head visible. The background was replaced with a homogenous grey. The original photographs were in colour. To artificially colour the face images, the images were first converted to greyscale. Two further copies of each face image were then made by colourising the image with either red or green, thus creating three versions of each facial image (greyscale, red and green). Each image was resized so that it was 384 pixels high in order to normalize face height and the background expanded to create a final image of 384x384 pixels yielding a visual angle of 4.87° when viewed from a distance of 60cm. Figure 4-1 shows an example of the faces used.

**Figure 4-1: Examples of the face stimuli used in Experiment 8. The three poses used (from left to right) were 0° (rotation away from full-face), 31° and 62°.**

The 18 faces were randomly placed into two sets of nine faces each before the experiment began and the faces remained in these groups for the all participants.

#### 4.2.1.4 Apparatus

The same apparatus as used the previous experiments was employed.

#### 4.2.1.5 Procedure

The experimental procedure was the same as Experiment 6, except that the lighting change was replaced by a colour change. Thus, during the two learning phases, participants learnt nine faces with one face presented in each pose/colour combination (pose/colour; 0°/greyscale, 0°/green, 0°/red, 31°/greyscale, 31°/green, 31°/red, 62°/greyscale, 62°/green and 62°/red). During the test phase, participants saw all nine targets and the nine distractors in all pose/colour combinations.

### 4.2.2 Results

The mean percent correct scores obtained for faces learnt in greyscale, green and red after transformations of colour and pose are shown in Figure 4-2, Figure 4-3 and Figure 4-4 respectively.

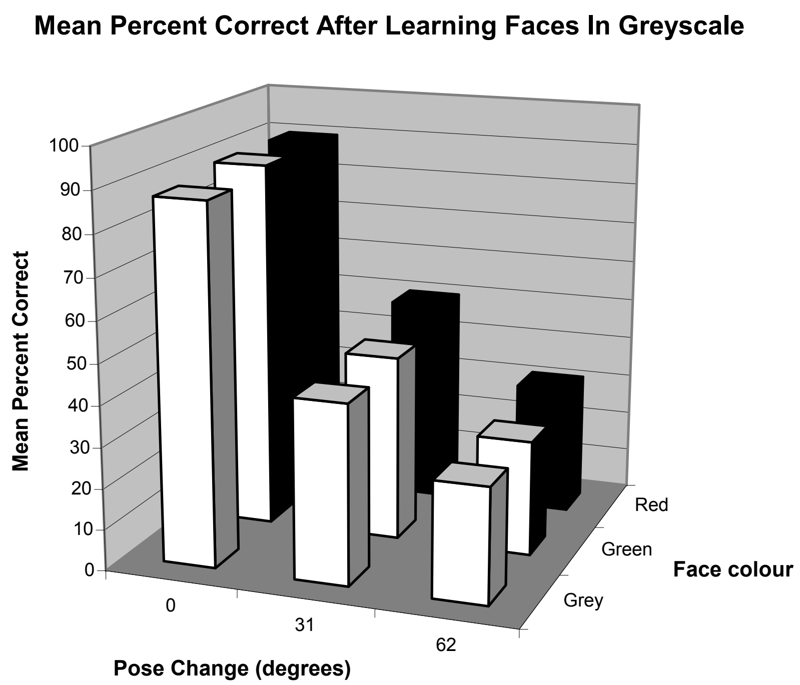


**Figure 4-2: Mean percent correct recognition obtained in Experiment 8 for faces learnt in greyscale. Columns represent mean percent correct. Pose change refers to the angle of rotation**

**away from the learnt view (i.e. a change of 0° indicates that recognition of the face was tested in the same viewpoint as it was learnt). Face colour indicates the colour of the face at test.**

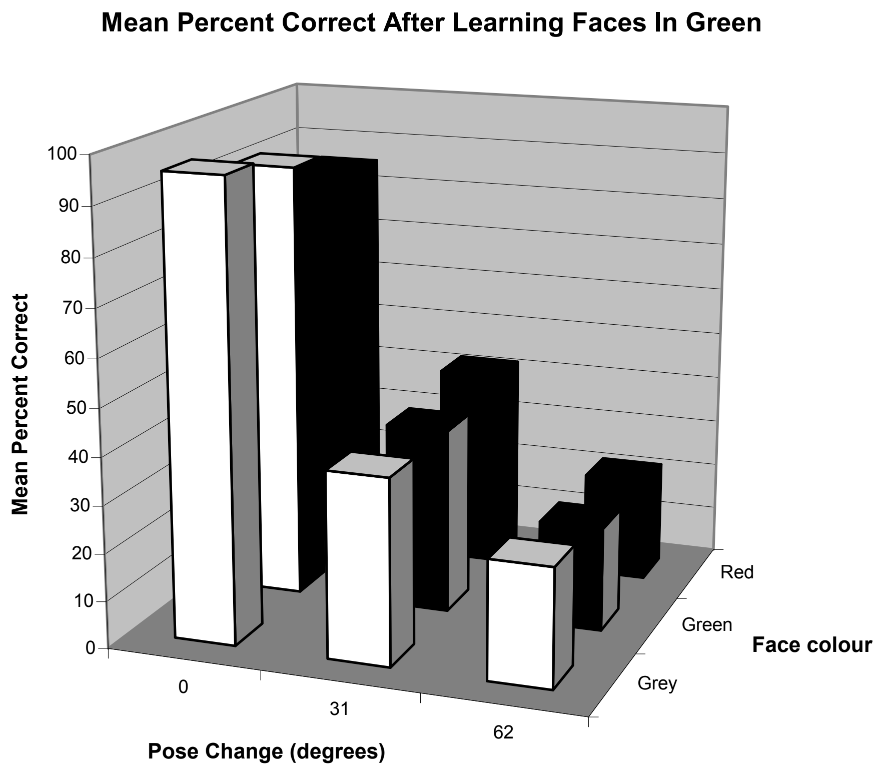


**Figure 4-3: Mean percent correct recognition obtained in Experiment 8 for faces learnt in green. Columns represent mean percent correct. Pose change refers to the angle of rotation away from the learnt view (i.e. a change of 0° indicates that recognition of the face was tested in the same viewpoint as it was learnt). Face colour indicates the colour of the face at test.**

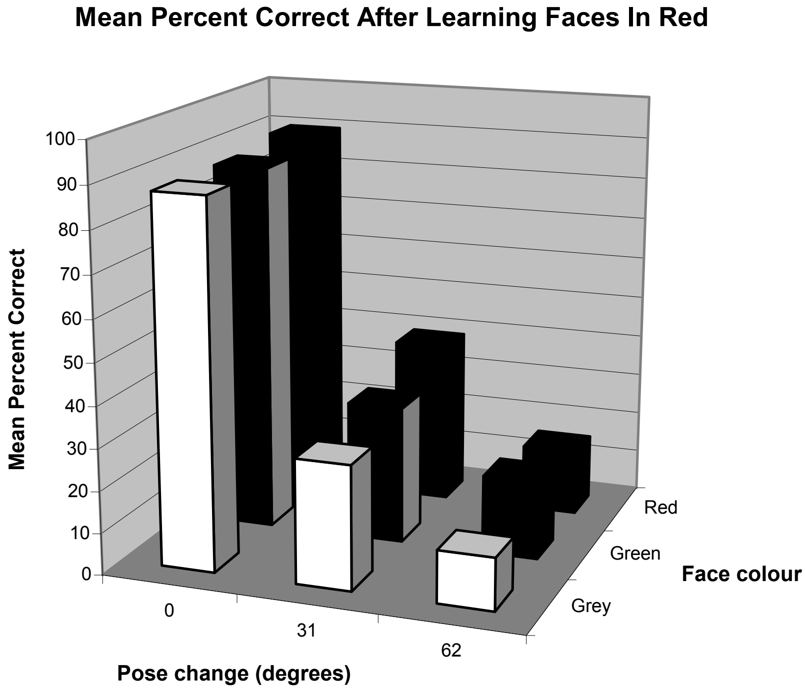


**Figure 4-4: Mean percent correct recognition obtained in Experiment 8 for faces learnt in red. Columns represent mean percent correct. Pose change refers to the angle of rotation away from the learnt view (i.e. a change of 0° indicates that recognition of the face was tested in the same viewpoint as it was learnt). Face colour indicates the colour of the face at test.**

As in Experiment 6, the number of data points making up a particular cell varied. Therefore, analysis was conducted on the percentage correct data. Prior to analysis, the data were transformed using the arcsine transformation. The resulting data were entered into a 3x3x3 within subject ANOVA with pose (0°, 31° and 62°), learning image colour (greyscale, green and red) and test image colour (greyscale, green and red) as independent variables. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There the expected significant main effect of pose change; $F(1.23,18.47) = 125.10$, $MSE = 0.49$, $p < .001$, $\eta_G^2 = 0.100$ (observed power

= 1.00). There were no effects of learn image colour; $F(1.57,23.59) = 2.66$, $p = .102$ or test image colour; $F < 1$, *ns*. There was no significant interaction between pose change and learn image colour, nor an interaction between pose and test image colour (both $F < 1$, *ns*). There was however an unexpected interaction between learn image colour and test image colour; $F(4,60) = 4,34$, $MSE = 0.07$, $p < .005$, $\eta_G^2 = 0.002$ (observed power =0.91). The 3-way interaction was not significant ($F < 1$, *ns*).

All the previous experiments revealed that the greater the angular change in pose between the study and test images, the poorer recognition accuracy became. Planned contrasts on the data of Experiment 8 revealed as expected, a 0° pose change was recognised better than a 31° or 62° change; $F(1,15) = 126.18$, $MSE = 29.51$, $p < .001$. Further, a 31° change was recognised better than a 62° change; $F(1,15) = 113.75$, $MSE = 0.94$, $p < .001$. Thus the expected pattern of results following a pose change was found.

To investigate the interaction between learnt image type and test image type, simple main effects analysis was conducted. It was found that faces were, overall, recognised less well if they had been learnt in red than either in greyscale or green; $F(2,30) = 7.57$, $MSE = 0.49$, $p < .005$. It was also found that faces were less well recognised if they were tested in greyscale compared to red or green; $F(2,30) = 7.27$, $MSE = 1.46$, $p < .005$.

### 4.2.3 Discussion

Previous research has suggested that colour information is not required for the recognition of familiar faces from good quality photographs (Kemp et al., 1996; Yip & Sinha, 2002) but that colour cues can aid face recognition when the stimulus is degraded through blurring (Yip & Sinha, 2002). The fact that we can recognise faces from unusual colourings (or lack of colour as in the case of greyscale images) should not be considered surprising. Without this ability, paling (e.g. through illness) or darkening (e.g. from a suntan) of the skin would result in difficulty to recognise known faces. How we can recognise faces despite changes in colour is not well understood however.

The colour invariance shown by familiar faces may be due to one of two possibilities. Firstly, it may be that through extensive experience with individual familiar faces, and in particular through the different lighting conditions which lead to subtle changes in the appearance of the colour of the face, each known face has been seen in a sufficient number of different colour shades. Secondly, it is possible that colour information is simply not required for the recognition process. The familiar faces used by Kemp et al. (1996) will almost certainly never have been seen after hue reversal, a transformation which is a long way from the normal day-to-day subtle changes in colour appearance which are experienced with familiar faces, suggesting a limited role for colour in the recognition process.

Experiment 8 revealed that faces learnt from a single photograph also demonstrate colour invariance such that faces are recognised well if a change in colour is made between study and test. Unlike familiar faces which have been seen in many subtly different colour shades, the faces seen in Experiment 8 had only ever been experienced in one unique colour. There was no opportunity during training to view the face in subtly different colours to assist participants in the recognition process. Therefore, it is unlikely that seeing the face in a number of subtly different colourings allows the recognition process to cope with a dramatic change such as hue reversal.

It has been suggested that faces are recognised well despite changes in colour because the transformation of colour does not affected shape-from-shading cues (Kemp et al., 1996; Bruce & Young, 1998). Thus, the three-dimensional structure of a face can be successfully reconstructed from an image of a face regardless of colour information, which is in turn used by the recognition process. However, the results of Experiments 1 through to 7 suggest that a three-dimensional structural model is not used for recognition. Instead, it seems to be the case that it is the image itself that is of importance. It would therefore appear that light intensity levels (i.e. a greyscale description) of an image of a face, are sufficient for successful recognition. The results presented in Experiment 8 therefore provide a boundary as to how similar the study and test images must be in order to enable successful recognition of faces learnt from photographs. It must be remembered, however, that the images used in Experiment 8 were of good quality. As suggested by Yip and Sinha (2002), colour may play a more prominent role in the recognition of faces from degraded images.

# CHAPTER FIVE

What cues present in a face are useful for face learning and recognition?

## 5.1 Introduction

The experiments reported thus far in this thesis have explicitly focused upon how faces, learnt from a single photograph or two photographs, are recognised after transformations of pose, lighting, size and colour. These transformations have been shown to have little or no effect on the recognition of familiar faces, yet the recognition of unfamiliar faces is adversely affected by most of them. Indeed, the best way to obtain high levels of recognition for unfamiliar faces is to perform the recognition task with the same images that were studied. It seems that, overall, faces are recognised primarily via pictorial codes over and above viewpoint-dependent and viewpoint-invariants codes; successful recognition requires that the image of the face presented for identification is similar to a stored representation of that person.

The idea that faces are recognised via pictorial codes would seem to suggest that the image is treated as a whole picture. However, particular features of the face might be more informative about identity than other features. Different parts of the face could be construed as distinctive for different people (e.g, a large nose or outlandish hairstyle) whilst some parts are important for social interaction. For example, the McGurk effect (McGurk & Macdonald, 1976), in which the perception of speech is influenced by the movement of the mouth, demonstrates the importance of the mouth for social communication. Likewise, the eyes receive a considerable proportion of attention as they provide an indication as to the focus of attention of a person we are interacting with. Indeed, eye tracking studies have revealed that both the eyes and mouth receive a higher proportion of fixations during recognition than other features of the face (Althoff & Cohen, 1999; Luria & Strauss, 1978; Stacey et al., 2005).

The features of the face can be grouped into two broad groups; the internal features, consisting of the eyes, eyebrows, nose and mouth and the external features, comprising the hair, ears and face shape. It has been shown that these two groups of features are of differing importance, depending upon whether the face to be recognised is familiar or unfamiliar. The internal features of familiar faces are processed more accurately (Ellis et al., 1979; Endo et al., 1984) and faster (Young et al., 1985) than the external features. Unfamiliar faces on the other hand show little or no advantage for either type of feature (with any advantage occurring for the external features). Several studies have also shown an internal feature advantage for newly learnt faces in a matching paradigm (Bonner et al., 2003; Clutterbuck & Johnston, 2002; Clutterbuck & Johnston, 2004; Clutterbuck & Johnston, 2005). However, no demonstrations have been reported of an internal feature advantage for newly learnt faces using a recognition paradigm.

This Chapter focuses on which particular features of the face are useful for recognition after a studied image is learnt to a high level of recognition. Experiment 9 examined whether faces learnt from a single photograph exhibit an internal feature advantage using a recognition paradigm. In addition, it investigated whether any advantage for the internal features that may arise is consistent across a transformation of pose. During Experiments 1 to 8, participants informally reported that of the external features, the found that the hair particularly useful for the training and recognition tasks. Experiments 10 and 11 investigated the role played by the hair in the recognition of faces.

## 5.2 Experiment 9 – An internal feature advantage for faces learnt from photographs?

Experiment 9 examined whether faces, learnt from a single view during an experimental procedure, demonstrate an internal feature advantage in a recognition task. Participants received either a single exposure to a face or were trained to achieve good recognition of a single photograph of a face. They were then tested for their recognition accuracy of the face from the whole image, the internal features of the face, and the external features. This was conducted for both the same viewpoint as originally learnt and an image of the face in a different pose. Presenting the face only once during learning is expected to leave the face relatively unfamiliar and thus, faces would not be expected to demonstrate an internal feature advantage during a recognition task. Faces that have been learnt well from a single view through training may show a benefit during recognition for the internal features. Such a result would suggest that the internal feature advantage can arise after experience of only a single image of a face and is not only the result of experience of multiple views of the face.

### 5.2.1 Method

#### 5.2.1.1 Design

The experiment had a similar design to that of Experiment 1. Participants were randomly allocated to one of two groups; a training group which received multiple exposures to the faces or no-training group who received only a single exposure to each face. After receiving exposure to the faces the participants were required to recognise newly learnt faces from two poses (a change of 0° and 31°) and from three types of image (the whole face, the internal features only and the external features only). The experiment therefore had a 2x2x3 design with training group, type of image and pose change as independent variables. The dependent variable was the number of faces correctly identified during the test phase.

#### 5.2.1.2 Participants

Twenty-four undergraduate students (4 males and 20 females) aged between 18 and 22 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision. None had participated in the previous experiments.

#### 5.2.1.3 Materials

Images of 20 faces (all Caucasian and male) from the PIE face database were used. None of the images depicted an individual with facial hair or wearing glasses. Each face was used in two poses and three image types (whole face, internal features and external features), resulting in a total of 120 images. The two poses were taken from directly in front of the model (0°) or to their right (31°). Each image was manipulated to remove all irrelevant background information, to leave only the head visible. The background was replaced with a homogenous grey. The original colour images were converted to greyscale. Each image was resized so that it was 384 pixels high in order to normalize face height and the background expanded to create a final image of 384x384 pixels, yielding a visual angle of 4.87° when viewed from a distance of 60cm. Two further variations of each image were made to show only the internal features (eyes, nose and mouth) or external features (hair, face shape and ears). The 20 faces were randomly placed into two sets of nine faces each before the experiment began and the faces remained in these groups for all participants.

#### 5.2.1.4 Apparatus

The same apparatus as used in the all the previous experiments was employed.

#### 5.2.1.5 Procedure

Participants were randomly assigned to either the single (no training) or multiple (training) exposure conditions. For half the participants, the first set of faces was allocated as the target set and the second was allocated as distractors. For the other half, this order was reversed.

All participants completed a first presentation phase and a test phase of the experiment. Those in the multiple exposures condition also received training between the first phase and the test phase. The participant sat in front of the computer screen at a distance of approximately 60cms and was given written instructions before the experiment began.

#### *5.2.1.5.1 First presentation phase*

During the first presentation phase, participants saw ten faces for a duration of 5 seconds each with 0.5 seconds between each face. The faces were evenly distributed across the two poses (full-face and three-quarter view) so that five faces were seen in each pose. Each individual face was presented to the participant once and was accompanied by a name, presented below the image of the face. These name/face pairings were randomly generated for each participant using a fixed set of ten names. All the facial images shown depicted the whole face.

#### *5.2.1.5.2 Training phases*

The training task was completed by the multiple exposures group only and took the same form as that used in the previous experiments. In the first part, the ten face photographs shown during the first presentation phase were divided into two blocks containing five faces each. In the second part, all ten face photographs were presented in a single block. As in the first presentation phase, only images showing the whole face were used.

#### *5.2.1.5.3 Test phase*

The testing phase was divided into six blocks of 20 individually presented faces, for which the participant made a "yes/no" decision as to whether they had seen the presented face during familiarisation. All the faces within each block contained faces presented in the same pose (0° or 31°) and image type (whole image, internal features only or external features only) combination. The order of the presentation of the blocks was rotated across participants.

For each block, ten of the individuals were those in the familiarisation set, whilst the other ten were taken from the distractor set. Faces were presented one at a time and two buttons (labelled “Yes” and “No”) were used for responses. Participants were required to click on “Yes” if they thought they recognised the individual as a member of the training set and “No” if they did not.

### 5.2.2 Results

An average percentage correct score was calculated for each participant based on the number of hits obtained during the test phase. These data are shown in Figure 5-1 and Figure 5-2 respectively. The number of hits (faces correctly recognised as being members of the familiarisation set) was also used for analysis.

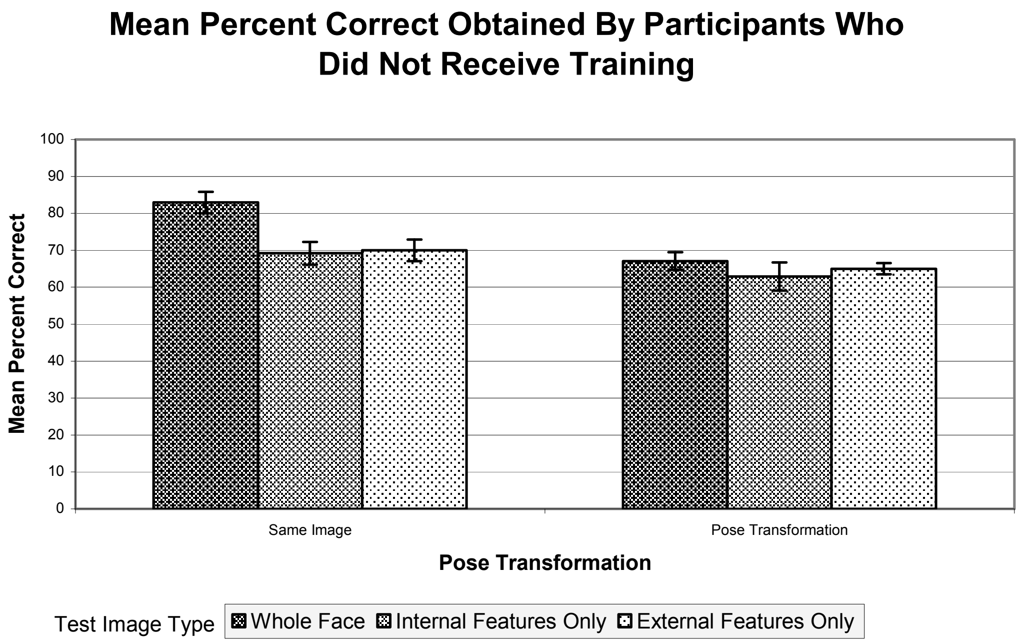


**Figure 5-1: Mean percentage correct recognition for participants in Experiment 9 who did not receive training for three facial image types (whole face, internal features only and external features only). Pose change refers to the difference in rotation between the learnt view and the test view measured in degrees. Bars represent mean percent correct and error bars represent standard error.**

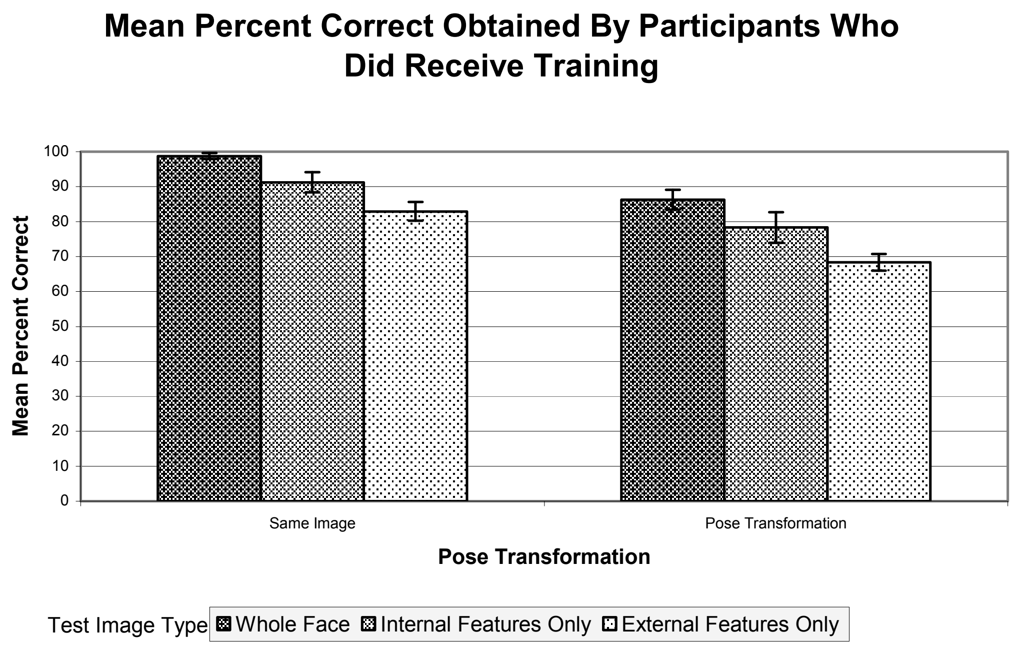


**Figure 5-2: Mean percentage correct recognition for participants in Experiment 9 who did receive training on whole face images for three facial image types (whole face, internal features only and external features only). Pose change refers to the difference in rotation between the learnt view and the test view measured in degrees. Bars represent mean percent correct and error bars represent standard error.**

The hit rates were entered into a mixed design 2x2x3 ANOVA with training condition (single or multiple presentation, between-subjects), pose change (0° or 31°, within-subjects) and image type (whole face, internal features and external features, within-subjects) as independent variables and number of hits obtained in the recognition task as the dependent variable. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There was a significant main effect of training; $F(1,26) = 15.262$, $MSE = 47.19$, $\eta_G^2 = 0.330$, $p < .01$ (observed power = 0.96), a significant main effect of pose; $F(1,26) = 51.35$, $MSE = 2.61$, $\eta_G^2 = 0.084$, $p < .001$ (observed power = 1.00) and of image type; $F(1.67,43.36) = 15.29$, $MSE = 2.26$ $\eta_G^2 = 0.038$, $p < .001$ (observed power = 0.99). There was a significant 3-way interaction between pose, image type and training type; $F(1.94,50.38) = 4.00$, $MSE = 1.34$, $\eta_G^2 = 0.007$, $p < .05$ (observed power = 0.68). A two-way interaction between pose and image type approached significance; $F(1.94,50.38) = 3.08$, $MSE = 1.34$, $\eta_G^2 = 0.005$, $p = .056$ (observed power = 0.56).

All the previous experiments revealed that the greater the angular change in pose between the study and test images, the poorer recognition accuracy becomes. Planned contrasts on the data of Experiment 9 revealed as expected that overall, a change in pose of 31° significantly decreased recognition accuracy across compared to no change in pose; $F(1,26) = 51.35$, $MSE = 15.65$, $p < .001$. It was also expected that in this analysis which includes data from both the training and non-training groups, the whole face image would be recognised better than either the internal features or the external features. This is because participants were familiarised with whole face images. Those in the untrained group will perform best for the same type of image as they saw during the first presentation phase, whilst those that received training will be expected to learn the training image close to 100% accuracy and a drop off in performance for the internal and external feature test conditions was likely. Planned contrasts confirmed this with the whole face image being recognised better than either the internal or external features; $F(1,26) = 53.87$, $MSE = 11.38$, $p < .001$. Overall, the internal and external features did not differ significantly. The recognition of the internal and external features before and after training was examined in a later analysis.

Simple main effects were calculated to examine the 3-way interaction between training group, pose change and image type. It was found that when no training was given, a pose transformation significantly decreased the recognition accuracy of faces from the whole image; $F(1,26) = 28.59$, $MSE = 1.71$, $p < .0001$ but made no change to the recognition of the internal features; $F(1,26) = 1.70$, $MSE = 1.34$, $p = .203$ or the external features; $F(1,26) = 2.01$, $MSE = 2.15$, $p = .168$. In addition, there was a significant effect of image type when no pose change was made; $F(2,52) = 16.06$, $MSE = 1.16$, $p < .001$ but there was no effect of image type after a pose change ($F < 1$,*ns*). Contrasts revealed that in the no pose change condition, the whole face was recognised significantly better than either the internal features or the external features; $F(1,13) = 48.53$, $MSE = 4.615$, $p < .001$. Recognition of the internal features and the external features did not differ ($F < 1$, *ns*).

When training was given, a pose change significantly decreased the recognition accuracy of all three test image types; $F(1,26) = 18.80$, $MSE = 1.71$, $p < .001$; $F(1,26) = 25.57$, $MSE = 1.34$, $p < .001$; $F(1,26) = 18.13$, $MSE = 2.15$, $p < .001$ for the whole face, internal features only and external features only. There was an effect of image type when no pose change took place between study and test; $F(2,52) = 8.13$, $MSE = 1.16$, $p < .005$. Contrasts revealed that the whole face was recognised significantly better than the internal or external features; $F(1,13) = 22.98$, $MSE = 3.81$, $p < .001$ and the internal features were recognised better than the external features; $F(1,13) = 5.52$, $MSE = 1.57$, $p < .05$. There was also a significant effect of test image type after a pose change was made; $F(2,52) = 5.99$, $MSE = 2.02$, $p < .01$. Contrasts showed that the whole face image was recognised significantly better than either the internal or external features; $F(1,13) = 14.97$, $MSE = 7.26$, $p < .005$ but the internal and external features were recognised equally well; $F(1,13) = 1.24$, $MSE = 9.76$, $p = .286$.

### 5.2.3 Discussion

This experiment was designed to examine whether the internal feature advantage typically seen for familiar faces, learnt and tested in a matching paradigm arises in a recognition task. As would be expected, it was found that participants who received training performed better (with a large effect size) than those that did not.

The three-way interaction revealed that, without training, the whole face image was recognised more accurately than either the internal or external features when the same image was used at study and test. After only a single exposure, the face would still be unfamiliar and the lack of difference between performance on the internal and external features is consistent with previous work (Ellis et al., 1979). The whole face image was recognised significantly better than either the internal or external features, suggesting a heavy reliance on pictorial codes for recognition. Whilst parts of internal and external images are identical to the training image, the removal of either the internal or external features appears to have altered the image enough that recognition of the face becomes much more difficult. After a change in pose overall performance dropped significantly and there was no difference between the three test image types. However performance was approximately 50%, which is chance level in a binary choice task, suggesting that participants found the task of recognising a face seen only once after a pose change too hard.

After training a very different picture emerges. When no pose change is made between study and test the whole face image is recognised significantly better than either the internal or external features. This again suggests that participants are relying on pictorial codes for recognition, as changing the image from the one originally studied reduced performance. Interestingly, however, there was a significant advantage in terms of accuracy for testing with the internal features over the external features, suggesting that a shift towards using the internal features for recognition had begun. This shift towards the internal features also suggests that through the training process the participants had started to become familiar with the faces.

Changing the pose of the face reduced performance significantly. After the pose transformation the whole face image was recognised more accurately than either the internal or external features. However, after the pose change, there was no statistical benefit for testing with the internal features over the external features. There are at least two potential explanations for this particular result. Firstly, from Figure 5-4 it appears that after a pose change has been made, the internal features are recognised better than the external features yet the data exhibit a greater amount of variance which leads to a non-significant statistical result. The second possibility is that an internal feature advantage formed from a particular view of a face does not generalise to novel views. Considering the results of the experiments in Chapters Two, Three and Four, in which face recognition was suggested to be heavily image dependent, this possibility is quite plausible.

In conclusion, the current experiments aimed to investigate whether the internal feature advantage associated with familiar faces can be observed after a training procedure in which the face is learnt from a single photograph. This result was indeed found, suggesting that faces learnt from photographs can display at least some of the properties of familiar faces. The internal feature advantage does appear however to be limited to the image of the face studied as it does not emerge after a pose change, suggesting that the internal feature advantage is view-specific.

## 5.3 Experiment 10 – The role of hairstyles in the recognition of faces

Experiment 9 demonstrated that the internal features of familiar or newly learnt faces can show an advantage in terms of recognition accuracy over the external features. However, the external features contain one very distinctive feature – the hair. Many participants in Experiments 1 through to 9 informally reported after the experiment that they felt the hairstyle helped them complete the training task. This could be considered a natural cue to use. When we meet new people, the hairstyle might be a primary cue for recognition as it is so distinctive. In the long term, as we get to know someone better, the internal features take a more prominent role in recognition as they are relatively stable in comparison to the external features (i.e. they are less likely to undergo a large transformation the drastically alter the appearance of the face).

Experiment 10 therefore examined the role of hair in the recognition of newly learnt faces. It may be that the hair, whilst an easy cue to extract, is actually misleading for the recognition process, resulting in a reduction in recognition accuracy in comparison to when either the internal features or the whole face is used. Removal of the hair from an image will force other parts of the image to be used for recognition instead. The features then used might display a greater degree of invariance in recognition. Experiment 10 was thus a replication of Experiment 3 except participants learnt faces with the hairstyle removed.

## 5.3.1 Method

### 5.3.1.1 Design

The experiment followed the same design as that of Experiment 3. A 3x3 within-subjects factorial design was used with learning view (full-face, profile or both) and test view (full-face, three-quarter or profile) as factors. The dependent variable was the number of faces correctly recognised during the test phase.

#### 5.3.1.2 Participants

Sixteen participants (14 females and 2 males) aged between 18 and 57 years from the University of York took part in the experiment in return for course credit or payment. All participants had normal or corrected-to-normal vision. None had participated in any of the previous experiments.

#### 5.3.1.3 Materials

The 24 face images from Experiment 3 were used, except that the hairstyle was removed from each of the faces. It was decided to remove the hair in a broad fashion rather than finely modifying each image by cutting along the hairline as this may create a highly distinctive shape around the top of the head that could be used for recognition. Examples of the images used are shown in Figure 5-3.

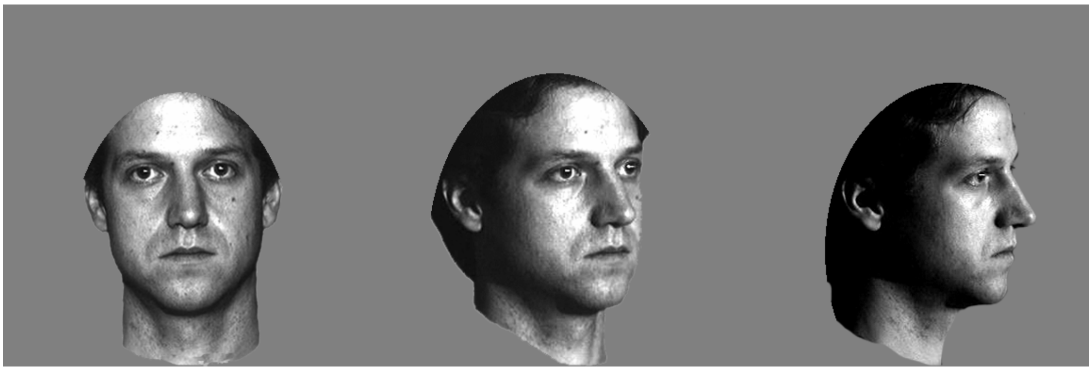

**Figure 5-3: Examples of the facial images used in Experiment 10.**

The faces were split into the same two groups of 12 faces as for Experiment 3.

#### 5.3.1.4 Apparatus

The same apparatus as used in the previous experiments was employed.

#### 5.3.1.5 Procedure

The procedure was identical to that of Experiment 3.

### 5.3.2 Results

Mean percentage correct scores were calculated from the number of hits for full-face only, profile only and both full-face and profile view learning conditions for the three test views. These data are shown in Figure 5-4. The number of hits (faces correctly recognised as being members of the familiarisation set) was used for analysis purposes.

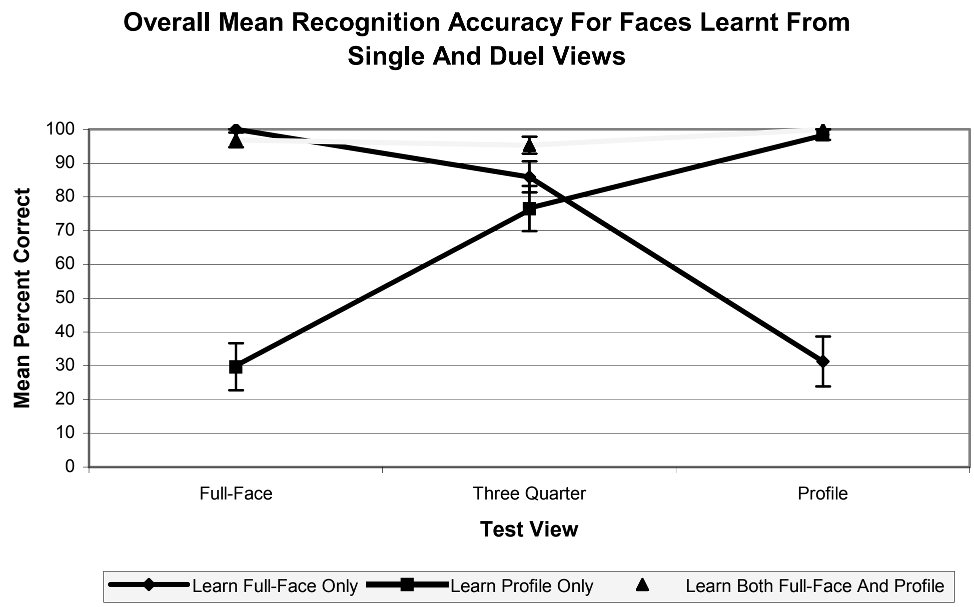


**Figure 5-4: Mean percent correct recognition obtained in Experiment 10 for faces learnt from with the hairstyle removed from the full-face view only, profile only and both full-face and profile views on the three test conditions. Lines represent mean percent correct. Errors bars represent standard error.**

A 3x3 repeated-measures ANOVA was conducted with learned view and test view as independent variables and number of hits as the dependent variable. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There were significant main effects of learning view condition; $F(2,30) = 33.95$, $MSE = 0.56$, $p < .001$, $\eta_G^2 = 0.165$ (observed power = 1.00) and of test view; $F(1,44,21.62) = 8.69$, $MSE = 0.40$, $p < .005$, $\eta_G^2 = 0.025$ (observed power = 0.89). There was also a significant interaction between learning type and test view; $F(2.89,43.42) = 66.81$, $MSE = 0.65$, $p < .001$, $\eta_G^2 = 0.394$ (observed power = 1.00).

The main effect of learning condition was expected as when two views are learnt, very high levels of performance are expected for both the two learnt views. When only one view (either the full-face or profile view) is learnt, performance is only high for one view. Therefore, the two view learning condition would be expected to produce, overall, higher levels of accuracy. Indeed, planned contrasts revealed that the two view learning condition led to overall higher accuracy rates; $F(1,15) = 53.37$, $MSE = 12.67$, $p < .001$ whilst the full-face only and profile only learning conditions did not differ significantly ($F(1,15) = 1.58$, $p = .228$).

The overall main effect of test view was also expected. This is because when a single view is learnt (either the full-face view or the profile view) and testing takes place on the most extreme change (i.e. the full-face view is learnt and recognition is tested for the profile view and vice-versa), performance is poor. This has the effect of lowering the average level of recognition on the full-face and profile views. Therefore, the three-quarter view would be expected to produced overall better levels of recognition. Contrasts revealed that the three-quarter test view was indeed, overall, recognised better than either the full-face or profile views; $F(1,15) = 11.70$, $MSE = 7.72$, $p < .005$ whilst the full-face and profile test views did not differ from each other ($F < 1$, *ns*).

A simple main effects analysis was carried out to analyse the strong interaction. Recognition on the profile view was poor after a full-face view alone was learnt, as was recognition of the full-face view after a profile view alone was learnt, a result which contrasts sharply with the high levels of accuracy obtained when the same view (and therefore image) was used for both learning and testing. Consequently, an expected effect of learning condition for both the full-face test view; $F(2,30) = 79.68$, $MSE = 0.51$, $p < .001$ and the profile test view; $F(2,30) = 81.83$, $MSE = 0.48$, $p < .001$ was found. In addition, an effect of learning condition was also found at the three-quarter test view; $F(2,30) = 4.35$, $MSE= 0.52$, $p < .025$. Contrasts were performed to find the source of this effect. It was found that the two-view learning condition led to higher levels of performance than either the full-face only or profile view only learning conditions; $F(1,15) = 5.87$, $MSE = 3.45$, $p < .05$. The full-face only and profile only learning condition did not differ significantly from either other; $F(1,15) = 2.46$, $MSE = 0.92$, $p = .138$.

As in Experiments 3, 4 and 5, generalisation slopes were found when a single view was learnt, indicated by an effect of test view for both the full-face view, $F(2,30) = 66.98$, $MSE = 0.50$, $p < .001$ and profile view, $F(2,30) = 48.59$, $MSE = 0.65$, $p < .001$ learning conditions. However, no effect of test view was found in the two view learning condition; $F(2,30) = 1.84$, $MSE = 0.08$, $p = .176$.

### 5.3.3 Discussion

Experiment 10 examined the role played by hairstyle in the learning of new faces. By removing the hair, which is a potentially misleading cue to use for recognition, it was hoped that greater invariance would be achieved after learning two views of a face due to the direction of attention during learning to more invariant features of the face. It was found that recognition on the crucial, previously unseen three-quarter test view was significantly higher after two views of the face had been learnt than if only a single view had been learnt. This indicates that participants were achieving a greater level of invariance after learning two views of a face in comparison to learning a single view. Furthermore, it was also found that performance on the three-quarter test view after two views of the face were learnt was not significantly below that obtained for the learnt views themselves, indicating that participants were demonstrating generalisation to a novel view of a face.

It is quite plausible therefore that the hairstyle could be a misleading feature. Removing the hair requires that participants focus upon other features of the face for both learning the faces and for performing the recognition task. If the hair is indeed being heavily relied upon, the drop in recognition accuracy typically seen after a face undergoes a pose transformation, is due the hairstyle taking on a sufficiently different enough appearance to suggest that the face belongs to a different person. To explore this issue further, Experiment 11 examines peoples' performance when they are asked to learn the hairstyle only and then recognise it after a transformation of pose.

## 5.4 Experiment 11 – Recognition of hairstyles

Experiment 11 extends Experiment 10 to further examine the role played by the hairstyle in the learning and recognition of new faces. In Experiment 10, it was suggested that the hairstyle might be a potentially misleading cue to use for recognition and might contribute to the difficulty people experience in identifying unfamiliar and relatively new faces after a transformation of pose. Experiment 11 examined the recognition of hairstyles over the transformation of pose. The experiment was a replication of Experiment 10 except the images presented during learning and test consisted only of the hairstyle. It is expected that when the same image is used in both the learning and testing phases then performance will be good, but this will drop off markedly when a transformation of pose is applied. If little invariant information is present in the hairstyle then a different amount of pose change will yield similar recognition rates as they will all be approximately equally difficult to identify.

### 5.4.1 Method

#### 5.4.1.1 Design

The experiment has the same design as that of Experiments 3 and 10. A 3x3 within-subjects factorial design was used with learning view (full-face, profile or both) and test view (full-face, three-quarter or profile) as factors. The dependent variable was the number of faces correctly recognised during the test phase.

#### 5.4.1.2 Participants

Twelve participants (14 females and 2 males) aged between 18 and 32 years from the University of York took part in the experiment in return for course credit or payment.

All participants had normal or corrected-to-normal vision. None had participated in any of the previous experiments.

#### 5.4.1.3 Materials

The 24 faces from Experiment 3 were adapted for use as stimuli. The images were prepared as in Experiment 3 except that all facial information *except* hairstyle was removed from each of the images. The hairstyles were created using a mask from the images used in Experiment 10. These images are thus the part of each image that the participants in Experiment 10 did not see. An example of the image used is shown in Figure 5-5.

**Figure 5-5: Examples of the images of hairstyles used in Experiment 11. Each image depicts the part of the whole face not seen in Experiment 10.**

The images were split into the same two sets of 12 people as used in Experiments 3 and 10.

#### 5.4.1.4 Apparatus

The same apparatus as used in the previous experiments was employed.

#### 5.4.1.5 Procedure

The procedure was identical to that used in Experiments 3 and 10.

### 5.4.2 Results

Average percentage correct scores were calculated from the number of hits and the number of correct rejections for front view learning only, profile view learning only and both full-face and profile view learning conditions for each of these test views. These data are shown in Figure 5-6. The number of hits (hairstyles correctly recognised as being members of the familiarisation set) was used for analysis purposes.

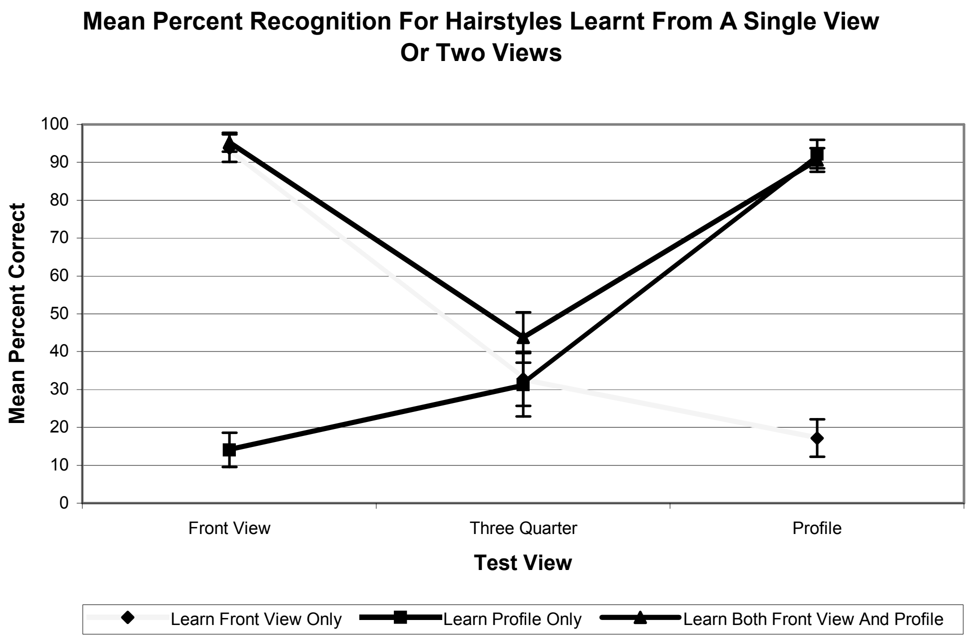


**Figure 5-6: Mean percent correct recognition obtained in Experiment 11 for hairstyles learnt from a front view only, profile only and both front view and profile views for the three test views. Lines represent mean percent correct and error bars represent standard error.**

A 3x3 repeated-measures ANOVA was conducted with learned view and test view as independent variables and number of hits as the dependent variable. The Huynh-

Feldt correction for departures from sphericity was used throughout the analyses and effect sizes are calculated using generalised eta-squared (Bakeman, 2005). Effect sizes are measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect and 0.26 for a large effect (Cohen, 1988). There were significant main effects of learning type; $F(1.88,28.21) = 38.13$, $MSE = 0.63$, $p < .001$, $\eta_G^2 = 0.120$ (observed power = 1.00) and of test view; $F(1,58,23.70) = 34.53$, $MSE = 0.92$, $p < .001$, $\eta_G^2 = 0.131$ (observed power = 1.00). There was also a significant interaction between learning type and test view; $F(4,60) = 84.92$, $MSE = 0.48$, $p < .001$, $\eta_G^2 = 0.327$ (observed power = 1.00).

As in Experiment 10, the main effect of learning condition was expected as when two views are learnt, very high levels of performance are expected for both the two learnt views. When only one view (either the front or profile view) is learnt, performance is only high for one view. Therefore, the two view learning condition would be expected to produce, overall, higher levels of accuracy. Planned contrasts revealed that as expected, two view learning led to better overall performance than learning either the front view or profile views only; $F(1,15) = 127.25$, $MSE = 6.38$, $p < .001$. Performance after learning the front view only or profile view only did not differ significantly ($F < 1$, *ns*).

Another common feature between the current Experiment and Experiment 10 is that the overall main effect of test view was also expected. However, the reasons behind this main effect are a little different to those that apply in Experiment 10. Just as in Experiment 10, when a single view is learnt (either the front view or the profile view) and testing takes place on the most extreme change (i.e. the front view is learnt and recognition is tested for the profile view and vice-versa), performance is poor. However, poor generalisation of hairstyles would mean that the three-quarter view would be poor, yielding accuracy levels close to those obtained after an extreme change in pose. Therefore, the average level of recognition on the front and profile views is actually higher (due to recognition accuracy obtained when the front and profile views are learnt) than that obtained on the unseen three-quarter view. Therefore, the three-quarter test view would be expected to produce *lower* levels of recognition than either the front or profile test views. It was indeed found that recognition of the three-quarter view was significantly poorer than either the front view or the profile views; $F(1,15) = 42.99$, $MSE = 20.93$, $p < .001$ whilst the front and three-quarter views did not differ significantly ($F < 1$, *ns*).

To analyse strong the interaction, a simple main effects analysis was conducted. Considering the poor performance obtained after the front view only was learnt and testing took place on the profile view and vice-versa, there was an expected effect of learning condition for both the front test view; ($F(2,30) = 188.60$, $MSE = 0.29$, $p < .001$) and profile test ($F(2,30) = 177.25$, $MSE = 0.27$, $p < .001$) views. There was no difference between the three learning types when tested with the previously unseen three-quarter view; $F(2,30) = 1.20$, $MSE = 0.99$, $p = .315$.

Again, due to the poor performance on the most extreme rotation after single view learning, there was a significantly effect of test view for faces learnt from the front view only ($F(2,30) = 41.90$, $MSE = 0.52$, $p < .001$) and from the profile view only ($F(2,30) = 75.95$, $MSE = 0.57$, $p < .001$). It was also found that recognition in the two view learning condition was also significantly affected by test view; $F(2,30) =$

35.26, *MSE* = 0.59, $p < .001$. Contrasts were conducted on a 1-way repeated measures ANOVA to determine the source of this result. It was found that when both the front view and profile views of a hairstyle were learnt, the three-quarter view was recognised significantly less well than the front or profile views; $F(1,15) = 45.97$, *MSE* = 5.40, $p < .001$. The front and profile views did not differ significantly ($F = 1$, *ns*).

### 5.4.3 Discussion

Experiment 11 investigated how well hairstyles can be recognised after a transformation of pose when learnt from either a single view or from two views of the hair. It was found that learning two views of a hairstyle did not lead to higher levels of accuracy on a previously unseen three-quarter test view than learning a single view (either the front view or the profile view). Indeed, the level of recognition accuracy on the three-quarter view in the current Experiment was much lower than that obtained in Experiments 3, 4, 5 and 10, indicating that generalisation of a hairstyle to novel views is poorer than that obtained with either the whole face (including hairstyle) or the face with the hair removed.

A strong possibility for the results found in Experiment 11 is that the level of information presented in the image is very limited. If, as suggested by Chapters Two, Three and Four, face recognition is mediated by pictorial codes over and above viewpoint-invariant and viewpoint-dependent codes, then the poor generalisation to novel views in the current Experiment is not surprising given the amount of information present in each image. However, the current Experiment does suffice to demonstrate that if participants in face learning experiments do choose to focus upon the hairstyle, reliance on a cue of such limited generalisation value is a suboptimal strategy.

It should be remembered however that the faces used in these experiments were chosen because the hairstyles were considered to be “typical” of the population. It is possible that some people will have a hairstyle (such as a Mohican) that will be recognisable due to its rarity. However, the same results may well occur if many faces with the same, albeit distinctive, hairstyle (such as a Mohican) were used.

## 5.5 General discussion of Experiments 9 – 11

Experiments 9 through 11 attempted to examine which of the cues present in a face are most useful for recognising newly learnt faces. It has been previously reported that the internal features of familiar (Ellis et al., 1979; Endo et al., 1984; Young et al., 1985) and newly learnt (Bonner et al., 2003; Clutterbuck & Johnston, 2002; Clutterbuck & Johnston, 2004; Clutterbuck & Johnston, 2005) faces are particularly useful for recognition. Unlike previous studies on newly learnt faces, Experiment 9 used a recognition task to assess familiarity. It was found that with this different task, the same pattern of results as previously reported emerged with the internal features of a face learnt from a single view proving more useful for recognition than the external features.

Expanding on this earlier work, Experiment 9 also demonstrated that this advantage for the internal features was also present when the face had to be recognised after a transformation of pose. This suggests that even after a change in the image, the internal features are still being used to aid recognition. However, the number of successful identifications in this condition is still significantly lower for the learnt image, indicating that the faces have not become sufficiently familiar that they display viewpoint-invariance. In contrast, participants who did not receive any training on the faces and only saw them once (leading to the face remaining unfamiliar), showed no preference for the internal features, with recognition accuracy being similar whether the internal or external features were given. Again this result is consistent with previous research.

Throughout Experiments 1 to 9 a number of participants were informally reporting after the experiment that the hairstyle was an important feature that they were using for recognition. It might, be thought, the case that the hair is a potentially misleading cue due to rotation of the hair being a difficult task to perform. Removal of the hairstyle would force participants to focus on other features of the face that might be more useful for recognition across poses. This was investigated in Experiment 10. It was found that recognition of the previously unseen three-quarter view was significantly better after two views of the face were learnt than when a single view was learnt. This result was in contrast to the results of Experiments 3, 4 and 5 which indicate that there is no benefit on the recognition of novel views after learning two views of a face over a single view. The only difference between Experiments 3, 4 and 5 and Experiment 10 is the presence of hairstyles in the face images. It therefore appears that forcing participants to use the internal features of a face for learning and recognition (through the removal of the hairstyle) can improve recognition performance after a change of pose.

If the hairstyle appears to be very different in full-face and profile views then this might remove any potential advantage to learning two views. Experiment 11 looked at this possibility by examining people's recognition of hairstyles on their own. It was found that recognition of the same image of a hairstyle was very good, but a change in pose had a significant impact on recognition accuracy. Levels of recognition accuracy for a previously unseen view were far lower in Experiment 11 than in the comparable Experiments (3, 4, 5 and 10) and sufficiently low (approximately 43% in the best case) to show that many hairstyles generalise poorly to novel views.

The results of Experiment 11 suggest that the hair, despite being potentially distinctive, isn't a particularly useful cue for recognising faces in novel views. This might also provide an explanation as to why recognition accuracy obtained after two view learning in Experiment 10 looks similar for all three test viewpoints. The hair, which may have been the primary cue used by participants in Experiments 3, 4 and 5, was removed in Experiment 10, forcing the participant to focus on other features of the face which are more reliable predictors of identity after a transformation of pose.

In conclusion, the three experiments reported in this Chapter have found that the internal features of a newly learnt face are more useful for recognition than the external features. The advantage applies even when a face is rotated into a novel pose. It is possible that the reason that people find it difficult to recognise faces from the external features is that the hair, whilst potentially a distinctive cue, is difficult to recognise after transformations of pose. Indeed, performance shows signs of improvements with the removal of the hair. Hence, whilst the internal features might be construed as a useful set of features for recognition, the hair can be a somewhat misleading cue to rely on to learn and recognise new faces.

# CHAPTER SIX

Conclusions and future directions

## 6.1 Recognising unfamiliar and familiar faces

Unfamiliar face recognition, whilst originally thought to be good (Hochberg & Galper, 1967; Galper & Hochberg, 1971) has been shown by subsequent research to be poor, with recognition accuracy and matching performance far lower than that for familiar faces. As such, the mental representation for unfamiliar faces is fragile. Unfamiliar face recognition seems to be highly image-specific (e.g. Liu & Chaudhuri, 2000) and any change in the image between the study and test phases of an experiment leads to a decrease in performance. Familiar face recognition, on the other hand, is much more robust such that they can be recognised despite changes in the image between study and test. Familiar faces can be recognised from poor quality images (Burton et al., 1999), after changes in colour due to hue negation (Kemp et al., 1996), and after changes in pose (Bruce, 1982), size (Brooks et al., 2002) and expression (Bruce, 1982).

The difference in performance between familiar and unfamiliar faces on recognition tasks has often thought to be due to the nature of the representation of the two types of faces. The recognition failures seen with unfamiliar faces when different images are used for learning and testing have been taken as an indication that unfamiliar face recognition is image dependent and indeed, it has been suggested that the greater the difference between the study and test images, the worse recognition becomes (Bruce, 1982). In stark contrast, familiar faces are recognised well from novel images. It has been suggested that the reason familiar faces are recognised well is that over time, a three-dimensional viewpoint-invariant structural model of the face is constructed (Bruce, 1982; Bruce & Young, 1986). During the recognition process, this model can be rotated or scaled to enable successful recognition despite image changes. The processes underlying familiar and unfamiliar face recognition are therefore hypothesised to be fundamentally different.

Despite the differences in the recognition of unfamiliar and familiar faces, the process by which individual faces become familiar has often been ignored. In part, this has been due to the difficulties in simulating the experiences we have with faces in everyday life in the laboratory. Typically, laboratory based investigations into face recognition have presented participants with a single photograph of a face to study and tested their recognition for either the same photograph or the face after a transformation (e.g. pose or lighting) has been applied. However, despite their critical role in laboratory studies, how faces become familiar through the use of photographs has not been extensively studied. Photographs are static two-dimensional stimuli whereas the faces we interact with everyday are dynamic three-dimensional objects. Indeed, it is unclear whether faces learnt from photographs can become truly familiar so that they exhibit invariance across various transformations. This thesis, over 11 experiments, therefore set out to examine how faces are recognised after they have been learnt from photographs. The experimental work was split into two sections. The first section (Chapters Two, Three and Four) describe work aimed at determining whether faces learnt from photographs are recognised via a viewpoint-invariant structural model or from an image-based approach such as viewpoint-dependent coding or pictorial coding. The second part (Chapter Five) looks at which identity cues present in a face image are used as the face becomes more familiar.

## 6.2 Summary of the main findings

Chapter Two reported a series of experiments that were designed to examine the extent to which faces learnt from either a single photograph (Experiments 1 and 2) or two photographs (Experiments 3 – 5) can be recognised after a transformation of lighting (Experiment 1) or pose (Experiments 1 – 5).

Typically, much of the previous research into face recognition has given a single exposure to the face during the learning task. Familiar faces have, however, been seen on more than one occasion. It is possible that a single exposure is not sufficient for the construction of a robust enough representation of the face to allow for successful recognition across changes in pose and lighting. Experiment 1 investigated whether providing multiple exposures during learning in comparison to a single exposure enables the face to be learnt sufficiently well that recognition is possible after a transformation of lighting or pose.

If multiple exposures to faces do aid their recognition, then the overall performance would be expected to be better than performance obtained after receiving a single exposure to the face. This was indeed found in Experiment 1. It was also found that, after seeing the face only once, recognition accuracy after a transformation of pose or lighting dropped significantly compared to the accuracy obtained for the recognition of the studied image. This result is consistent with previous findings which demonstrate the image dependence of unfamiliar face recognition. Performance after a transformation of pose or lighting was also significantly lower than the performance on the studied image when the face had been learnt from multiple exposures. The drop in performance after a transformation of pose or lighting was similar for both the single and multiple exposure groups, suggesting that, despite the increased learning provided by multiple exposures to a face image, the same process was being used for recognition, regardless of whether the face was seen once or many times.

The results of Experiment 1 indicate that even after an image of a face has been learnt to a very high level (participants who learnt faces from multiple exposures scored nearly 100% correct during the recognition task for the studied image), recognition is not invariant across transformations of lighting and pose. This implies that in these circumstances, recognition is mediated via an image-based process such as viewpoint-dependent codes or pictorial codes, as suggested by image-based models of recognition, over and above any three-dimensional viewpoint-invariant structural model.

There are two possible mechanisms by which an image-based system of face recognition may operate; viewpoint-dependent coding and pictorial coding. The results obtained after the lighting change used in Experiment 1 can help identify which of these two mechanisms is being used. Changing the lighting direction of a face between the study and test phases of an experiment keeps the viewpoint constant but alters the image. If recognition is achieved via the use of viewpoint-dependent codes (which enable recognition within a set viewpoint), recognition accuracy after a change in lighting should be no different to that obtained when testing the recognition of the studied image, as the viewpoint is the same in both images. A drop in accuracy would however indicate that recognition is being affected because the image has changed, suggesting that pictorial codes are used for recognition. Experiment 1 found that when a lighting change was made, performance dropped significantly, suggesting that recognition was being mediated by pictorial codes.

Despite the indications that recognition after such changes is mediated via image-based codes, and in particular pictorial codes, the use of a viewpoint-invariant structural model as the means of recognition cannot be ruled out. This is because when testing takes place with the studied image, a combination of both viewpoint-invariant and pictorial codes may be used for recognition (as the picture is identical to the learnt image), having the effect of boosting performance. After a transformation of pose, only the viewpoint-invariant model is available to the recognition process. In Experiment 1, this model will have been created from only a single view; it is thus likely to be impoverished and inaccurate. When recognition must be achieved using such a model on its own (such as after a transformation of pose), performance will suffer due to the model's inaccuracy.

Experiment 2 extended Experiment 1 to investigate the possibility that the reduction in accuracy after a change in pose is due to the use of an impoverished three-dimensional structural model learnt from a single image as opposed to image-based codes. An impoverished three-dimensional structural model will contain errors that will be present in all viewpoints. Therefore, if recognition is mediated via a structural model then performance will be similar across all angles of rotation. On the other hand, recognition via the use of image-based codes will produce a generalisation gradient in which recognition decreases the greater the difference between the study and test images becomes. This difference will be expected to increase the further the face is rotated away from the learnt view. Experiment 2 used different amounts of angular change between the study view and test view to examine recognition accuracy after increasing differences between the study and test images. It was found, just as in Experiment 1, that recognition accuracy for the learnt view was very good. The further the face was rotated away from the learnt view, however, the poorer recognition accuracy became. The results indicate that recognition accuracy is a function of the rotation of the face away from the learnt view. This suggests that recognition is mediated via the use of image-based codes rather than a three-dimensional viewpoint-invariant structural model.

The results of Experiments 1 and 2 suggest that image-based codes, and particularly pictorial codes, are used for the recognition of faces learnt from photographs. However, in both experiments, the face was learnt from a single view. It is possible that learning the face from only one view limits the construction of a three-dimensional model. Experiments 3, 4 and 5 examined if learning two different views of a face produces any recognition advantage on the recognition of a novel view of a face over learning a single view. Experiment 3 found that if participants learnt two views of a face (the full-face view and the profile view), recognition of a novel view (a three-quarter view) was not statistically better than if the full-face view only or the profile view only had been learnt. Experiment 3 also revealed that the number of exposures required to learn each image presented of a face was approximately the same, regardless of whether the face was being learnt from one or two views. This result suggests that image-based cues are being used for recognition and does not support the idea of the creation of a three-dimensional viewpoint-invariant structural model.

During the training phase of Experiment 3, faces that were to be learnt from two views had the same name given to each of the views. Therefore, participants had ample opportunity to tie the two images together to construct a single structural model of the face, as they received confirmation throughout training that the two views depicted the same individual. To provide an indication as to whether participants were treating the two views as separate people, Experiment 4 added a further condition in which two views of a face could be learnt with either the same name or different names applied to each view. If participants combined the two views of an individual into a single structural model that representing that individual, then learning two views of a face with different names applied to each view would lead to confusion during the learning phase, and therefore require more exposures to each view, than if both views are given the same name. Experiment 4 revealed that, in contrast, participants required approximately the same number of exposures to each view of a face to learn that view, regardless of whether they learnt the faces either from the full-face view only, the profile view only, both views with the same name applied to each view, or both views with different names applied to each view. This indicated that people have no extra difficulty in learning two views of the same face when different names are given to each view.

Two potential issues were identified with Experiments 3 and 4. Firstly, 12 participants were used in both experiments which could have the effect of limiting the power of the experiments to detect any benefit of learning two views of a face. Secondly, many of the data points in both experiments were close to ceiling, possibly masking any effects present. Experiment 5 addressed these two issues by firstly increasing the number of participants to increase the power of the experiment. In order to deal with the ceiling effects, latency data were analysed as reaction times can be a more sensitive measure when the data are close to ceiling.

The accuracy data from Experiment 5 produced a similar result to that obtained in Experiments 3 and 4. Again it was found that recognition accuracy on a previously unseen three-quarter view was the same, statistically, regardless of whether the face was learnt from the full-face view only, the profile view only, both views with the same name applied to each view, or both views with different names applied to each view.

To analyse the latency data, it was necessary to combine the full-face only and profile only learning conditions into a new, single view learning condition, in which response times for the full-face view testing condition were taken from when the full-face view was learnt and response times for the profile view testing condition were taken from when the profile view was learnt (i.e. the views that were learnt during training). To determine the response times to be used for the unseen three-quarter view testing condition, the data from the learning condition which yielded the highest accuracy on this test view were used. These data came from the full-face only learning condition. It was found that response times to the critical previously unseen three-quarter view at test were significantly longer than to either the full-face or profile views. Furthermore, it was also found that response times across the three test views did not differ whether a single view or two views of the face were learnt. It seems that during the recognition task, participants primarily rely upon one of the learnt views to perform recognition. This provides further support to the idea that in face learning experiments in which the face is learnt from a photograph, participants learn something about the picture, not the face itself.

To summarise, Experiments 1 to 5 in Chapter Two suggest that when faces are learnt from photographs, people readily learn something about the image but do not easily acquire invariant information about the face. It seems that a structural three-dimensional viewpoint-invariant model is not formed and used; instead recognition is mediated by image-based codes which may be either viewpoint-dependent or pictorially-dependent.

Chapter Three reported two experiments which were designed to disentangle whether viewpoint-dependent codes or pictorial codes are used for recognition. Performance after a change in lighting direction, as used in Experiment 1, suggests that recognition is more likely to be achieved via pictorial codes than viewpoint-dependent codes. Experiment 6 extended Experiment 1 in two ways. Firstly, recognition accuracy after three levels of lighting change was compared instead of the single change used in Experiment 1. The extra lighting changes used in Experiment 6 allowed for an examination of recognition accuracy after increasing amounts of lighting change, to see if performance decreases as the lighting change becomes greater. As the lighting change between the learnt image and the test image increases, the difference between the two images increases. If pictorial codes are used for the recognition of faces then as the difference between the images increases, the poorer performance will become. It was indeed found that recognition accuracy decreased the greater the lighting change was. This is an important result in attempting to separate viewpoint-dependent coding and pictorial coding as changing the lighting direction of a face between study and test changes the image yet leaves the viewpoint of the face unchanged. Therefore, it appears that the recognition of faces learnt from photographs is mediated by pictorial codes over and above viewpoint-dependent codes.

The second issue addressed in Experiment 6 was whether combining two transformations (a pose change and a lighting change) together yielded poorer performance than a single transformation. Bruce (1982) suggested that the key component to successful recognition using pictorial codes is how similar two images are, not the type of change through which the face has gone. Thus it would be expected that making two changes to an image between study and test would reduce recognition accuracy to a greater extent than if only one change was made. It was found that making a pose change and a lighting change together produced poorer performance than if either change was made on its own.

Experiment 7 used another transformation to further investigate the relative roles of viewpoint-dependent and viewpoint-invariant codes; the transformation of size. It is possible that changing the lighting direction of a face between study and test (as used in Experiments 1 and 6) might not only alter the image, but could also potentially alter the viewpoint. This is because features that were present during the learning phase may no longer be visible during the test phase, producing what could be considered a new viewpoint. The transformation of size leaves not only the viewpoint unchanged whilst altering the image properties, but also leaves the features of the face visible after the transformation. The results of Experiment 7 indicated that size did adversely affect the recognition of faces learnt from photographs. This suggests, in line with the results of Experiments 1 and 6, that the recognition of faces is mediated by pictorial codes rather than viewpoint-dependent codes.

Chapter Five changed the focus from an examination of the roles of three-dimensional viewpoint-invariant structural model and image-based model approaches to face recognition to look at changes in the way different parts of the face are processed as they become more familiar. Previous research has demonstrated that the internal features of familiar faces are processed more accurately (Ellis et al., 1979) and quickly (Young et al., 1985) than the external features. A possible explanation for this difference is that for unfamiliar faces, the most distinctive feature (and therefore most accessible) of the face is used for recognition. Many of the participants in Experiments 1 to 8 informally commented that they were using the hairstyle (an external feature) to help them learn the faces as it was the feature that "stuck out" most to them. Hairstyles change regularly though and so would not be a useful cue to use for people we know well and see often. Hence, the internal features might constitute a more reliable set of features with which to perform a recognition task for familiar faces. The experiments in Chapter Five therefore examined not only the relative usefulness of the internal and external features for the recognition of newly learnt faces, but also how useful a cue the hairstyle is for recognition.

Experiment 9 extended previous work by Bonner, Burton and Bruce (2003) who demonstrated that faces learnt and tested using a matching paradigm show an internal feature advantage after three days of training with the faces. In Experiment 9, participants either received a single exposure to a face (and thus the faces were still relatively unfamiliar to them) or were trained on a series of faces. Both groups were then tested on their recognition for the whole facial image, the internal features only and external features only for both the same pose as originally studied and also after a pose change of 31°. It was found that participants who received only a single exposure to a face demonstrated no difference in terns of recognition accuracy between the internal and external features. Those who received training did however exhibit an advantage at test for the internal features. This internal feature advantage was only found in the same pose condition, suggesting that it arises within viewpoints rather than across a general representation of the face. This is the first time such an advantage has been reported after such a short training procedure (on average the whole experiment took approximately 20 minutes).

Throughout Experiments 1 to 8 a number of participants informally reported after completing the experiment that they found the hairstyle to be of use to complete the training task. However, Experiment 9 suggests that the external features, of which the hair forms a major component, may not be particularly useful for the recognition of faces learnt from a single view. Experiment 10 investigated whether removing the hair from the photographs might increase performance, as this would force participants to use other features (particularly the internal features) to learn and recognise faces. The experiment was identical to Experiment 3 (in which participants learnt faces in the full-face view only, the profile view only and both full-face and profile views and recognition accuracy was tested for the full-face view, profile view and a previously unseen three-quarter view) except the photographs of the faces were altered to remove the hair. For single view learning, performance levels in Experiment 10 were similar to Experiment 3. However, generalisation to the unseen three-quarter view after two views had been learnt was notably higher than in Experiment 3. Furthermore, in Experiment 10, learning two views of a face led to significantly better performance than if only the full-face view or profile view was learnt. It appears the people are better at using two images together to generalise to a novel view when the hair, a potentially distracting feature, is removed.

Experiment 11 set out to examine further the role of the hair in face recognition. If the hair has poor generalisation to novel views then if the hairstyle alone is learnt, recognition after a pose change should fall dramatically. Experiment 11 found that this was indeed the case, with good levels of performance (over 90% accuracy) when the same hairstyle was used at both study and test, but recognition falling down to approximately 30% on the unseen three-quarter view after the full-face view or profile only views of a hairstyle were learnt. Performance after both views were learnt on the three-quarter test view was approximately 42%. This higher level of performance was not statistically significant however. Overall, it appears that relying primarily on the hairstyle to learn new faces is a suboptimal strategy to employ as the recognition of a rotated hairstyle is difficult, reducing the chances of successful recognition after a pose change.

## 6.3 How do faces become familiar?

#### 6.3.1.1 Face learning and recognition from photographs in laboratory experiments

Familiar and unfamiliar faces have been shown to exhibit very different levels of performance on both recognition and matching tasks. Familiar face recognition is very robust and familiar faces are recognised well despite changes in pose, expression, poor image quality and hue. Unfamiliar face recognition, on the other hand, is error prone after such changes and even matching two different images of an unfamiliar face is difficult.

An important theoretical question is why familiar and unfamiliar faces yield such different levels of recognition accuracy and matching performance. After all, all faces, whether they are familiar or unfamiliar, are the same class of object (i.e. approximately round in shape, two eyes above a nose which, in turn, is above a mouth). It therefore seems strange that people's ability with the two types of face is so different. The most prominent theory on why unfamiliar and familiar faces differ so much with regard to how well they are recognised relates to the nature of the internal representation of familiar and unfamiliar faces.

Theories of object recognition have suggested that objects may be recognised via a viewpoint-invariant structural representation of the object (Biederman, 1987; Marr, 1982; Marr & Nishihara, 1978). This representation, built up over time with experience of the object, can be rotated to any new viewpoint to allow for successful recognition. Bruce and Young (1986) proposed that such a viewpoint-invariant model (termed a structural code) is formed for a familiar face through repeated experience of that face. It is this structural code provides invariant recognition across natural changes (such as pose or lighting) (Bruce, 1982).

In contrast, unfamiliar face recognition is much more fragile. Changing the image through a transformation of pose (Bruce, 1982; Krouse, 1981; Hill et al., 1997), lighting (Braje et al., 1996; Braje et al., 1998; Braje, 2003; Hill & Bruce, 1996; Liu et al., 1999) or expression significantly decreases recognition accuracy. Bruce (1982) and Bruce and Young (1986) argue that this fall in performance is due to the nature of the coding used for unfamiliar faces. They suggest that the recognition of unfamiliar faces is mediated via pictorial codes that are specific to the image studied. As long as the presented image is sufficiently similar to a stored pictorial code, then recognition will be possible. If the image is sufficiently different, however, recognition becomes much more error prone and difficult.

The predominant view of familiar face recognition is that in order for a face to become familiar, a structural representation of the face must be formed. However, both the work presented in this thesis and previous research suggest that we need to look harder at whether a three-dimensional viewpoint-invariant structural model is required for familiar face recognition. For example, there have been reports of viewpoint-dependence with familiar faces (Bruce et al., 1987; Troje & Kersten, 1999). Results that suggest viewpoint-dependence for familiar faces are at odds with a structural account of familiar face recognition. In fact, a closer examination of some of the properties of familiar faces suggests that viewpoint-dependent codes are a sufficient explanation for the robustness of familiar face recognition. Familiar faces have not only been seen a number of times (i.e. more than a single exposure) but they have also been seen from a number of different views. It is therefore plausible that the reason familiar faces demonstrate invariance over changes in, for example, pose and lighting is that we have previously seen, and remembered, familiar faces in a number of different poses and lighting conditions. As long as the currently presented image of a known face is sufficiently similar to a stored pictorial or viewpoint representation of the face, then recognition and identification of that face will be possible. The newly encountered image of the face can then be stored for future reference.

In the object recognition literature, image-based models (Bulthoff et al., 1995; Gauthier & Tarr, 1997; Poggio & Edelman, 1990) have been suggested as an alternative to viewpoint-invariant models for recognition and can explain the viewpoint-dependency seen in some studies of familiar face recognition. These models predict that recognition accuracy will be highest for the view that is originally studied with performance falling the greater the difference is between the studied view of an object and the test view. For totally invariant recognition, image-based models require a sufficient number of views of the object (or face) to enable it to be recognised in a novel view. If the differences between the stored representation and the presented image are small enough then the object or face will be recognised.

Hill et al. (1997) demonstrated that, for unfamiliar faces, the greater the angular change was between the studied image and the test image, the lower performance became.

In an effort to understand not only the nature of familiar and unfamiliar face recognition but also the process through which faces become familiar, the predominant method of investigation has been the use of photographs of faces. Photographs are inherently different to the faces we experience in everyday life – they are two-dimensional and static as opposed to the three-dimensional, animated objects we interact with. Therefore, there is a strong need to understand how people learn and recognise faces from photographs.

This thesis examined the recognition of faces learnt from photographs. In an important modification of previous research, participants were required to learn the face images to a high level. Much of the earlier work into face learning has provided only a single exposure to a previously unfamiliar face during a study phase. The problem with such a technique is that it is possible that a single exposure does not provide sufficient experience with the face for a robust representation of the face (either structural or image-based) to be formed. It may not be considered surprising therefore that many studies have obtained results consistent with a viewpoint-dependence or pictorial dependence model of unfamiliar face recognition.

Experiments 1 to 8 provided participants with multiple exposures to individual faces in order to give the participants an opportunity to form a structural code. Despite the fact that recognition accuracy of the studied view of the face (i.e. when testing occurred with the same picture as originally studied) was high (approaching 100% accuracy in most of the experiments), performance after the image was changed by a transformation of pose, lighting or size indicated that the primary method of recognition used by participants was via the use of pictorial codes. These codes are highly image-specific and may contain little information about the face itself. Instead, a pictorial code can be based on other information about the image, such as a mark on the photograph.

#### 6.3.1.2 Face learning and recognition in everyday life

It would appear that when faces are learnt from photographs, pictorial codes specific to the picture studied play a critical role in the recognition process. However, in everyday life, faces are learnt from encounters with people in which many views of their face are seen, as opposed to specific images. Therefore, it is unlikely that pictorial codes play a major role in recognition in everyday, real life interactions. However, the critical role of pictorial codes in the recognition of faces learnt from photographs may mimic a real world face recognition mechanism which may be instance-based. That is, instead of storing a series of particular images of people, a snapshot, relating to a particular instance in which a face is encountered, is stored. Such snapshots will not contain some of the details of a pictorial code (e.g. marks on a photograph would not be stored) but will be specific to a particular instance (e.g. they will be specific to viewpoint, lighting direction, hairstyle etc).

If such snapshots are used for familiar face recognition, then it follows that in order for robust recognition, a sufficiently large number of instances of the face have to be stored. The number of viewpoints needed for successful recognition needs to be sufficiently high that the difference between the stored representation and a snapshot that can be created from the current instance of the face is small. The common consensus is that at least three viewpoints are required; the full-face view, the three-quarter view and the profile view. The number of instances *within* each viewpoint needed for recognition has two interesting alternatives. Firstly, as has been suggested by the work presented in this thesis, a number of discrete instances of each individual known to us need to be stored. As long as the current instance of a face is sufficiently similar to a stored representation, then successful recognition is possible. Once recognition is achieved and the identity of the individual confirmed, then the instance seen during the recognition process can be stored as a new snapshot of that person.

An alternative to such an approach has been proposed by Burton et al. (2005). This approach proposes that an average image for each viewpoint is calculated from the seen exemplars. In both tests on human participants and on computer simulations run using PCA, Burton et al. consistently found that the greater the number of exemplars that made up the average image, the better the accuracy of the PCA system and the faster the recognition by human participants. Such an approach is appealing as during the process of averaging facial images, lighting and contrast changes, as well as slight variations in pose are eliminated. For example, consider two images of a face that are illuminated from opposite directions. The average of these two images will cancel out much of the effect of the change in illumination direction. In addition, small changes in pose are also cancelled, out so that only one average image of the face needs to be stored per viewpoint.

Despite the appeal of a system that functions on the average of a series of facial images, the results of the experiments in this thesis support a face recognition system that is instance-based which operates on exemplars of faces created during social encounters. However, it must be remember that all experiments reported here only provide participants with a most, one image per viewpoint. Therefore there is no opportunity for the formation of an average face image. To repeat the experiments conducted by Burton et al. (2005) with unfamiliar faces, a database of unfamiliar faces which depicts faces over a period of time (approximately 5 years), with the images taken with different cameras under different lighting conditions would be needed. A number of images per person would be required (the maximum number of images used per person in Burton et al.'s study was 20). Unfortunately, such a database, whilst possible to create through the use of family photo albums, does not exist today.

## 6.4 Practical applications of the current research

The work presented in this thesis has implications for current applications of face learning and recognition from photographs. In particular, photographs are used extensively by the police force in appeals to the public for help in searching for an individual. Experiments 1 to 8 suggest that providing only a single photograph of an unfamiliar individual is not an optimal strategy. Even a slight change in appearance, which will happen naturally through changes in expression, lighting and so on, will result in difficulties for people unfamiliar with the person to identify them. It is therefore preferable, if possible, to present multiple images of the individual to maximise the general public's chances of identifying the person. Of course, a potential problem with this idea is that people may see the different images as different people. However, it appears that providing a single image of a person makes it difficult for someone unknown to recognise him or her in any other situation other than from a near-identical image.

The experiments in Chapter Five do, however, suggest that people can focus upon features of the face that may be considered more stable over transformations such as pose (the internal features of the face) and that this can occur after training with a limited number of views. It is therefore not impossible that police officers could be trained with photographs of an individual to heighten their chances of apprehending a suspect.

## 6.5 Future directions

The current research has identified two key areas for future research. Firstly, to examine the possibility of the face recognition system employing the use of a prototype face image, the recognition of faces learnt from multiple images within the same viewpoint (e.g. two or more different images depicting the full-face view of an individual) could be investigated. Also, photographs are a two-dimensional stimulus. Research using three-dimensional stimuli might reveal new aspects of face learning.

### 6.5.1 Learning faces from a number of separate images

Burton et al. (2005) suggest that familiar face recognition is achieved, within a number of discrete viewpoints, via the use of an average image of a face created from multiple images. Previous research into face learning (and the experiments presented in this thesis) primarily use a very limited number of images for learning. For example, all experiments in this thesis have only presented a single image within each viewpoint during learning. Therefore, the average image, if it is indeed used for recognition, remains simply the image that was learnt, leading to the emergence of pictorial codes for recognition.

A first step in looking into whether an average is used for recognition would be to train participants to recognise two images that depicted a face in the same viewpoint (e.g. the full-face view) and two different lighting conditions, for example profile view lighting and three-quarter lighting (the two rightmost images in Figure 6-1) and comparing recognition accuracy for these two (previously seen) images against the face in a novel, unseen, lighting condition; full-face view lighting (the leftmost image in Figure 6-1). However, a potential problem with this approach is that the face could be recognised either via an average of the two images or from one particular part of one or both of the images, leaving the roles of pictorial coding and viewpoint-dependent average image coding unresolved.

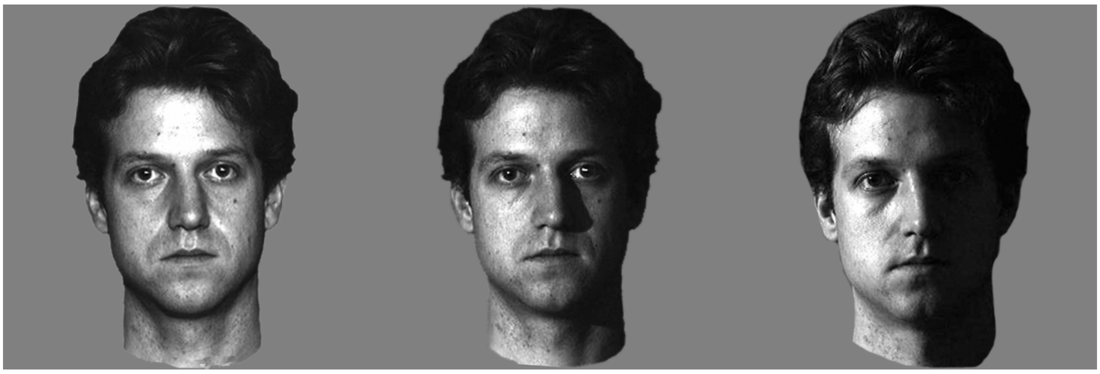

**Figure 6-1: The same face from three different lighting conditions (lighting direction from left to right; full-face, three-quarter and profile).**

A clearer distinction as to whether recognition is achieved using the pictorial codes from particular exemplars or an average of all the training images could be achieved by training participants to recognise a series of different images of the same person within the same viewpoint with the images taken at different times, which in turn, results in different looking images. After training, participants' recognition of a third, novel image (presented in the same pose as the training images) could be tested. If recognition is achieved via the use of pictorial codes, recognition will be hampered by the use of a different looking image at test and the fact that two images were used for training would not be a great advantage over using a single image. However, an average image of the face, created from two views of the face, would be recognised better than if only a single image was learnt (which would imply that the average would in fact be the studied image!) as the average would hold more similarities with the novel test image than the training images.

### 6.5.2 Two-dimensional versus three-dimensional images

One criticism that can be levelled at research into face learning that employs the use of photographs is that photographs are a two-dimensional stimulus whereas faces are three-dimensional. Photographs, even whilst they depict a face, may not be learnt in the same way as three-dimensional objects. It may even be argued that it is not surprising that people, when asked to learn faces from photographs, appear to learn the photograph over and above the face itself. Conversely, criticism can also be aimed at studies in which participants interact with live actors. Such experiments can suffer from large amounts of variation in participants' experience with the actor, so that the level of control in the experiment is low in comparison to a typical laboratory study.

A way of enabling participants to interact with three-dimensional faces whilst keeping the experience for all participants the same is to use virtual reality (VR).

The use of stereoscopic displays allows faces to appear three-dimensional whilst they are displayed on a two-dimensional computer screen. Computer software exists that allows the face to be rotated on the screen and change the lighting direction in real time (Ward & Liu, 2005). Early results in this field have suggested that stereoscopic information may not aid facial recognition (Liu, Ward, & Young, 2006) but it may be that three-dimensional cues are used in certain situations (Liu & Ward, 2005; Liu & Ward, 2006).

The results of the experiments in Chapter Five present a caution for work using VR. Experiments 9 to 11 highlight the role that hair plays in the learning and recognition of new faces. The faces used in VR systems must comprise a three-dimensional structural model, typically created by laser scanning a real human head. Whilst these laser-scanning systems can produce very accurate models of human faces, they cope much less well with the detail presented in hair. Therefore, the face models used in VR experiments normally comprise a "mask" that is devoid of any information about the hair (see Figure 6-2). As demonstrated by Experiments 9 to 11, the hairstyle might be an important cue during the early stages of face familiarisation and the effect of losing that hairstyle as a result of laser-scanning needs to be considered in experimental design.

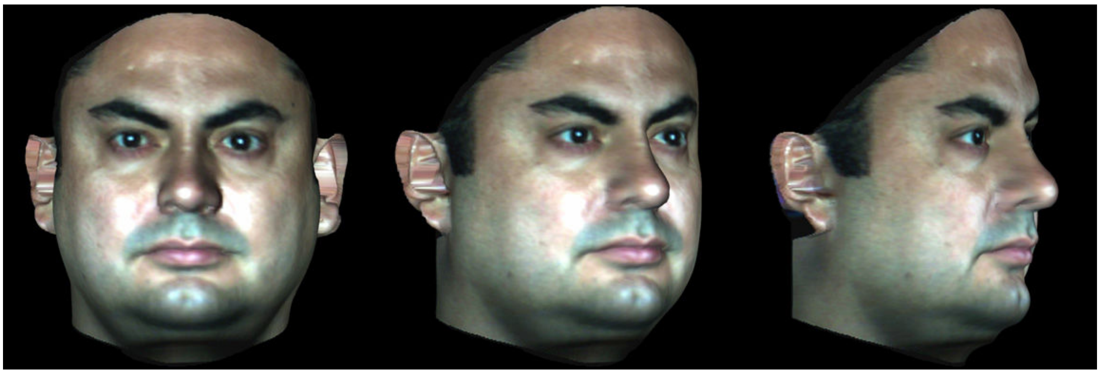

**Figure 6-2: Example of the types of faces that are created using laser-scanning techniques. The images were rendered from a database of models obtained from the University of South Florida.**

## 6.6 Conclusions

Research into face learning and recognition has heavily employed the use of photographs. Typically, only a single exposure to a face is used during face learning studies, which limits the possibility of forming a robust representation of the face. This thesis adds to existing knowledge by examining how faces are recognised after the image of the face has been learnt very well and demonstrates that changing the image when testing recognition of the face leads to a significant decrease in performance. The lack of invariance appears not to be due to the lack of the formation of a viewpoint-invariant representation of the face, but rather the lack of experience with a sufficient number of images of the face. It seems that the recognition of faces learnt from photographs is image-based. The current work also shows that recognition of faces learnt from a single photograph, in a particular viewpoint, is likely to be mediated by pictorial codes rather than viewpoint dependent codes. The work presented in this thesis also demonstrates that colour consistency is not required for recognition, suggesting that a low-level image intensity map is sufficient for the recognition process. Additionally, the present work also demonstrates, for the first time, that the internal feature advantage typically associated with familiar faces can occur in a recognition task after faces have been learnt from a single photograph, suggesting that under certain circumstances familiarity with faces can be obtained in the laboratory using photographs in a short space of time. Finally, the current work identified several new directions of research. To successfully simulate the real world face learning experience in the laboratory, more images than a single view per viewpoint could be used. Also, virtual reality has been identified as a potential tool for overcoming the limitations imposed by photographs, a two-dimensional stimulus, for learning a three-dimensional stimulus – the human face.

## References


Adini, Y., Moses, Y., & Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*, 721-732

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 997-1010

Angeli, A., Bruce, V., & Ellis, H. D. (1998). Getting familiar with faces. *Perception 28*, Suppl. 115.

Baddeley, A., & Woodhead, M. (1983). Improving face recognition. In S. M. A. Lloyd-Bostock, & B. R. Clifford (eds) *Evaluating eyewitness evidence*. UK: John Wiley and Sons

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*, 379-384

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

Biederman, I. (1987). Recognition-by-Components - A theory of human image understanding. *Psychological Review, 94*, 115-147

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39*, 2885-2899

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects - evidence and conditions for 3-dimensional viewpoint invariance. *Journal of*

*Experimental Psychology: Human Perception and Performance, 19*, 1162-1182

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

Braje, W. L. (2003). Illumination encoding in face recognition: effect of position shift. *Journal of Vision, 3*, 161-170

Braje, W. L., Kersten, D., Tarr, M. J., & Troje, N. F. (1996). Illumination and shadows influence face recognition. *Investigative Ophthalmology & Visual Science, 37*, 839

Braje, W. L., Kersten, D., Tarr, M. J., & Troje, N. F. (1998). Illumination effects in face recognition. *Psychobiology, 26*, 371-380

Brooks, B. E., Rosielle, L. J., & Cooper, E. E. (2002). The priming of face recognition after metric transformations. *Perception, 31*, 297-313

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

Bruce, V. (1983). Recognizing faces. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 302*, 423-436

Bruce, V. (1995). Perceiving and recognizing faces. I. Roth, & V. Bruce *Perception and representation*. Buckingham: Open Unviersity Press.

Bruce, V., & Burton, M. (2002). Learning new faces. In M. E. Fahle, & T. E. Poggio (eds) *Perceptual learning*. Cambridge, MA, US: MIT Press

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

Bruce, V., & Valentine, T. (1986). Semantic priming of familiar faces. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 38*, 125-150

Bruce, V., & Valentine, T. (1988). When a nod's as good as a wink: The role of dynamic information in face recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (eds) *Practical aspects of memory: Current research and issues. Vol. 1*. Chichester: John Wiley & Sons

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

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

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38*, 109-144

Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of mass in line drawings of faces. *Applied Cognitive Psychology, 6*, 619-628

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

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

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

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

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89*, 60-64

Bulthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are 3-dimensional objects represented in the brain. *Cerebral Cortex, 5*, 247-260

Burton, A. M. (1994). Learning new faces in an interactive activation and competition model. *Visual Cognition, 1*, 313-348

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

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

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

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

Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: extension and limits. *Memory & Cognition, 27*, 139-151

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

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

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory & Cognition, 26*, 780-790

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception, 31*, 985-994

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition, 11*, 857-869

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17*, 97-116

Cohen, J. (1988). *Statistical power for the behavioural sciences* (2nd ed). Hillsdale, NJ: Erlbaum

Davies, G., & Thasen, S. (2000). Closed-circuit television: How effective an identification aid? *British Journal of Psychology, 91*, 411-426

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

Dehon, H., & Bredart, S. (2001). An 'other-Race' effect in age estimation from faces. *Perception, 30*, 1107-1113

Diamond, R., & Carey, S. (1986). Why faces are and are not special - an effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117

Ellis, H. D., & Florence, M. (1990). Bodamer's (1947) paper on prosopagnosia. *Cognitive Neuropsychology, 7*, 81-105.

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

Endo, M., Takahashi, K., & Maruyama, K. (1984). Effects of observer's attitude on the familiarity of faces: Using the difference in cue value between central and peripheral facial elements as an index of familiarity. *Tohoku Psychologica Folia, 43*, 23-34

Galper, R. E., & Hochberg, J. (1971). Recognition memory for photographs of faces. *American Journal of Psychology, 84*, 351-354

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*, 1673-1682

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

Harries, M. H., Perrett, D. I., & Lavender, A. (1991). Preferential inspection of views of 3-d model heads. *Perception, 20*, 669-680

Hay, D. C., & Young, A. W. (1982). The human face. A. W. Ellis *Normality and pathology in cognitive functions* . London: Academic Press.

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception, 15*, 595-602

Hayward, W. G. (2003). After the viewpoint debate: where next in object recognition? *Trends in Cognitive Sciences, 7*, 425-427

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1511-1521

Hayward, W. G., & Tarr, M. J. (2000). Differing views on views: Comments on Biederman and Bar (1999). *Vision Research, 40*, 3895-3899

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science, 11*, 7-12

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

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance,* *22*, 986-1004

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201-222

Hochberg, J., & Galper, R. E. (1967). Recognition of faces I: An exploratory study. *Psychonomic Science, 9*, 619-620

Hosie, J. A., & Milne, A. B. (1996). The effect of experimental design on memory for typical and distinctive faces. *Memory, 4*, 175-197

Hsu, T. C., & Feldt, L. S. (1969). The effect of limitations of the number of criterion score values on the significance level of the F-test. *American Educational Research Journal, 6*, 515-527

Inn, D., Walden, K. J., & Solso, R. L. (1993). Facial prototype formation in children. *Bulletin of the Psychonomic Society, 31*, 197-200

Jeffery, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. *Psychological Science, 17*, 501-505

Johansson, G. (1973). Visual-perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201-211

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1-19

Johnston, A., Hill, H., & Carman, N. (1992). Recognizing faces - effects of lighting direction, inversion, and brightness reversal. *Perception, 21*, 365-375

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition , 13*, 289-303

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

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

Knapp, B. R., Nosofsky, R. M., & Busey, T. A. (2006). Recognizing distinctive faces: A hybrid-similarity exemplar model account. *Memory & Cognition, 34*, 877-889.

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition, 4*, 265-273

Kolers, P. A., Duchnicky, R. L., & Sundstroem, G. (1985). Size in the visual processing of faces and words. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 726-751

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology, 66*, 651-654

Laeng, B., & Rouw, R. (2001). Canonical views of faces and the cerebral hemispheres. *Laterality, 6*, 193-224

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: evidence from picture matching. *Journal of Experimental Psychology:*

*Human Perception and Performance, 22*, 395-416

Lewis, M. B. (2001). The lady's not for turning: Rotation of the Thatcher illusion. *Perception, 30*, 769-774

Lewis, M. B. (2003). Thatcher's children: development and the Thatcher illusion. *Perception, 32*, 1415-1421

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

Liu, C. H., & Chaudhuri, A. (2000). Recognition of unfamiliar faces: Three kinds of effects. *Trends in Cognitive Sciences, 4*, 445-446

Liu, C. H., & Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. *Cognition, 83*, 31-48

Liu, C. H., & Ward, J. (2005). Advantages of three-dimensional Methods for Face Recognition Research in Humans. *Lecture Notes in Computer Science, 3723*, 244-254

Liu, C. H., & Ward, J. (2006). The use of three-dimensional information in face recognition. *Vision Research, 46*, 768-773

Liu, C. H., Collin, C. A., Burton, A. M., & Chaudhuri, A. (1999). Lighting direction affects recognition of untextured faces in photographic positive and negative. *Vision Research, 39*, 4003-4009

Liu, C. H., Ward, J., & Young, A. W. (2006). Transfer between two- and three-dimensional representations of faces. *Visual Cognition, 13*, 51-64

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. *Applied Cognitive Psychology, 1*, 53-69

Luria, S. M., & Strauss, M. S. (1978). Comparison of eye movements over faces in photographic positives and negatives. *Perception, 7*, 349-358

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 and Psychiatry, 45*, 820-822

Marotta, J. J., McKeeff, T. J., & Behrmann, M. (2002). The effects of rotation and inversion on face processing in prosopagnosia. *Cognitive Neuropsychology, 19*, 31-47

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shape. *Philosophical Transactions of the Royal Society of London: Series B, 200*, 269-294

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

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*, 3-35

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*, 553-566

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R. & Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science, 10*, 419-422

Moses, Y., Ullman, S., & Edelman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception, 25*, 443-461

Norton, D. W. (1952) *An empirical investigation into some effects of non-normality and heterogeneity on the F-distribution.* Unpublished dissertation, State University of Iwoa, 1952.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning Memory and Cognition, 14*, 700-708

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 3-27

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

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

O'Toole, A. J., Edelman, S., & Bulthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research, 38*, 2351-2363

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long, & A. D. Baddeley (eds) *Attention and performance IX*. Hillsdale, NJ: Lawrence Earlbaum Associates

Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296*, 1321-1323

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 406-417

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London Series B-Biological Sciences, 223*, 293-317

Phillips, R. J. (1972). Why are faces hard to recognize in photographic negative. *Perception & Psychophysics, 12*, 425-426

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition, 4*, 409-437

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3-dimensional objects. *Nature, 343*, 263-266

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 41*, 797-828

Rae, G. (1976). Table of A'. *Perceptual and Motor Skills, 42*, 98

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

Roark, D. A., Barrett, S. E., Spence, M. J., Abdi, H., & O'Toole, A. J. (2003). Psychological and neural perspectives on the role of motion in face recognition. *Behavioural and Cognitive Neuroscience Reviews, 2*, 15-46

Roark, D. A., O'Toole, A. J., & Abdi, H. (2003). Human recogntion of familiar and unfamiliar people in naturalistic video. *IEEE International Workshop on Analysis and Modelling of Face and Gestures,* 36-43

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception, 32*, 285-293

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156-163

Shepherd, J., Davies, G., & Ellis, H. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (eds) *Perceiving and remembering faces*. London: Academic Press

Siéroff, E. (2001). Feature processing and superiority of three-quarter views in face recognition. *Brain and Cognition, 46*, 272-276

Sim, T., Baker, S., & Bsat, M. (2002). The CMU Pose, Illumination, and Expression (PIE) Database. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, May 2002*

Solso, R. L., & Mccarthy, J. E. (1981). Prototype formation of faces - a case of pseudo-memory. *British Journal of Psychology, 72*, 499-503

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity: evidence from eye-movement data. *British Journal of Psychology, 96*, 407-422

Sturzel, F., & Spillmann, L. (2000). Thatcher Illusion: Dependence on angle of rotation. *Perception, 29*, 937-942

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

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences, 7*, 23-30

Tarr, M. J., & Pinker, S. (1989). Human object recognition uses a viewer-centered frame of reference. *Bulletin of the Psychonomic Society, 27*, 506

Tarr, M. J., Kersten, D., & Bulthoff, H. H. (1998). Why the visual recognition system might encode the effects of illumination. *Vision Research, 38*, 2259-2275

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1*, 275-277

Thompson, P. (1980). Margaret Thatcher - A new illusion. *Perception, 9*, 483-484

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

Troje, N. F., & Kersten, D. (1999). Viewpoint-dependent recognition of familiar faces. *Perception, 28*, 483-487

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

Valentin, D., Abdi, H., & Edelman, B. (1997). What represents a face? A computational approach for the integration of physiological and psychological data. *Perception, 26*, 1271-1288

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

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

Valentine, T., & Bruce, V. (1986a). The effects of distinctiveness in recognizing and classifying faces. *Perception, 15*, 525-535

Valentine, T., & Bruce, V. (1986b). Recognizing familiar faces - the role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40*, 300-305

Vician, C., & Desanctis, G. (2000). The impact of role training in a user-driven group support system environment. *Group Decision and Negotiation, 9*, 275-296

Wallis, G., & Bulthoff, H. H. (2002). Learning to recognize objects. In M. Fahle, & T. Poggio (eds) *Perceptual Learning*. London, England: MIT Press

Ward, J., & Liu, C. H. (2005). VRVision: a new tool for the display of 3-d images in behavioral research. *Behavior Research Methods, 37*, 464-469

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. (1979). On training people to recognize faces. *Ergonomics, 22*, 333-343

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

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception, 31*, 995-1003

Young, A. W. (1998). *Face and mind*. Oxford: Oxford University Press

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*, 495-523

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*, 737-746

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