Stability from variety: The prototype effect in face recognition

Janelle E. Renfrew

March 2008

Thesis submitted for the degree of PhD

Department of Psychology
University of Stirling

Abstract

The central goal of the current thesis was to increase our understanding of how representations of individual faces are built from instances that vary. The *prototype effect* was used as a tool to probe the nature of our internal face representations. In face recognition, the prototype effect refers to the tendency to recognize, or find familiar, the average image of a face after having studied a series of similar face images. The experiments presented in this thesis investigated the modulating role of different variables on the prototype effect in face recognition. In the study phase, two or more different exemplars based on the same identity were presented. In the test phase, one of the seen exemplars, the unseen prototype, and an unseen exemplar of each studied identity were presented one at a time, and participants were asked to make a recognition judgement about the prior occurrence of either the exact image or the person's face. Variants of each face identity were either unaltered images of real people's faces, or they were created artificially by manipulating images of faces using several different techniques. All experiments using artificial variants produced strong prototype effects. The unseen prototype image was recognized more confidently than the actually studied images. This was true even when the variants were so similar that they were barely perceptually discriminable. Importantly, even when participants were given additional exposure to the studied exemplars, no weakening of the prototype effect was observed. Surprisingly, in the experiments using natural images of real people's faces, no clear recognition advantage for the prototype image was observed. Results suggest that the prototype effect in face recognition might not be tapping an averaging mechanism that operates solely on variations within the same identity.

Acknowledgements

I would like to thank NSERC and the Department of Psychology at the University of Stirling for funding the current research. Above all, I thank and acknowledge Josephine Ross for overseeing the printing and binding of the initial examination version of this thesis. Without her, none of this would have been possible. Thanks to my supervisors, Vicki Bruce for her advice and encouragement, and Peter Hancock for always being more than just a *second* supervisor. I am very grateful to Steve Lindsay for his supervision while I was conducting experiments at the University of Victoria. Also, my thanks go to Mike Masson for enduring through all the lengthy question-answer sessions. Thank you to my parents for their love and support. Last but not least, thanks to everyone in the face lab but especially to Yoriko, Ed, Robbie, and Anna for being unforgettable friends.

Table of Contents

Chapter 1: Introduction

The overall aim of the present work is to elucidate how representations of faces are built from varying instances. Specifically, I am interested in how identity is internally represented as we learn new faces. To investigate this, the *prototype effect* will be used as a tool to probe the nature of the underlying face representations. Although the effect has been studied in other areas of cognition, in regard to face recognition the prototype effect “refers to the tendency to recognize the face corresponding to the central value of a series of seen faces…” (Cabeza, Bruce, Kato, & Oda, 1999; p. 139). A better understanding of the prototype effect will lead to a better understanding of memory for faces or at least promote cautious and thoughtful interpretation of theoretical and experimental work concerning the nature of face representation.

In this review of the literature, material on the recognition of familiar and unfamiliar faces, the prototype effect, and the implications of different models of memory will be covered. Following this, the current study will be introduced and the structure of the thesis will be outlined. Finally, a general methods section is provided at the end of this chapter.

Literature Review

Recognizing Familiar and Unfamiliar Faces

Unless one is familiar with a face, it can be difficult to recognize it across changes in viewing conditions or appearance (Bruck, Cavanagh, & Ceci, 1991; Burton, Bruce, & Hancock, 1999; Hancock, Bruce, & Burton, 2000; O’Toole, Deffenbacher, Valentin, & Abdi, 1994). Even matching identity from images of unfamiliar faces can be difficult (Bruce, Henderson, Newman, & Burton, 2001; Henderson, Bruce, & Burton, 2001). This is true even when images in the matching task are presented simultaneously rather than having to be retrieved from memory. However, recognition of familiar faces is quite resistant to image degradation (Burton, Wilson, Cowan, & Bruce, 1999; Harmon & Julesz, 1973; James, Humphrey, Gati, Menon, & Goodale, 2000; Liu, Seetzen, Burton, & Chaudhuri, 2003). Megreya and Burton (2006a) suggested that this difficulty is due to poor initial encoding of unfamiliar faces. They suggest that an unfamiliar face does not automatically engage the same perceptual processes as familiar faces. In agreement, Buttle and Raymond (2003) assert that the perceptual processing of familiar faces proceeds in a more efficient manner and requires less attentional resources than unfamiliar faces

Internal/External Shift

One of the most well known effects of familiarity on face processing is the shift of the relative importance of internal and external features in face recognition. The importance of the internal features relative to the external features is higher for familiar faces compared to less familiar faces (Bonner, Burton, & Bruce, 2003; Bruce, Henderson, Greenwood, Hancock, Burton, & Miller, 1999; Clutterbuck, & Johnston, 2002; Ellis, Shepherd, & Davies, 1979; Haig, 1986; Nachson, Moscovitch, & Umiltà, 1995; Young, Hay, McWeeny, Flude, & Ellis, 1985). For example, Ellis et al. found that unfamiliar faces studied from full face photographs were recognized equally well at test from either their internal or external features alone. In contrast, they also found that famous faces were both identified and recognized better from internal features only compared to external features only. Bonner et al. (2003) found that in a face matching task, unfamiliar faces were matched more accurately from their external features compared to their internal features, but after these same faces were made more familiar, matches were made from internal and external features with equal accuracy. Does this suggest that familiar faces are processed differently to unfamiliar faces in some fundamental way? Some believe this to be the case (see Ellis et al., 1979; Hancock et al., 2000; Megreya & Burton, 2006a; Megreya & Burton, 2006b; Ryu & Chaudhuri, 2006). In fact, the shift toward an advantage for internal features has been proposed for use as an empirical indicator of familiarity (Clutterbuck, & Johnston, 2002).

Inversion is a manipulation commonly believed to disrupt configural processing with little effect on featural processing of faces (this issue is reviewed in more detail later on in this chapter). Additionally, the detrimental effect of inversion is increased when only internal features of a face are visible (Leder & Bruce, 1998; Moscovitch & Moscovitch, 2000). Therefore, it is possible that a major difference between familiar and unfamiliar faces is how much their processing relies on precise configural information (Buttle & Raymond, 2003).

The Correlation Between Familiar and Unfamiliar Face Processing

Further evidence of different processing for familiar and unfamiliar faces comes from studies analyzing the correlations between recognition of familiar and unfamiliar faces. If recognition of these two types of faces uses the same process, one might expect to find a positive correlation between a participant's ability to recognize familiar and unfamiliar faces. Conversely, recognition deficits of brain injured patients reveal no such correlation suggesting a dissociation between the underlying processes (Warrington & James, 1967; see also Benton, 1980). Moreover, Megreya and Burton (2006a) found that while face matching performance on upright and inverted faces does show a positive correlation when using unfamiliar faces, no correlation across orientation exists when identifying or matching familiar faces.

However, results from other studies conflict with those of Megreya and Burton. Yin (1969) found the opposite relationship in a recognition task using unfamiliar faces. Participants were split into two groups based on their performance on inverted faces. Yin found that the group that was "worse" on inverted faces was more accurate at recognizing upright faces than the "better" group was. Still different results were found by Phillips and Rawles (1979). Phillips and Rawles found no correlation between the recognition of upright and inverted unfamiliar faces, but a positive correlation between upright and inverted familiar faces in an identification task. The inconsistencies among findings may partially be due to differing task demands. For example, Megreya and Burton only used test of simultaneous matching or immediate memory. In contrast, both Yin and Phillips and Rawles' studies were divided into separate study and test phases (when using unfamiliar face stimuli) so faces would have to be retained in memory for a much longer amount of time. Also most of Megreya and Burton's experiments used a line up task where participants tried to match the target out of an array of ten images of different peoples' faces. In half the trials, the target was present in the array, and in half the trials the target was absent. In Yin and Phillips and Rawles' experiments, participants were presented with two test faces and were asked to select the one from the study phase. The target was always presented in each test pair. Finally, the experiments of Yin and Phillips and Rawles used the same images in both the study and test phases.

Viewpoint and Image Dependency in Face Recognition

Changing viewpoint is another manipulation used in experimental tasks that distinguishes recognition of familiar faces from unfamiliar faces. Once made familiar with a face, one can recognize it with ease from any viewpoint. However, if a face is unfamiliar, recognizing it from a novel viewpoint may be difficult or unlikely (Bruce, 1982). If asked to decide whether two images taken from different viewpoints depict the same person or not, participants perform more accurately if images are of familiar rather than unfamiliar faces (Hill & Bruce, 1996, experiment 1). One possible explanation is that the unfamiliar faces have representations that are relatively unformed compared to familiar faces. Especially in an experimental context where unfamiliar faces are only presented as a single still image, the representation formed may be largely image based since the studied image is really all one has to go on (see Bruce & Young, 1986). If the representation of an unfamiliar face only consists of memory for a small number (possibly only one) of images, then the representation must be heavily viewpoint dependent by necessity (Ryu & Chaudhuri, 2006). Consistent with this hypothesis, Kemp, Pike, White, and Musselman (1996) found that recognition of familiar faces was not affected by manipulations to the colour of images, but this did affect recognition of unfamiliar faces. Thus, if memory for an unfamiliar face is, in some cases, akin to picture memory, then colour information may be important in image recognition but less so for face recognition in general. On the other hand, even pictures of familiar faces are still stored as separate memories, and in certain experimental tasks, may influence recognition decisions or response times (Bruce, 1982; Bruce & Valentine, 1985).

A viewpoint/image-based representation differs from what some researchers have proposed for familiar faces (e.g. Bruce & Young, 1986). The representation of a familiar face differs from that of an unfamiliar face mainly in that it is more abstract and more interconnected to other representations in memory. A familiar face representation is thought to be view-point invariant (Ryu & Chaudhuri, 2006).

Another similar possibility is that a familiar face representation may include a small number of examples of or rules for transformations such as viewpoint (Ellis et al., 1979). Representations of the appearance of a familiar person's face are also thought to be linked to "semantic" information about the person including their name (e.g. Bruce & Young, 1986).

Megreya and Burton (2006a) warn against trying to explain identity processing of familiar and unfamiliar faces under a single theory. They acknowledge, of course, that unfamiliar faces sometimes do become familiar through experience. Thus, a major challenge for future theoretical work will be modelling the process by which newly familiarized faces are added into memory.

Stevenage, Lee, and Donnelly (2005) have gone a step further by arguing that even the initial perception of a face is altered once it is made familiar. Familiarity with a face can affect processing prior to activation of its representation in memory. They contend that the differential effects of familiarity found in studies of caricature and categorical perception are evidence of this.

Caricature Effects

Caricaturing images of faces was a popular technique used in the 1980's and 90's in studies of facial distinctiveness. A *caricature* exaggerates the distinctive aspects of a person's face. Caricatures can be made using computer software programs, but of course the term "caricature" comes from the portrait style used by some artists. Computer-based image transformations let experimenters produce their own caricatured images in a more controlled manner than artists could achieve.

The technique developed by Benson and Perrett (1991) will be used as an example to illustrate how a caricature transform works. To produce a caricature, first, a large number of images of different peoples' faces are averaged together. This will involve manually plotting reference points on each face and the use of morphing software to determine the average face shape and texture. The image selected for caricaturing has also had the same reference points plotted on it. Next, the software calculates the difference between the selected image and the average face and applies this difference to the selected image. The result is an image that has been morphed such that any deviations from normal in the original image are now made even more deviant. The degree of caricature can also be controlled, so experimenters can choose how much exaggeration they want to apply to a face. It is also possible to make *anti-caricatures* by reducing rather than exaggerating the differences from the average face (see Figure 1.1).

Figure 1.1. Example of an anti-caricature (left) and caricature (right) of the actor, Harrison Ford. The middle image is a reproduction of the original photograph (taken from Lee & Perrett, 2000).

In experiments, *caricature effects* are usually studied using tasks of identification, recognition, or judgements of best likeness. Performance on caricatures can be compared to performance on the original, or *veridical*, images. For example, if the task is identifying famous faces, half of the images would be caricatures and half would be veridical images of the famous faces. If faces are named faster or more accurately when they were presented as caricatures compared to veridical, a caricature effect is said to be observed. To study caricature effects of unfamiliar faces, one would need to give participants some amount of minimal exposure to a set of faces before the test phase.

Of interest is the finding that recognition of unfamiliar faces shows little or no benefit from caricaturing whereas familiar faces often do benefit from caricaturing. Rhodes, Brennan, and Carey (1987) found veridical line drawing of unfamiliar faces were rated as better likenesses of the people they depicted compared to both caricatures and anti-caricatured line drawings of those same people. This was true when participants were recalling an earlier presentation of the unfamiliar face from memory or when a photograph of the same person was presented on the screen for comparison. Rhodes and Moody (1990) also reported a failure to find caricature effects in a recognition task using unfamiliar faces presented in an earlier study phase. Likewise, Deffenbacher, Vetter, Johanson, and O’Toole (1998) reported that pilot tests could not produce a caricature effect in a recognition task unless participants had a minimum of about one minute prior exposure to the face.

Benson and Perrett (1991, experiment 1) investigated the effects of caricaturing photographs of famous faces on judgements of best likeness. On each trial, participants were presented with four morphed images of the same face. The range of caricaturing across the four images went from anti-caricature to caricature and also included a veridical image of the face. Participants were asked to rate the familiarity of the person depicted as well as select one image as being the best likeness of that person. Benson and Perrett found that the degree of caricature selected as best likeness correlated positively with how familiar the participant was with the depicted person. Thus, the more familiar a face, the more likely a caricature of it would produce an image of better likeness. Stevenage et al. (2005) interpret these results as indicating “that caricaturing only comes to play a role in face processing when the faces are familiar” (p. 1105). Their interpretation is more of a summary of past findings than an illuminating discussion of their implications.

However, Stevenage et al. (2005) do make an important point about the lack of explanatory power of Valentine’s (1991) face space model. A slight caricature should improve recognition a small amount by reducing the competition of other nearby face representations in the space. This should hold true for both familiar and unfamiliar faces. I can only speculate that perhaps the trade off between encoding specificity (i.e. the closeness in match between the internal representations of the test and study images) and reduced competition from other nearby faces is affected by familiarity because an unfamiliar face may rely more on picture-to-picture matching compared to familiar faces. Recognition of familiar faces should be able to draw on a number of varying stored representations of previous encounters with that face. Therefore, recognition of familiar faces may tolerate a larger degree of transformation before the lack of encoding specificity outweighs the improvement due to reduction of competing similar looking faces.

Categorical Perception

Categorical Perception is the perception of a categorical structure imposed onto stimuli that vary along a continuum. It is the opposite of continuous perception. The classic example used to illustrate categorical perception is colour. Our perception of a rainbow is not a smooth continuum of colour. We see bands of colour when in fact, physically, the light frequencies change steadily across the span of the rainbow. Thus, categorical perception seems to stretch out the psychological similarity space in some regions and/or compress it in others to make stimuli of the same category appear more similar and stimuli of different categories appear more distinct.

In operational terms, categorical perception is said to occur whenever perceived differences are more difficult to detect within-category and/or more easily detected between-category, compared to a baseline. For established categories, the baseline is usually the physical size of the differences. In studies of categorical learning, categorical perception is usually assessed after learning and compared to a baseline taken before learning. The standard methodology, as applied to faces, for testing categorical perception was used by Beale and Keil (1995) and will be described next in more detail.

Beale and Keil (1995) sought to determine whether familiar faces are perceived categorically. To do this, they paired up photographs of famous faces and created a morphed continuum of blended faces between the two faces in a pair. The continuum consisted of 11 faces morphed at 10% increments (exception: the two original faces were blended by 1% to ensure all stimuli had undergone the morphing process). For example, in the Kennedy/Clinton continuum, the first face would be 99% Kennedy/1% Clinton; the second face would be 90% Kennedy/10% Clinton; the third face would be 80% Kennedy/20% Clinton, and so on. The procedure had two parts: a discrimination task and a categorization task. In the discrimination task, three face images were presented consecutively. The third image was the target face and was identical to one of the two previous images. The first two images always differed by 20% (two steps on the continuum). The task was to indicate whether it was the first or the second image that matched the third image. In the categorization task, all images in the continuum were presented one at a time and participants were asked to classify each one as either more like Kennedy or more like Clinton.

To test for categorical effects using artificial categories, one must determine where the categorical boundary lies along the continuum and compare discrimination at this point with performance along the rest of the continuum. If categorical perception is present, then discrimination performance should peak on the two-step pair that straddles the categorical boundary. The categorization task is used to determine the categorical boundary. Beale and Keil (1995) chose 33% and 66% as the cut-off points for the categorical boundary. This meant, for example, that images judged Kennedy on more than 66% of trials could be considered members of the Kennedy category. Images judged to be Kennedy on less than 33% of trials could be considered members of the Clinton category. Any other images were, on average, not consistently judged as belonging to either category, so they would be right on the categorical boundary. According to Beale and Keil's results, images four, five, and six were judged as Kennedy approximately 90%, 45%, and 15% respectively. Therefore, the two-step pair that straddles the categorical boundary would be pair 4/6. Discrimination performance on pair 4/6 was higher than the average performance on all other pairs combined. Thus, the hallmark of categorical perception was observed since the discrimination of differences was poorer within a category than between categories.

Beale and Keil (1995) did indeed find categorical effects for famous faces of different identities. They also found some support for an effect of familiarity on categorical perception. The effect of familiarity was investigated by collecting familiarity ratings on a set of famous faces and then creating morphed continua between pairs of faces of different levels of familiarity (faces belonging to the same pair were matched for familiarity). The level of familiarity of a face pair correlated positively with the strength of categorical perception. However, more compelling evidence would have been obtained by using the same stimuli in both the familiar and unfamiliar conditions and manipulating the degree of learning. This would at least rule out the possibility that the differences in the categorical perception task were being driven by some systematic difference between the faces rated as most familiar and the faces that were rated as less familiar.

Since Beale and Keil (1995) first demonstrated categorical perception of facial identity, several studies have attempted to examine the same phenomenon throughout the course of learning within an experiment. However, results have been inconsistent. Angeli, Davidoff, and Valentine (2001) and Campanella, Chrysochoos, and Bruyer (2001) both concluded that the identity of unfamiliar faces are not perceived categorically (unless the identities are highly distinctive). In contrast, other studies have reported categorical perception using unfamiliar faces (Campanella, Hanoteau, Seron, Joassin, & Bruyer, 2003; Levin & Beale, 2000; Viviani, Binda, & Borsato, 2007). McKone, Martini, and Nakayama (2001) and Stevenage (1998) have demonstrated the emergence of categorical perception after extensive familiarity training on previously unfamiliar faces. It should be noted that the standard methodology of testing for categorical perception requires at least a low level of familiarity with the stimulus. In fact, learning from the test phase alone is enough to produce a categorical perception effect (Viviani et al., 2007). Therefore, the important finding consistent across studies is that familiarity with the stimuli modulates categorical perception of faces. Categorical perception effects are stronger for more familiar faces than less familiar ones, and this is expected since a face, when seen for the first time, will have no categorical representation of its identity already established in memory. Trying to make all or none statements about the presence or absence of categorical perception will likely lead to confusion since the literature uses the term "unfamiliar" to refer to the familiarity status of a face *before* the start of an experiment, but learning of the face takes place throughout the experiment.

Face Classification: Evidence from Thatcher Faces

Stevenage et al. (2005) present their own experimental findings regarding the effect of familiarity on perceptual or structural encoding processes. They make use of Thatcherized faces in which the eyes and mouth of a face are inverted while the rest of the face remains in place (see Figure 1.2). This Thatcher task basically serves as an alternate version to the face/no face task which uses faces with scrambled features as distractors. The Thatcher task is more difficult though, particularly when the stimuli are inverted. In experiment 1, a two alternative forced choice task was used that pitted a Thatcherized version of a face against a normal photograph of that same person's face. Half of the trials used famous faces, and half used unfamiliar faces. Also, half of the trials were upright, and half were inverted. In experiment 2, the same stimuli were presented one at a time, and participants judged whether the face was normal or odd. Stevenage et al. expected an advantage for familiar faces to be revealed within the inverted condition due to the increased difficulty of the task. Indeed, in both experiments, results showed similar performance at both levels of familiarity when faces were upright, but when faces were inverted, performance was higher on familiar faces than unfamiliar faces. The problem with their argument is that they assume detecting whether the face is intact is a perceptual process that does not require accessing a stored memory (at least a memory of that particular face or a similar face). They assume that this is an early perceptual process performed during structural encoding and, according to the Bruce and Young (1986) model, should be completed before the familiarity of the face is signalled. Stevenage et al. then explain how their results could be explained if one assumes "First, the decisions of familiarity and face-ness must proceed in parallel. Second, the familiarity decision must be achieved earlier than the face classification decision. Third, completion of the familiarity decision must facilitate the face classification decision" (p. 1110). If one makes these assumptions, then by necessity accessing stored representations of familiar faces is not a later process; if anything, it is the earlier process. In a sense, their argument seems to be that familiarity can affect earlier face processing if those earlier face processes are not really earlier after all.

Figure 1.2. Examples of normal and Thatcherized faces. Bottom shows an unfamiliar familiar face and top shows the famous face of the actress Jamie Lee Curtis (taken from Stevenage et al., 2005).

Levels of Familiarity with Faces

How much experience is needed with a face for it to become familiar? O'Donnell (2003) claimed that 18 minutes of exposure can produce familiarization equal to that of personally known familiar faces. Tong and Nakayama (1999) assert that thousands of exposures (rather than hundreds) is needed to make a face as familiar as a face of a close friend or relative. Results from experiments reported by Buttle and Raymond (2003) support this assertion since famous faces were found to produced better performance on a change detection task than newly familiarized faces even after hundreds of exposures. Ryu and Chaudhuri (2006) emphasise the importance of encoding across variable viewing conditions and also the accumulation of semantic information in attaining familiarity with a person's face.

For familiar faces, the distinction is often drawn between personally familiar faces (friends, family) and familiar famous faces. There may be differences between these groups if familiarity has been induced in different ways (e.g. TV, photos, vs. real life interaction). O'Donnell (2003) found evidence that newly familiarized faces (in a lab setting) yielded different patterns of manipulation detection than personally familiar ones. Buttle and Raymond (2003) found that with newly familiarized faces, participants were slower to detect changes than with famous faces using a task similar to those used in attentional blink studies. This held true despite participants being given ample opportunity to learn semantic information about the newly familiarized faces. Therefore, it is likely that the information stored and/or retrieved for very familiar faces is slightly different than newly familiarized faces.

Effects of Familiarity on Other Face Perception Tasks

Familiarity with a face has been shown to improve performance on perceptual classification tasks such as gender (Ganel & Goshen-Gottstein, 2002; Rossion, 2002; Rossion, Schiltz, Robaye, Pirenne, & Crommelinck, 2001), age (Bruyer, Lafalize, & Distefano, 1991; Bruyer, Mejias, & Doublet, 2007), ethnicity (Bruyer, Leclere, & Quinet, 2004), emotional expression (Ganel, Valyear, Goshen-Gottstein, & Goodale, 2005; Schweinberger, Burton, & Kelly, 1999; Schweinberger & Soukup, 1998;), and lip reading (Schweinberger & Soukup, 1998).

Other factors, besides our past experience with that face, may affect how familiar a face is perceived to be. While smiling faces are judged as more familiar than faces with a neutral expression (Baudouin, Gilibert, Sansone, & Tiberghien, 2000), faces with a negative expression are judged as less familiar (Lander & Metcalfe, 2007).

Representing Faces as Parts and Wholes

Researchers have reached a general consensus that information about the individual features of a face and their configuration are both important in recognizing faces. A more contentious issue is whether face representations are feature based or holistic, that is, not decomposed into discrete features or parts. Part of the difficulty of settling this debate is the lack of precise, consistent definitions of the terms *feature*, *configuration*, and *holistic* (Tanaka & Farah, 1993).

In the context of this debate, the terms feature, part, and component can usually be used interchangeably and most commonly refer to the most obvious, intuitive, and nameable features of the face including eyes, nose, mouth, etc. Configuration usually refers to the spatial relationships or positioning of features within the face. Use of the term holistic is notably inconsistent. Holistic and configural processing often refer to the same thing; however, the distinction is sometimes made (McKone, Kanwisher, & Duchaine, 2007). When the distinction is made, it is usually to emphasize the lack of parts in an undifferentiated holistic representation whereas representing the configuration implies that there are discrete features that vary subtly in their spatial arrangements. What configural and holistic processing both emphasize is the importance of relational (i.e. spatial) information that integrates so called features.

A lot of the confusion surrounding configural and holistic processing stems from the intuitive assumption that faces can be thought of as a set of discrete features which can be arranged spatially into different configurations (Bruce & Burton, 2002). So if faces are assumed to be composed of parts and part relations what alternative hypotheses have been put forth? The main alternative view is the holistic face hypothesis (Farah, Wilson, Drain, & Tanaka, 1998; Tanaka & Farah, 1993) which assumes the opposite. According to this view, faces are not processed in terms of their component features, so by necessity, the feature relations cannot be processed either (because spatial relationships are among specified component features). Bruce (1979) expressed this debate as being a question of "whether faces are processed analytically (feature-by-feature) or globally, as Gestalten." (p. 374).

Early research on the relative importance of feature processing focused on whether facial features were processed serially or in parallel. Bradshaw and Wallace (1971) reasoned that serial processing of features is consistent with the idea that faces are represented by their component features, and parallel processing is consistent with the proposal of a more holistic representation of faces. They used Identikit faces in a sequential matching task and measured the time it took participants to make a "same" or "different" decision. Reaction times varied according to the number of features that differed between the two faces being compared, with faster responses being associated with a higher number of differing features. Bradshaw and Wallace interpreted this result as evidence of serial self-terminating search for differing face features and, therefore, feature-based processing.

Using the same approach as Bradshaw and Wallace (1971), Matthews (1978) used a simultaneous matching task again using Identikit faces. He found that the pattern of results depended on which features differed between the two faces of a pair. Specifically, Matthews found that changes to the eyes, chin, and hair were spotted faster and processed in parallel; whereas changes to the eyebrows, nose, and mouth were spotted slower and were processed serially, feature-by-feature. The overall conclusion Matthews came to was that some features of a face are processed serially and others, in parallel (possibly holistically). However, Sergent (1984) pointed out that the finding of equal reaction times across changes in eyes, chin, and hair may have been an artefact of analyzing data averaged across participants. If participants used different serial scanning strategies, this would have been hidden by averaging the data; and thus, equal response times across feature changes does not rule out serial processing of features as the sole underlying process driving all of the results.

Smith and Nielsen (1970) found evidence that matching faces from short term memory involves more than one style of processing. They used a sequential matching task which varied the length of delay (1, 4, or 10 sec) between presentation of the first and second face of each pair. In addition to manipulating the number of differing features within each pair, they also manipulated the number of task relevant features in each block of trials. Reaction times for correct “different” and “same” responses produced slightly different patterns of results. As the number of differing features in a pair increased, reaction times decreased. This pattern is consistent with the use of some type of self-terminating, feature-by-feature comparison (either serial or parallel), and this pattern held true regardless of the length of the delay interval. For “same” responses, reaction times increased as the number of relevant features increased. This pattern was also interpreted as having involved self-terminating, feature-by-feature comparisons. However, this pattern was only true for “same” responses at long lags (10 sec). In contrast, after a 1sec delay, reaction times of “same” responses were not affected by the number of task relevant features. Smith and Nielsen argued that this pattern of results was evidence that participants were treating the faces as unitary wholes when making comparisons on “same” trials with short delay intervals

Sergent (1984) tested the related hypothesis that facial features are processed independently of one another. She used Photofit faces in a simultaneous matching task. Her rationale was that if features are processed independently, then change detection should not be faster to a pair of faces differing by two features than a pair differing only on the most salient feature. She used three features: eyes (plus eyebrows), chin contour, and a feature Sergent termed “internal space” (the closeness of the internal face features), and each feature could take one of two values. Results indicated that chin contour was the most salient feature while eyes and internal space were less salient than the chin but equally salient compared to one another. Contrary to the prediction made when assuming independence of feature processing, face pairs differing by both the chin and the internal spacing were more quickly judged as “different” than pairs differing by the chin alone. This interactive pattern was not found for the eyes. Thus, Sergent concluded that some features may interact with one another and be processed holistically, but other features may be processed independently of one another. This could be achieved by separate featural and configural processes operating simultaneously on different properties of the facial stimuli. The obvious problem with this study is that the feature “internal spacing” may not be treated in the same way as the more conventional features (eyes, nose, mouth etc.), and since it was the only feature that interacted with other features, conclusions are severely limited in that respect.

The main problem with the four studies just described is that they based their hypotheses on inappropriate assumptions about the models they are testing. For example, Smith and Nielsen (1970), Bradshaw and Wallace (1971), and Matthews

(1978) all assume that the number of features in a face or the number of differing features between two faces should have no effect on reaction times in a matching task if faces are processed and represented as holistic Gestalts. This is not the case since holistic representations can vary in the amount of information they contain which could affect how quickly they could be used. Furthermore, the number of differing features between two faces would affect how different their holistic representations are, which in turn would affect how quickly a difference could be detected (note: the term features is referring to the features as defined and manipulated by the experimenters. Of course, if faces were represented as holistic Gestalts, they would not contain any countable features). Also, whether faces were matched using a serial or parallel search strategy does not distinguish holistic Gestalt processing from configural processing of the spatial relations of features since both would likely involve parallel search. It is even possible that features could be processed in parallel without necessarily representing their spatial relationships (Tanaka & Farah, 1993).

The next major line of related research was spurred on when presenting stimuli upside down was found to cause more impairment in recognition and perceptual tasks using faces compared to other types of objects and visual patterns (Yin, 1969). This finding generated quite a lot of interest into what might be causing inversion effects and whether face perception involves special processes used only on faces. Thus, the question switched from what is the relative importance of featural vs. configural processing in overall face recognition to is face perception more dependent on configural/holistic processing relative to featural processing than is the perception of other kinds of visual objects (Tanaka & Farah, 1993)? If inversion impairs the perception of faces more than other types of objects, then determining more precisely the locus of impairment will shed light on how facial processing might differ from the processing of most other objects.

Most researchers agree that upright and inverted faces are not processed in the same way (Goffaux & Rossion, 2007). In general, inverting a face makes it more difficult to recognize (Yin, 1969) and makes distortions or alterations applied to it more difficult to detect (Bartlett & Searcy, 1993; Thompson, 1980). The predominant view is that while information from the local features and their configuration are both important for processing upright faces, the processing of configural information is greatly impaired when faces are inverted (Bartlett & Searcy, 1993; Barton, Keenan, & Bass, 2001; Freire, Lee, & Symons, 2000; Leder & Bruce, 1998; Leder & Bruce, 2000; Leder & Carbon, 2006; Le Grand, Mondloch, Maurer, & Brent, 2001; Mondloch, Le Grand, & Maurer, 2002; Murray, Yong, & Rhodes, 2000; Rhodes, Brake, & Atkinson, 1993; Rhodes, Hayward, & Winkler, 2006; Searcy & Bartlett, 1996; Sergent, 1984).

One of the most common methods used to investigate the effects of inversion is measuring participants' sensitivity to manipulations of the featural vs. configural information within upright vs. inverted faces. However, manipulations of featural or configural information often affect one another to some extent, so entirely pure measures of featural or configural processing are not possible. Even still, through careful manipulation of stimuli, sets of faces can be produced that vary the configural information with little effect on the information about the local features and vice versa. To manipulate the configural information, the standard approach is to keep the individual features the same, but to alter their spacing or positioning within the overall face shape (e.g. Bartlett & Searcy, 1993; Freire et al., 2000; Hosie, Ellis, & Haig, 1988; Kemp, McManus, & Piggot, 1990; Le Grand et al., 2001; Leder & Bruce, 2000; Leder & Carbon, 2006; Leder, Candrian, Huber, & Bruce, 2001; Mondloch et al., 2002; Murray et al., 2000). Less obvious is how features can be altered without disturbing the configuration among them (i.e. the precise distances between features).

Past attempts at manipulating featural information have included swapping the facial features among images of different faces (Freire et al., 2000; Goffaux, Hault, Michel, Vuong, & Rossion, 2005; Goffaux & Rossion, 2007; Le Grand et al., 2001; Mondloch et al., 2002) and also making surface alterations like colour and brightness to the local features (Barton et al., 2001; Leder & Bruce, 2000; Leder & Carbon, 2006; Murray et al., 2000). The obvious problem with the feature swapping method is that it will slightly alter the precise spatial relationships among the features. Any manipulation that alters the shape of the facial features will have this problem by necessity. Thus, researchers turned to surface manipulations of the features that would not alter the shape of the individual features or their spatial configuration. One problem with the latter method is that featural processing (if it exists) supposedly would process the shape of the features, or at least would process more than just their hue and brightness, so it is unclear whether detecting changes to just the surface properties of features can provide a valid and representative measure of featural processing. An alternative technique is to blur (Sergent, 1986) or add random noise (McKone et al., 2001) to images of faces to remove fine detailed information about the local facial features.

Studies using manipulations of configural and/or featural information to investigate the face inversion effect have most commonly used a face discrimination task. For example, in a typical experiment, pairs of faces are presented on a screen, and participants are asked to judge if the two images are exactly the same or not. Half of the face pairs are presented upright, and half are inverted. The two faces in a pair are made different by either altering the positioning of the eyes and mouth or by swapping the eyes and mouth with the eyes and mouth of a different person's face. Accuracy and reaction times are analyzed to compare the relative contribution of processing the featural vs. configural information between upright and inverted faces. The consistent finding is that inversion reduces accuracy and increases reaction times much more so when discriminating changes to the configural rather than featural information in a face (e.g. Freire et al., 2000; Le Grand et al., 2001; Mondloch et al., 2002). Another task that has been used by several researchers is distinctiveness ratings (Leder & Bruce, 1998; Murray et al., 2000; Searcy & Bartlett, 1996). In this method, face stimuli are made to look more distinctive or bizarre by one of two methods. Some faces have had their features distorted in some way (e.g. blackening teeth, whitening or blurring pupils, bushy eyebrows), and other faces have had the spacing of their features altered so as to appear unusual. Participants are asked to rate how unusual each face looks. Half of the faces are presented upright, and half are presented inverted. Compared to upright, inverting faces that have been made configurally distinctive lowers bizarreness ratings much more so than those that have been made featurally distinctive.

Of course it may not be that face specific processes are impaired by inversion. Faces may just happen to belong to a class of stimuli whose properties make them more difficult to be perceived upside down. In contrast to most other types of objects, faces are very similar in appearance to one another (i.e. visually homogenous), are almost always seen in the same orientation (upright), have bilateral symmetry, and belong to a class of stimuli with which we have had extensive visual experience (Goffaux & Rossion, 2007).

The studies, described thus far, that examined the face inversion effect did not set out to distinguish the view that faces are represented by their features plus the spatial relationships among those features from the view that faces are represented as holistic Gestalts. Since featural processing was contrasted with configural information specifically about the spacing of the features, these experiments do not provide direct evidence for or against the hypothesis that upright faces are represented and processed as nondecomposable wholes. So what evidence is there about the holistic nature of face processing?

In an experiment examining the efficacy of Photo-Fit faces as a facial composite system, Ellis, Shepherd and Davies (1975) found that participants had difficulty reconstructing a Photo-Fit face by selecting the identical features (e.g. the same mouth) from a pool of variants on those features (i.e. from a pool of different mouths). Even when the target face that they were trying to build remained in view, participants still made errors when trying to select the identical features. Bruce (1979) interpreted Ellis et al.'s results as indirect evidence of faces being processed as wholes without parts. Presumably, Bruce was implying that if facial features are represented explicitly and discretely, then participants would have been better at picking out the correct features. In fact, much of the evidence for holistic processing of faces rests on the key assumption that it would be more difficult to recognize a facial feature presented in isolation (e.g. a particular nose) if the underlying face representation is an undifferentiated whole than if it is based on the individual facial features.

Tanaka and Farah (1993) point out that the holistic vs. featural distinction does not have to be completely dichotomous, since object recognition may rely on both types of representations to differing degrees for different classes of objects. For this reason, they designed a series of experiments to test whether the degree of holistic processing used in face recognition is greater than that used in recognizing other objects and visual patterns. They chose scrambled faces, inverted faces, and houses as the objects to contrast with normal upright faces. The eyes, nose, and mouth were the features of the normal, inverted, and scrambled faces. The door, a big window, and a small window were the features of the house stimuli. The procedure was as follows: First, participants learn to name a set of normal upright faces and a set of one type of contrast stimuli. Then, two types of two alternative forced choice recognition tasks were conducted. In the isolated part test condition, two isolated features were presented simultaneously for comparison. Participants were asked “which is Jim’s nose?” In the full face test condition, two full faces were presented. Participants were asked “which is Jim?” The full face condition was designed such that the target and foil differed only with respect to the particular feature being tested. Thus, if Jim was the target, and the feature being tested was the nose; then, the foil would be Jim’s face with his nose replaced with a different person’s nose (see Figure 1.3). The test was the same for the contrast stimuli as participants treated inverted and scrambled faces as though they belonged to different people and houses as though they were owned by different people. Results showed that normal faces demonstrated a greater disadvantage for recognizing the isolated parts vs. the wholes compared to inverted faces, scrambled faces, and houses. Again, this was true despite the only difference between the stimuli used in the part and whole test conditions was the feature being tested. This result has become known as the *part-whole effect* in face recognition.

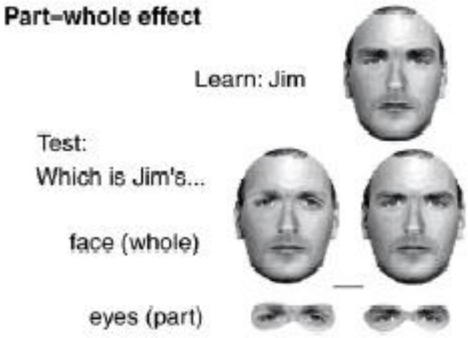


Figure 1.3. Example stimuli of the Part-whole task (taken from Goffaux & Rossion, 2007).

First demonstrated by Young, Hellawell, and Hay (1987), the *composite effect* is the most well known and arguably the most convincing evidence of holistic face processing. Young et al. constructed face stimuli by combining the top half of one famous face with the bottom half of another famous face. When the two halves of the faces were aligned, participants found it more difficult to identify the top or bottom face compared to when the two halves were offset (see Figure 1.4). This difference between the aligned and offset condition is known as the composite effect and is thought to be due to perceiving the aligned halves as a single whole face which interferes with the perception of the two halves. Using a matching task, Hole (1994) found that unfamiliar faces also produce composite effects. Interestingly, the composite effect disappeared when the same stimuli were presented upside down. This implies that the two faces could be identified by their local features, but alignment in the upright condition either interfered with extracting information from the local feature or led to the extraction of configural properties arising from the conjunction of the two original faces. Inversion may have made it easier to process the faces in terms of their component parts and/or suppressed processing of the misleading configural information.

Taken altogether, the research to date provides no persuasive evidence that faces are represented only in terms of the nameable features of a face. Whether faces are represented as a set of features plus their spatial configuration or as a holistic Gestalt is less clear. These two viewpoints are difficult to distinguish as evidence for one view can often be taken as evidence for the other view as well. Furthermore, since tests of the holistic hypothesis are centred around the presence or absence of featural processing, any evidence against featural processing can be taken as evidence for holistic processing. This reasoning is flawed even though it underlies the interpretation of almost all of the relevant experimental findings. The problem is that whether a face is treated as having features or not entirely depends on how one defines the features. Thus, the original contrast between "parts" or no "parts", would be more correctly framed as "*these* parts" or not "*these* parts". This would at least leave open the possibility that facial representations could be decomposable but not in terms of the nameable features of a face.

Figure 1.4. The composite effect. Left: an aligned composite. Right: a composite with the upper and lower halves offset. Tony Blair (top) and David Cameron (bottom).

Theories of Concept Representation: Prototypes vs. Exemplars

Theories have been put forward that store different identities as prototype representations. In its simplest form, each person’s face is stored as the central tendency of its variants. The notion is that the within face variation can be overcome by storing a central face representation that emphasizes the shared properties of its constituent variants while washing away the variation that is inconsistent among them (Burton, Jenkins, Hancock, & White, 2005). A purely prototype account of face recognition would posit that only the prototypes need be stored in long term memory for subsequent recognition. However, most current prototype accounts allow for the storage of the individual variants, or *exemplars*, in addition to the prototype representation of a person’s face.

The opposing viewpoint is the purely exemplar account which maintains that there is no need to have a special unitary representation that reflects numerous previous encounters with a particular face. Instead, models of the purely exemplar type only store individual exemplars and do not build up abstract category representations of any sort in long term memory. In its extreme form, a type of exemplar model known as episodic models of memory stores every experience with a face anew regardless of whether that exemplar has already been seen before or not.

The prototype vs. exemplar distinction has received some attention in the face recognition literature (e.g. Bruce et al., 1991). However, it is largely ignored by most of the models of face recognition that are currently popular (e.g. Valentine,1991; Burton & Bruce, 1993). Models often use a stable unitary data structure to represent different face identities without giving any specification at all as to how these seemingly abstract representations came to be. Granted, on intuitive and computational grounds, it is easier to conceptualize a representation of a person's face as being a single entity stored in memory. Nonetheless, unlocking the secrets of the human mind is not expected to be easy.

Although largely ignored in face recognition, the question of how information is internally represented has fuelled a long lasting debate in the broader domain of memory and cognition. A review highlighting this debate will follow. First, I will focus on studies of prototype effects found in non-face studies of recognition and categorization; then, I will focus specifically on the prototype effect in face recognition.

The Prototype Effect in Non-face Domains

Prototype Effects Using Dot Patterns

The debate between prototype models and exemplar-based models has been the concern primarily of those developing theories of concept formation and general memory mechanisms rather than the specific case of face recognition. Posner and Keele (1968; 1970) brought the issue into the limelight through their research on prototype abstraction using a random dot pattern classification task. In most of their experiments, 3 patterns served as the prototypical dot patterns of 3 different categories. Each of these patterns consisted of 9 dots filled in on a 30x30 grid (see Figure 1.5). Distortions of each category prototype pattern were generated. Distortions included a training set as well as ones that were only presented in the transfer phase. In the training phase, participants learned to classify the training set of pattern distortions into the three categories. Participants received feedback after each trial, and the task continued until they reached a set criterion level of performance. The transfer phase required participants to classify, without feedback, some of the "old" patterns from the training phase, the three category prototypes, and several new patterns that were also members of the learned categories. Their basic finding was that in the transfer phase, participants were better at classifying prototype patterns than new exemplars of the learned categories. Participants classified the prototypes equally as well as the old exemplars even though the prototype patterns never appeared as part of the training set. Posner and Keele interpreted their results as being most consistent with a system that stores both the abstract prototype category representations and the individual exemplars. Indeed, the prototype effect that Posner and Keele obtained in this study has been replicated (e.g. Homa, Cross, Cornell, Goldman, & Shwartz, 1973; Homa & Cultice, 1984; Palmeri & Nosofsky, 2001) and cited many times as offering strong evidence for the storage of abstract representations (e.g. Homa, Sterling, & Trepel, 1981; Rosch, 1977).

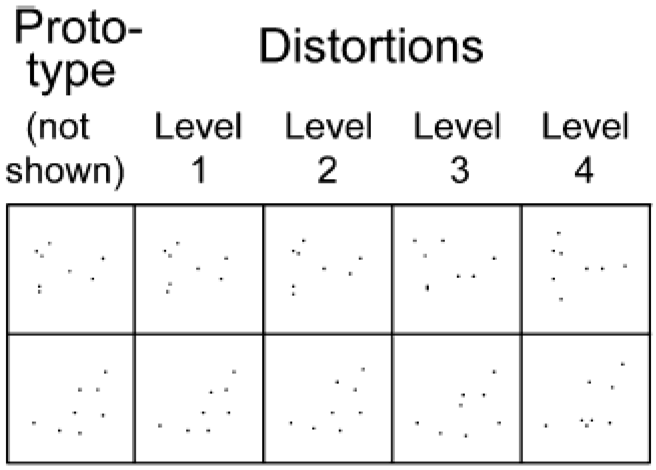


Figure 1.5. Examples of two prototype random dot patterns along with exemplars across four levels of distortion (taken from Winkielman, Halberstadt, Fazendeiro, & Catty, 2006).

Posner and Keele (1968) argued for a mixed exemplar plus prototype representation system based on two main findings. Two experiments they reported manipulated the level of distortion used to generate the training patterns. Half the participants trained on low level distortions, and half the participants trained on high level distortions. Both experiments found that after training to criterion on high level distortions, participants subsequently showed better classification transfer to new exemplars of the learned categories compared to participants who trained on the low level distortion exemplars. The new distortions were medium level distortions. If only the category prototype was stored, new distortions should be classified equally as well in high and low distortion training groups. This is because, between the two groups, the prototypes and the new distortions were the same, so the only difference was the distortion level of the original training exemplars; therefore, if the old training patterns were not stored, then the contents of memory would be the same between the high and low distortion groups. Furthermore, if within category variability did affect the prototype representation in some way, one might expect abstraction to proceed with great ease and encoding precision when learning “tight” categories as opposed to more “loose” categories. Since it was the training on highly variable category members that led to increased generalization to new category exemplars, information about the individual training exemplars must have been stored. Posner and Keele rejected a purely exemplar-based representational system on the basis that original training patterns were controlled to be equally similar to both the category prototype and the new distortions. The authors argued that if prototypes were recognized better than new distortions, then it must have been because the prototypes were being abstracted and represented in some fashion.

Posner and Keele (1970) also investigated whether the prototypes they believed were being stored in memory during classification tasks in their prior experiments (Posner & Keele, 1968) were in fact abstracted during learning. They tested whether classification of prototype patterns would show less of an effect of forgetting than the old exemplars participants were trained on. Half the participants were tested on the transfer patterns immediately, and half were tested one week after they were trained. The pattern of results did indeed show evidence of differential forgetting. This effect was only significant on the first run of the transfer pattern list (the list was repeated four times in transfer phase). In immediate testing, old patterns were better classified than prototypes, but after one week, olds were about equal to prototypes. The authors argued that this pattern of results was strong evidence for the abstraction of prototypes during learning claiming that memory for central items is more enduring that memory of peripherally stored items (in some sort of perceptual pattern space). According to their reasoning, prototypes must have been formed during learning because if they were formed during retrieval, memory deficits on old items should be associated with memory deficits on prototype items since prototypes would be recognized through generalization from the now weakened stored exemplars.

However influential the work of Posner and Keele (1968, 1970) may have been in the advancement of theories of concept abstraction and categorization (e.g. attribute-frequency models (Neumann, 1974, 1977; Medin & Schaffer, 1978), there are some critical flaws in the methodology which render their interpretation of the results unfounded. I will briefly touch on a few. First is the obvious problem of training participants to criterion before the transfer phase. Since learning to classify is more difficult when the category exemplars are highly variable, more learning and more practice will be required by participants in the high distortion training group compared to the low distortion group. Better generalization to new category exemplars in the high distortion training group could simply be a consequence of those participants having undergone more demanding training procedures. Hintzman (1986) also has pointed out that, in these experiments, variability of within-category exemplars is confounded with degree of learning.

Accounting for differential forgetting of old items over prototypes is challenging, but not impossible, for a purely-exemplar based model. Medin and Schaffer's (1978) context theory of classification is able to predict the differential forgetting effect. Recognition of old items relies heavily on a close match between the old item as a probe and its representation in memory from a previous encounter. In contrast, recognition of the prototype does not rely so much on being very similar to any one trace in memory. Rather, it is likely to be moderately similar to many stored traces and recognized on that basis. Forgetting can cause the contents of stored representations to degrade which will prevent old items from matching closely to their stored representation after a delay. This will have less of an impact on the prototype because it only needs to be moderately similar to many traces, and after a delay it likely still will be. The nonlinearity of the similarity metric has a lot to do with this because small changes in similarity of highly similar items will affect activation levels much more than small changes in similarity of items that are only moderately similar to start out. Delay has the effect of flattening out the curve associated with the similarity-proximity function. This means that in immediate testing, recognition is dominated by items that are highly similar. Traces of moderate similarity to the probe have more influence on recognition after a delay relative to their influence during immediate testing.

Hintzman's Minerva II (1986), an episodic exemplar-based model, can also account for differential forgetting by essentially the same reasoning. In immediate testing, old items have an advantage over prototypes in the quality of information (is able to activate strongly a trace with near identical feature values thus the echo content will be very clearly a match, see section on models for more information), but after a delay, when the quality of the stored exemplars is impoverished, prototype recognition has the advantage due to redundancy in the number of traces that match on each property.

Rejecting an exemplar view due to prototypes being classified better than new exemplars despite having equal similarity to the training patterns is also premature. By virtue of being the central tendency of the exemplars of the same category, the average distance (in a Euclidean similarity space) between a category prototype and its constituent exemplars must be smaller than the average distance between any given new distortion and the exemplars of the same category. This is important because many current models compare the similarity of the-to-be classified (or recognized) stimulus (which could be a face) to multiple stored exemplars or even to all stored exemplars rather than to just one. The bottom line is that although the prototype advantage over new category members observed in visual classification paradigms was originally thought to be evidence that some sort of abstract prototype is stored in memory (perhaps in addition to the individual exemplars), more recent endeavours aiming to explain this effect within a purely exemplar-based framework have been met with considerable success (e.g. Nosofsky, 1988; Palmeri & Gauthier, 2004).

Prototype Effects Using Other Visual Forms or Patterns

While random dot patterns have been the most widely used stimuli because their properties can be precisely controlled, investigations of prototype effects using other visual stimuli have been made. Geometric forms have been focussed on in several studies (e.g. Franks & Bransford, 1971; Homa, 1978; Homa, Goldhardt, Burruel-Homa, & Smith, 1993; Homa, Sterling, & Trepel, 1981; Solso & Raynis, 1979). Conclusions are difficult to draw. In general, the prototype form or shape is recognized more than new forms, but not necessarily more than old forms.

Prototype Effects in Verbal and Non-Visual Memory

The prototype effect has been used to gain insight into how words are represented and recognized. Interestingly, the impetus for examining memory for prototype words was to explain why words are sometimes falsely remembered rather than to explain how they are correctly remembered. Using the standard experimental procedure introduced by Deese in 1959, Roediger and McDermott (1995) read aloud to participants lists of words that are highly related to a nonpresented word. For example, the study list could contain: thread, pin, sewing, sharp, point, haystack, pain, and injection. In the test phase, participants often recalled or recognized the word “needle” as it is the common associate of the words in the study list. The rates of recall (~50%) and recognition (~80%) of prototype words were approximately equal to those of the actually studied words[1].

Roediger and McDermott (1995) also found that participants will just as often report having conscious recollection of the prototype word compared to the studied words in a remember-know task (see Tulving, 1985 for information about the remember-know paradigm). Interestingly, Payne, Elie, Blackwell, & Neuschatz (1996) found that when participants were asked if they remember which of two speakers had earlier spoken the words they produced in free recall, participants reported a speaker on 87% of the unspoken prototype words. In some of the word lists, one speaker read aloud all of the words associated with one prototype and another speaker read all of the words associated with a second prototype. Participants only chose the “correct” speaker of prototype words on 53% of the trials where they reported a speaker (participants left blanks when they were not sure of the speaker). False recognition of the prototype word has even occurred following rapid presentation of words in the study list (Solso, Heck, & Mearns, 1993). Thus, if prototype representations are being formed, they can be formed rapidly and possibly never make it to long term memory in some cases. This tendency to recognize the prototype word may be due to the same kind of mechanism as prototype effects observed in other domains. The key difference is that this type of prototype is based on the meaning of the study words rather than their

---

[1] Some studies similar to Roediger and McDermott’s 1995 study have reported slightly higher rates of recall for studied words compared to prototype words or a strong but nonsignificant trend in that direction (e.g. Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996). Therefore, the general finding is that studied and prototype words tend to show similar rates of responding.

appearance (although the prototype need only be an associate of each of the studied words, but not necessarily similar in meaning). There is some evidence to suggest that prototype effects using associated word lists are related to prototype effects using visual forms. Category size (Shiffrin, Huber, & Marinelli, 1995) and differential forgetting effects (Payne, Elie, Blackwell, & Neuschatz, 1996) show the same pattern of results for both words and visual forms (e.g. random dot patterns, schematic faces, etc.).

Words can, however, show non-semantic prototype effects. Using lists of phonologically similar words, Schacter, Verfaellie, and Anes (1997) found essentially the same pattern of results as Roediger and McDermott (1995) did using semantically associated words.

In addition to the auditory modality, prototype effects have be observed among kinaesthetic movements by Solso and Raynis (1979). In the study phase, the experimenter would move the participant’s arm to learn 10 exemplar movements. This task was almost like mimicking the movement of drawing a figure on a blackboard. In the test phase, the experimenter again moved the arm of the participant. Now the task was to rate how confident you are the movement was one of the training movements. Even though the prototype movement was not part of the study phase, there was a striking increase in confidence for the prototype compared to both old and new movements. Solso and Raynis concluded that kinaesthetic movements are represented in memory by a set of prototypical movements. Their claim seemed to be based merely on the observation that the prototype of the studied series of movements was highly, and falsely, recognized. Old movements were also recognized more confidently than new movements. Likewise, Solso and Raynis interpreted this finding as evidence that old items were also being stored in memory in addition to prototypes. This is a good example of an assumption behind the logic of many recognition experiments. The assumption is that something is recognized because it is like something stored in memory. Thus we recognize previously unseen prototypes because they are “like” the representation we have stored in memory. No one would disagree with this assumption altogether. The major theoretical difference is whether the stimulus can be recognized because it is like many memories rather than like just a single memory.

The Prototype Effect in Face Recognition

A typical experiment investigating the prototype effect in face recognition is divided into separate study and test phases. In the study phase, several sets of variants around different face identities are presented to participants. Unlike in a standard test of face recognition, multiple exemplars of each identity are presented rather than learning each face from a single image. The average of a set of seen exemplars corresponds to the prototype stimulus for that identity. In the recognition test phase, a mix of prototypes, seen exemplars, and unseen exemplars are presented. The prototype effect has been observed even when no prototypes were presented during the study phase. Thus, it seems that multiple instances of the same person’s face are stored and/or retrieved in such a way that benefits recognition and classification of a person’s prototype face (Cabeza et al., 1999).

Prototype effects are often inferred as evidence of some sort of averaging or blending mechanism performed among face instances to form a single prototype representation that is being accessed when participants judge whether a face is known to them. Prototype effects in face recognition experiments are often defined as a recognition advantage for prototypes over seen exemplars. Note that studies of categorization and memory illusions use the term prototype effect to describe any pattern of data where the prototype is recognized or categorized more accurately than other non-prototypical new items.

Previous Research of Prototype Effects in Face Recognition

The first experiment investigating prototype effects in face recognition aimed to extend the previous research using dot patterns and geometric figures to a more natural and complex class of visual stimuli, faces. Solso and McCarthy (1981) used Identikit faces (Identikit is a device used by police to construct a likeness of a suspect's face) which are constructed by selecting different of combinations of line drawn features from a premade pool. Three prototype faces were constructed such that none of them shared any features with either of the other two prototypes. Distortions of each prototype were made by swapping one or more features with a different feature from the pool (e.g. the mouth selected in the original prototype could be swapped with a different mouth from the pool of available mouths). The four face feature types that were varied were: hairstyle, eyes plus eyebrows, nose plus chin, and mouth (see Figure 1.6). Distortions were made swapping 1, 2 or 3 features of the prototype with new features (75%, 50%, or 25% distortions respectively). Distortions never shared any features with other distortions that were not shared with the prototype faces. Participants attempted to memorize three 75%, four 50%, and three 25% distortions of one of the prototypes. A recognition test was given after a delay of either 5 minutes or 6 weeks. The recognition test included a Yes/No decision as to whether the test face had been presented in the study phase followed by a confidence rating. The stimuli in the recognition test phase included the never seen prototype and both old and new distortions at each of the three levels (1, 2, or 3 features different from prototype) as well as one face that shared no features with any other faces. Results showed highest recognition confidence for prototype faces compared to both old and new distortion faces. Old faces were called “old” more than new faces. The pattern of results was the same regardless of the length of delay. In fact, the 6 week delay did not seem to have much of a detrimental effect on memory performance for any of the stimuli types (which may be why no analyses of differential forgetting effects were reported). The authors concluded that because the prototype face was recognized more than the old and new distortions, the prototype must have been stored in memory. A similar study using essentially the same materials and procedure as the immediate testing condition was conducted on children aged 3-6 years old (Inn, Walden, & Solso, 1993). The same pattern of results was found in 6 year olds as was found in adult participants. However, children younger than 6 did not falsely recognize the prototype face more than old and new faces. The younger the child was, the less they recognized the prototype.

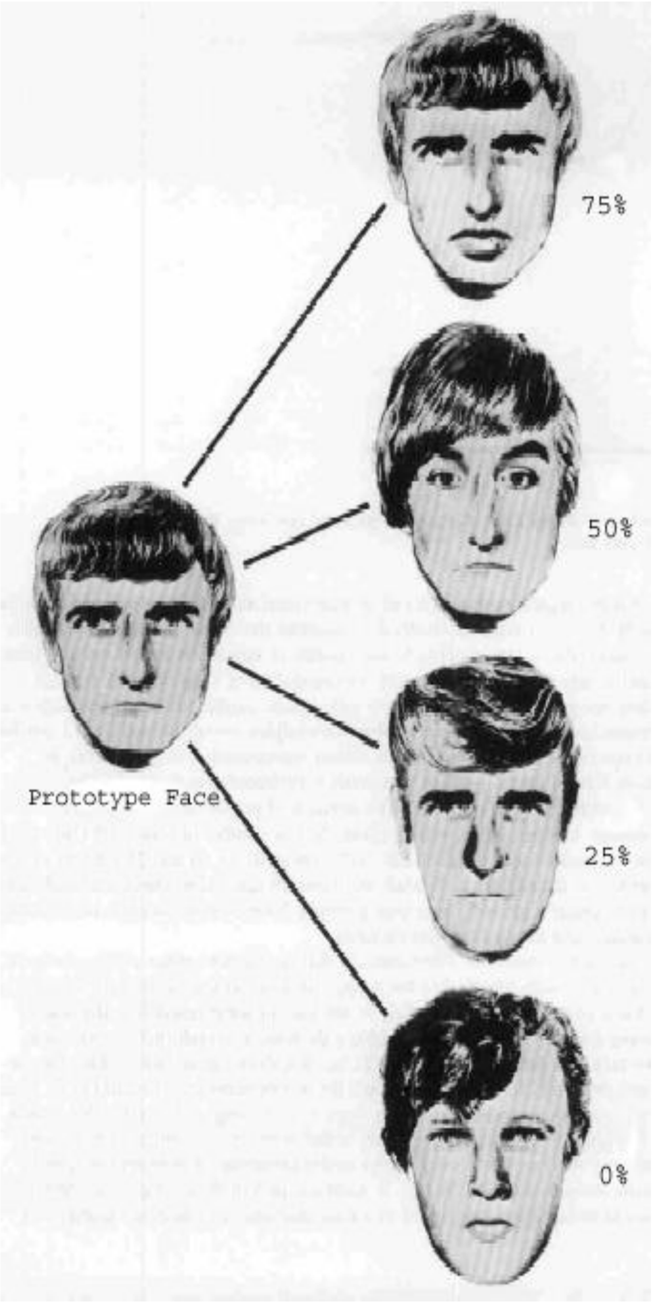
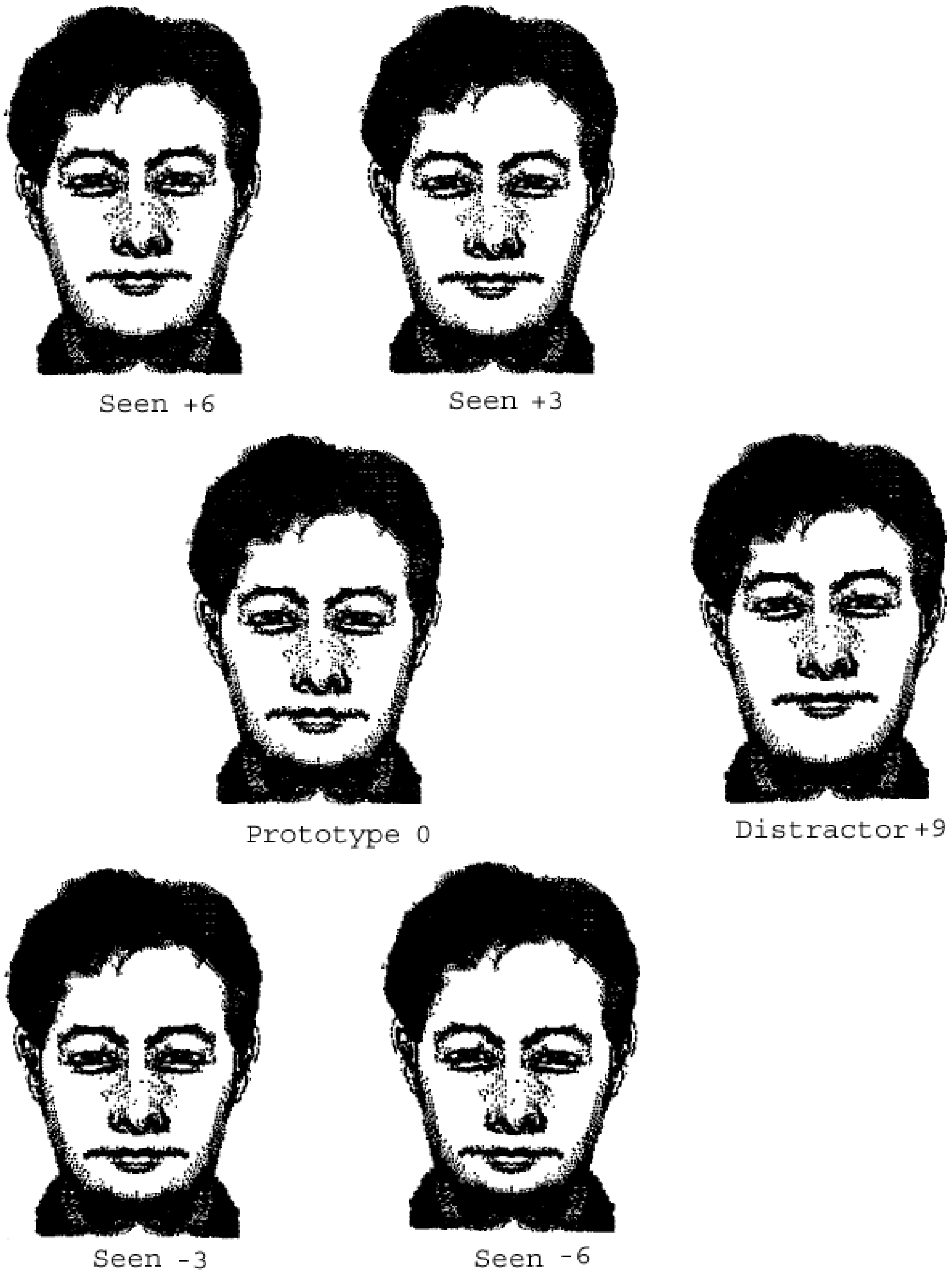


Figure 1.6. Example stimuli from Solso and McCarthy's 1981 experiment.

These abstract prototype memories are thought to be functionally similar to memories of the actual experiences, but may provide a more salient representation of a face during retrieval (Solso & McCarthy, 1981). The finding that old faces were still able to be distinguished from new faces was interpreted as support for the view whereby a prototype is formed and stored in addition to the memories of individual experiences. Just why these prototypes are more salient or how they are formed in the first place was not answered. It is likely that the task demands are partially responsible for the observed pattern of results. The only thing distinguishing the old and new faces (excluding the prototype) was that in the test phase, old faces had no entirely new features, and new faces had at least one feature that had never been seen at any time before the study phase or earlier in the test phase. This would lead to more of a reliance on a frequency based method of recognition because the task could more or less be reduced to a frequency judgement on each feature where the choice was between zero or more than zero. The higher frequency of presentation of prototype features both in the study phase and from previous trials in the test phase protect those features from being judged as new features. The swapped features in the old items were seen only once in the study phase, so they have a higher chance of being judged as entirely new. The more swapped features an old face has, the higher the likelihood that the face as a whole will be judged as new. Likewise, the new features in the new faces are actually entirely new features, so they are even more likely to be judged as new features. The more new features a new face has the more likely it will be called "new". However, this study provided an intriguing initial result for further enquiry into the nature of prototype representation of faces in memory.

One of the most important contributions to the investigation of prototype effects in face recognition was made by Bruce, Doyle, Dench, and Burton (1991). The results of several experiments were reported. They used Mac-a-Mug faces as stimuli which are very similar to Identikit faces in appearance (line drawn) and in the way they are constructed (selecting features from a pool). However, in Bruce et al.'s experiments, the distortions around the prototype were not produced by swapping facial features but by varying the configuration (i.e. placement) of the features on the face outline. Their initial experiments shifted the feature placements up or down by varying amounts to mimic an aging transformation in a crude way. They produced 4 exemplars around each of the 10 original prototype images which had no features shifted. Participants made incidental ratings of the faces in the study phase and completed a forced choice task in the test phase which followed immediately after the study phase. The forced choice task always pitted two images against one another that were based on the same original face. The distractor was an unseen version of that face which had all of its internal features shifted up or down as a single unit by 10 pixels. Results found that when the prototype image had been seen in the study phase (in addition to the four other exemplars), participants picked the prototype at a level nearing ceiling (90%). When the prototype was not shown in the study phase, participants still chose the prototype more often than the distractor (79%). Seen exemplars were picked ~70% of the time against the same type of distractors.

The remainder of Bruce et al.'s (1991) experiments constructed the variations around the prototypes by simply shifting the internal features as a unit up or down on the face outline (see Figure 1.7). This transformation left the configuration among the internal features intact, but it affected how high or low the internal features sat on the face shape. Since the experiments so far all used the original image as the prototype, they repeated the procedure but this time shifted the features of the prototype and built the study phase exemplars around the now shifted prototype. The original unshifted image was used as the distractor. A prototype preference (82%) was found on the forced choice task even when the prototype was now an altered image and the distractor was now the original image. The final forced choice experiment pitted a seen exemplar against the unseen prototype which resulted in chance responding (prototypes=54%). The final two experiments switched to an image recognition confidence rating task at test. They also compared intentional vs. incidental learning and whether exemplars of the same face were presented consecutively in the study phase or not. Three exemplar shift patterns were used to create the stimuli. The first two, old and young, were essentially equivalent. The only difference was that the features of the old faces were shifted up by 3, 6, 12, 15 and the young were shifted downwards by the same amounts. The third type of exemplar pattern, the bimodal condition, was studied as two repetitions each of two different images (up 9 and down nine pixels). Results indicated that learning instructions affected confidence ratings in the bimodal condition only. After seeing the -9 and +9 exemplars two times each, the prototype effect was stronger for incidental learners compared to intentional learners. Whether the exemplars of a particular face were presented consecutively during the study phase or not made a difference only to the young and old conditions and not the bimodal condition. For young and old faces, there was a stronger prototype effect when the exemplars were randomly distributed throughout the series of study faces rather than blocked by identity set.



Seen +6 Seen +3

Prototype 0 Distractor +9

Seen -3 Seen -6

Figure 1.7. Internal facial features have been shifted up and down on the face outline in increments of 3 pixels (Adapted from Bruce, 1994).

That study instructions and blocking, in Bruce et al.'s (1991) experiment, affected recognition differently depending on whether the exemplars followed a 3, 6, 12, 15 pattern or a -9, -9, +9, +9 pattern is very interesting. This result demonstrates that participants' readiness to falsely recognize the unseen prototype of a set of related exemplars is influenced not only by what they see but also by the circumstances of learning. The explanation behind this interaction is not obvious. However, it provides the grounds for examining the effects of similarity relations among exemplars of the same person's face on empirical measures of prototype abstraction. This will not be an easy mystery to unravel. Consider that the bimodal and old/young conditions produced comparably strong prototypes effects even though the +/- 9 exemplars actually were seen before in the bimodal condition but were unseen in the old/young condition. This alone would lead one to expect smaller prototype effects in the bimodal condition. On top of that, the prototype was less similar to its studied exemplars in the bimodal condition which again might lead one to expect prototype effects to be weakened.

Bruce et al.'s 1991 work was an important contribution to this area of research because it was the first study to use complex face stimuli (i.e. not schematic faces) to examine prototype abstraction using the natural category of identity. Previous research using complex faces imposed artificial categories onto the set of faces. This meant that participants were learning what a group of different peoples' faces had in common that made them different than another group of people's faces. Obviously it is more informative for face recognition researchers to examine recognition of identity or prior exposure.

Another notable finding from Bruce et al.'s (1991) work was that participants were not learning the prototypical configural pattern of all faces in the study phase. Rather, they showed sensitivity to the prototype configuration specific to each identity. Participants were not trained to classify the exemplars into identity based categories, yet their strong responses to the prototypes of these identities indicates that the abstraction taking place was based on the exemplars of shared identity. Thus, a prototype abstraction mechanism would seem to be most useful if it averaged across exemplars of the same person but not across exemplars of different peoples' faces.

Cabeza et al. (1999) attempted to extend knowledge of the prototype effect in face recognition using more realistic stimuli and experimental tasks. They reported a series of experiments which followed the same basic methodology. The experiments all had two phases: study and test. In the study phase, participants made incidental ratings about the appearance of each face. The test phase asked participants to either rate their confidence of having seen that exact image of the face in the study phase or to make an Old/New judgement as quickly as possible about the person's face in general. In the task of rating recognition confidence of the specific exemplars, all the identities had been seen before in the study phase. Test stimuli included both new and old exemplars of each studied identity. In the Old/New task, a mix of old and new identities were used. Both types of test phases presented each studied identity in three different versions: old exemplar, new exemplar, and the unseen prototype.

The first experiment manipulated feature locations as in Bruce et al. (1991). However, Cabeza et al. (1999) used high quality colour photographs of real people's faces. Internal features were shifted up or down using Morph software which does not leave a noticeable joint at the edges. Results showed prototypes were recognized more than both old and new exemplars. This was true in both the exemplar recognition task and the Old/New face recognition task. Thus, even though, intuitively, the demands of the two tasks seem quite different, they both produced the same pattern of results.

Cabeza et al. (1999) found no effect of prototypical view on the prototype effect (front vs. 45°-left view). They also found some evidence that the prototype effect is more sensitive to changes in viewpoint angle than to variation in feature locations within a single view. Cabeza et al. argued that an averaging mechanism was likely operating on instances of the same face from the same viewpoint but not across different viewpoints. Instead, approximation to a limited set of stored views might be needed to recognize a face from different viewpoints.

Cabeza et al. (1999) also used morphing technology to vary the shape and texture of the face features elastically. This manipulation was a novel approach, at the time, to investigating prototype effects in face recognition. They used the Morph software to vary the similarity among exemplars of the same face identity. Results showed that prototypes were recognized as well as seen exemplars when the studied exemplars were similar enough to plausibly be perceived as the same person. However, when the studied exemplars (of the same identity set) no longer resembled the same person, the prototype was recognized less than the seen exemplars.

Busey and Tunnicliff (1999) provide partial support for the existence of a prototype mechanism that operates on instances belonging to different people's faces. They used morphing to create a controlled set of stimuli of photographic quality that could be used to compare competing accounts of the prototype effect in face recognition. Their method was to blend, or "morph", two images of different individuals together to create a 50%/50% morph that was equally similar to both of the original images, termed "parents". As exploring the role of similarity among exemplars was of particular interest, the perceived similarity among the original faces was assessed and used as a guide to choose the parents in each pairing. Similar and dissimilar parent pairs were selected. This method resulted in a tendency for the dissimilar parent faces to look more distinctive than similar parent faces. Results showed that after participants studied just the parent faces, the false positive rate to the prototype was higher than the hit rate to the parents, but this was true only for similar looking parents. Dissimilar parents, on the other hand, were recognized more than their associated prototype morphs. Pure exemplar-based models had difficulty predicting morphs of similar parents would be more recognized than their parents whilst simultaneously predicting that dissimilar parents would be more recognized than similar parents. Busey and Tunnicliff concluded that the human data coming from their experiments were better explained by an exemplar plus prototype model. In this model, when encoding a new item, another trace would be laid down at the averaged location between the new item and any sufficiently similar previously stored items. In this case, the more similar the parent faces are to one another, the stronger the memory trace of the abstracted prototype.

When comparing the adequacy of exemplar-based vs. prototype models in explaining their experimental results, Busey and Tunnicliff (1999) treat a purely exemplar approach as the null hypothesis. If their data cannot be adequately explained by a purely exemplar-based model, then the pure exemplar approach can be rejected in favour of a mixed model. If a mixed model is found to provide a better account for the human data, then this can be taken as evidence for the abstraction and storage of prototypes in memory. This was in fact what the authors reported. However, caution is urged when comparing theories based on tests of a limited number of models that do not represent the full range of models within those particular theoretical perspectives.

Although Busey and Tunnicliff mention it, it should be emphasized that the exemplar approach was rejected based only on models that used summed similarity to predict recognition scores. They tested three exemplar models which were variants of either Nosofsky's (1986) generalized context model (GCM) or the search of associative memory (SAM) model (Gillund & Shiffrin, 1984).

Consider a purely exemplar-based model that computes the "oldness" of a probe by summing the activation it produces across all stored exemplars in memory. If activation of a stored exemplar is dependent on its similarity to the probe, then the function that relates activation to similarity is very important. If a probe produces considerable activation of exemplars that are moderately similar, predictions about prototype effects will differ dramatically from a situation where activation by a probe is effectively limited to only extremely similar exemplars. This is especially relevant to Busey and Tunnicliff's (1999) work because the multi-dimensional scaling (MDS) solution used to model similarity relations was based on pairwise similarity ratings made by participants after they had completed the experiment. This might make it difficult for an exemplar model to predict prototype effects if preexposure to the stimuli reduces the probability of false alarming to the prototypes (e.g. familiarity with faces could affect the scaling parameter c, or possibly the attentional weights applied to the individual dimensions; model parameters are explained in a later section).

In response to Busey and Tunnicliff (1999), Zaki and Nosofsky (2001) used the exact same stimuli in a mock subliminal recognition test. Black rectangles were presented so quickly that all participants could detect was a flash on the computer screen. Participants were told before the start of the mock subliminal phase that faces were going to be presented, but they would only appear as a flash. Nonetheless, they should attend to the stimuli because a recognition test would follow. Of course, in reality, no faces had been presented in the subliminal phase. In the test phase, faces were presented one at a time, and participants were asked to rate how likely the face had been seen before in the previous phase. It was suggested to participants that they should base their ratings on things like gut instinct or a sense of familiarity. Results showed that faces that served as the morphs in Busey and Tunnicliff's experiments were rated higher than the faces that served as the parents. Thus, Zaki and Nosofsky called into question Busey and Tunnicliff's use of explicitly stored prototype traces in their modified exemplar model to account for higher oldness ratings to morphs compared to parent faces. Again, caution is needed when interpreting prototype effects as evidence of stored prototypes. Zaki and Nosofsky obtained a prototype-like effect without any possibility of an abstraction process operating in the study phase. Yet, the morph advantage in Busey and Tunnicliff's study was deemed to be unaccountable unless storage of prototypes was assumed.

Averages Vs. Prototypes

Confusion arises from the use of the term *prototype* in norm-based models. Norm-based models are a subset of models based on a multidimensional Euclidean face space (Valentine, 2001). Norm-based models are usually discussed in contrast to purely exemplar-based models. In norm-based models, the prototype face is the average of *all* faces in the space and is located at the origin. Faces are encoded relative to this norm face. This is quite different to what I refer to as a prototype account because a norm-based model of face space does not explicitly address the issue of how our cognitive system represents a stimulus which constantly varies slightly in appearance. Rather, in face space different facial identities are represented as a single point in multidimensional space (pure exemplar-based model) or as a vector between that point and the origin (norm-based model). A face space made from a Voronoi diagram does conceptualize a face identity as a region of variation, but these regions are constructed based only on a single image of each known face stored in memory (Lewis & Johnston, 1999). These types of face space models do not conflict with a prototype model since one could think of the faces stored in the space as being the product of some abstraction mechanism that produces a prototype for each known face. The main thing is to remember that this thesis concerns the recognition of prototypes of different identities not the recognition of faces varying in distance from the overall average, or ‘origin,’ of face space (i.e. typicality effects).

Another ambiguity in the discussion of prototype effects is what exactly is meant by the term “prototype”. Initially, I defined it as “the central value of a set of varying exemplars”. Indeed, almost all work in the area of visual perception conceptualizes prototypes within this description. However, prototypes in studies of prototype effects in face recognition are often assumed to be the average face of a set of varying exemplars. If prototypes really are represented internally, they do not necessarily have to correspond to the average value of a set. Maybe prototypes correspond to the modal value of a set instead (see Neumann, 1974). It is also possible that the mind may use more than one method of representing the central face depending on the circumstances.

Goldman and Homa (1977) investigated whether the value of a prototype represents the means or the modes of the underlying distribution of feature values of the instances. They found stronger evidence for a modal prototype over an average prototype. However, when the differences between individuals were made unclear, participants responded more to an average prototype (see also Homa et al., 2001).

Conjunction Faces

A related line of research has been conducted using *conjunction faces* (Hannigan & Reinitz, 2000; Reinitz & Hannigan, 2001; Reinitz, Lammers, & Cochran, 1992; Reinitz, Morrissey, & Demb, 1994). Conjunction faces are face stimuli that have been constructed by combining half of the features from one studied face with half the features from another studied face. The result is a face whose features are all “old” as they have all been seen in the study phase. However, the features of a conjunction face were never seen in that particular combination before since they were taken from two different faces.

In a typical experiment, Identikit faces are used to make the stimuli as shown in Figure 1.8. Exemplars in the study phase share no features with one another. In the test phase, a recognition test is given. Target faces, conjunction faces, feature faces (half the features are old half are new) and new faces (all features are new) are presented for recognition. The general finding is that conjunction faces are recognized more than feature or new faces. Although conjunction errors are never higher than the hit rate for target stimuli, participants do not treat the conjunction stimuli the same as new stimuli and have difficulty rejecting them as being one of the training faces (Reinitz et al., 1992).

Figure 1.8. Two different Identikit faces are shown in the left and middle positions. On the right is a conjunction of the other two faces (taken from Reinitz et al., 1994).

Reinitz et al. (1994) proposed that attention is needed to adequately store interrelations among features of stimuli in memory. Little attention is needed to store the features themselves. Thus, conjunction errors may be due to retrieval of the features of the studied faces in the absence of retrieving information about their co-occurrence with other features. Reinitz et al. (1994) found that dividing attention during the study phase drastically lowered the reported "oldness" of target stimuli but had little effect on conjunction, feature, or new faces. Using a remember know paradigm, divided attention lowered remember responses to target stimuli down to a level equal with conjunction stimuli. Thus, sometimes an unseen stimulus can give rise to the feeling of conscious recollection if it is composed of features that have been seen.

Reinitz et al. (1994) suggest that faces are stored as features and feature relations separately. When recognizing a conjunction face, a recollective reconstruction process ensues that may give rise to a "transient visual whole" if relational information is lacking.

Reinitz and Hannigan (2001) investigated whether the proximity between two studied faces influenced how likely their combined features would give rise to conjunction errors. They found that conjunction errors did increase when pairs of faces were presented simultaneously. Surprisingly, no proximity effects could be found when pairs of faces were presented one after the other even when the pairs themselves were separated by a time lag from other pairs in the series. Even when participants were instructed to remember which faces went together, sequentially presented face pairs never led to increased conjunction errors compared to conjunctions of faces from different pairs. A clever final experiment in this study shed some light on the lack of proximity effects observed for sequential pairs. Each face in each pair was presented two times. For a pair AB there were two possible presentation orders at study: AABB (repeating) or ABAB (alternating). The key difference between these two conditions is that in the alternating condition, attention can switch back and forth between the two faces in the pair, but in the repeating condition, this is not possible. Results indicated a large effect of attention switching. Alternating pairs gave rise to larger conjunction errors than did faces from repeating pairs. Therefore, proximity effects in the simultaneous condition might have been due to attention being switched back and forth between the two faces on the screen in a similar manner as when face repetitions were presented in an alternating fashion. This attention switching could be important in binding memories.

Two main explanations of conjunction errors have been put forward (Reinitz & Hannigan, 2001). The first is the *familiarity hypothesis*. Old features of the conjunction stimulus give rise to familiarity, but the original faces they occurred in cannot be recalled. Because they seem familiar and the relational information that may have led to a rejection was not recalled, conjunctions are judged as old on the basis of familiarity alone. This could also be understood as an explanation based on the stimulus similarity because familiarity, in this sense, is a function of the summed similarity of the test item to all items stored in memory.

The second explanation of conjunction errors is the *binding (failure) hypothesis*. When two things are bound in memory, they will be retrieved together. Conjunction errors in memory happen when the features of a face are encoded but are poorly bound together, or when information about the binding relationships of features is forgotten faster than the features are forgotten, or when features of different faces are bound together so that they are later retrieved together. The main idea is that when features are poorly bound, they are more vulnerable to being miscombined with features from other faces. It is also possible that features from different faces may be “bundled” in the initial encoding. In both cases, the binding process is thought to have failed because its purpose is to bind the features within a single face together to distinguish it from other faces with similar features.

The familiarity explanation of conjunction errors cannot account for proximity effects as observed by Reinitz and Hannigan (2000; 2001). Thus, obtaining strong evidence of a proximity effect is critical to the debate. Very recently however, Jones, Bartlett, and Wade (2006) conducted a series of experiments using naturalistic face photographs. They found absolutely no effect of proximity using the simultaneous presentation method or the alternating repetition method. Just why Jones et al. failed to replicate Reinitz et al.’s (2001) pattern of results using real face photographs is still an open question. The effect is probably not confined solely to simple line drawn faces though, as McKone and Peh (2006) have replicated proximity effects using naturalistic photographs of unfamiliar faces that were masked to remove external features. It should be noted that the general pattern of old>conjunction>feature(when included)>new was found in all experiments reported.

Megreya and Burton (2006b) found that studying two unfamiliar target faces simultaneously decreased participants' ability to pick out the target face in a line-up task compared to studying one target face. This was found in test of immediate recall and face matching with targets presented concurrently with the line-up. Thus, the authors argue that the presence of another face on the screen had a detrimental effect on encoding a target face. Although the tasks are very different from those used to probe conjunction errors, it is consistent with the notion that impoverished encoding will lower recognition of studied faces.

A study by Cabeza and Kato (2000) reported a similar pattern of results using high quality colour photographs of faces. Conjunction (or "prototype") faces in this case were made from four different studied faces. Two different prototype faces were constructed for each set of four faces: a featural prototype and a configural prototype. The featural prototype was equivalent to the conjunction faces made in the studies by Reinitz and colleagues, in that it was made by combining a different feature (eyes, nose, mouth, and face outline) from each face. The configural prototype was built by averaging four studied faces using a morphing technique. Results indicated that old stimuli were by far the best recognized, but participants were still more likely to report having seen the unseen prototype than to have not seen it. Distractors were prototypes made by the same methods from novel faces, and they were more likely to be judged as new than as old. Featural and configural prototypes were recognized more than distractors by about the same degree. This could suggest several things. Cabeza and

Kato concluded that this must mean that both features and configuration are represented separately in memory by what they call a “dual code”. It is also possible participants may form more than one prototype by different processes or take a different approach to recognize the two prototypes. Of course the simplest explanation is that both featural and configural prototypes were approximately equal in terms of their perceptual similarity to the studied faces.

The study by Busey and Tunnicliff (1999) described in the previous section is important in understanding how the research on conjunction errors relates to that on prototype effects. Busey and Tunnicliff set out to replicate the pattern of results found using Identikit conjunction faces with realistic photographs. Busey and Tunnicliff reason that if the studied faces are subject to some sort of binding or blending mechanism, then an averaged face will provide better evidence of such a mechanism because an averaged face will match more closely with a blended representation than a conjunction face will. It might be helpful at this point to distinguish between composites and compromises. A composite recollection combines intact features of (at least) two different sources and leads to false recognition of conjunction stimuli in recognition tasks. A compromise recollection does not contain the intact features of stored items; rather, a blend of (at least) two sources is recalled as in studies using averaged prototypes (Tanaka & Schooler, 1991). It is not agreed whether these two phenomena can be explained by the same mechanism operating on the same type of representations. However, binding as Reinitz and his colleagues discuss it is very different than blending as in, for example, Metcalfe’s (1990) composite holographic association recall model (CHARM) or in any type of prototype model that represents concepts by their central tendencies. Reinitz and colleagues maintain only that features that are bound together will be retrieved together. The features themselves are stored intact. The miscombination of features from different faces is thought to occur during retrieval.

The relationship between recognition of conjunction and prototype faces is an interesting issue to consider. It circles back to the debate of whether faces are represented in terms of their features or as Gestalten. It helps link prototype studies that used feature swapping (Solso & McCarthy, 1981) to studies that manipulated the spatial relations among discrete facial features (Bruce et al., 1991) to studies using morphed face blends (Cabeza et al., 1999). Additionally, it raises the question of whether prototypes should only form from exemplars of the same person's face or if they are likely to occur between different people's faces at least in some circumstances.

Relationships Among Tasks: Categorization, Identification, and Recognition

Prototype effects have been investigated using two experimental tasks: categorization and, less commonly, recognition. Exactly how these tasks relate to one another is not well understood, but the issue deserves some consideration. Likewise, identification is a task frequently employed in the face recognition literature and shares an interesting relationship with both recognition and categorization.

Identification, recognition, and categorization all require the visual system to overcome the influences of lighting and viewpoint on the initial percept. Beyond that, the visual system faces the additional challenge of having to ignore certain changes in appearance while discriminating among others depending on the task and stimuli. Identification is primarily a task of discriminating similar stimuli, but categorization involves generalizing across different stimuli. Furthermore, categories can be part of larger categories which in turn require even greater degrees of stimulus generalization. Faces provide a good illustration of this. While I can identify a face as my friend Ralph under different viewing conditions, I can also categorize his face being that of a male, a young adult, a Caucasian, or a human being.

In prototype models, categorization can proceed by matching to the stored category central tendency (Posner & Keele, 1968). Recognition of specific exemplars would require additional representations of events (Whittlesea, 1987). On the other hand, in exemplar models, categorization, identification, and recognition can all be achieved using a single representational system of stored instances of exemplars (Kahneman & Miller, 1986; Medin & Schaffer, 1978; Whittlesea, 1987). Categorization and identification are based on the relative similarity of a probe stimulus to stored experiences with exemplars of the same or different categories or identities respectively. Recognition is based on a signal of overall familiarity derived from a probe's similarity to all stored instances regardless of their identity or category membership (e.g. Hintzman's Minerva II, 1986). However, to successfully account for all three tasks, a model must be able to use stimulus dimensions flexibly (Palmeri & Gauthier, 2004). For example, the task demands should either constrain dimensions to ones relevant to the task (e.g. Kahneman & Miller, 1986) or alter the metric of the dimensions (i.e. stretch or shrink multidimensional space) to optimize task performance (Nosofsky, 1988).

Cognitive Models

There are a number of ways the numerous models of categorization and recognition could be divided up for comparison with one another. Since one of the main themes of the current thesis is the nature of the underlying representations stored in memory, I chose this division as a starting point for discussion. Models of categorization as well as recognition are relevant to understanding prototype effects in face recognition, so both classes will be reviewed. Note that the use of the term ‘category’ will be used frequently, and facial identity can be thought of as the category of interest. The term ‘exemplar’ is best thought of as analogous to an example of a particular person’s face rather than representing an entire identity by itself.

Theoretical Context

One of the easiest ways to distinguish memory models is by the representational role of individual instances. This distinction gained importance through two theoretical debates. One of the debates concerns whether memory requires two separate systems or if all aspects of memory can be explained using a single system. The semantic-episodic theory of memory maintains that there are separate memory systems for the storage of decontextualized general knowledge and specific events. In contrast, episodic theories explain the representation of all information in memory as being based on the storage of individual events.

The other debate deals with how experience leads to the formation of concepts. One view is that concepts are formed by summarizing information across individual experiences. This is the basis of prototype theories of concept formation. Alternatively, concepts are not contained within a single representation. Rather, instance theories assert that a conceptual representation is distributed across many stored examples.

The distinction between abstract and event related information in the semantic-episodic hypothesis of memory is echoed in prototype theories of concept formation. Because it attempts to explain all memory through memories of events, the episodic hypothesis mirrors instance-based theories of concept formation (Whittlesea, 1987).

Prototype and Other Summary-based Representational Models

Pure Prototype-based Models

In most prototype models, a category is represented by only a single representation, the central tendency of its members (Elio & Anderson, 1981; Franks & Bransford, 1974; Homa et al., 1981; Neumann, 1977; Posner & Keele, 1968; Reed, 1972). A prototype representation can be abstracted from a series of related or similar exemplars which, under normal circumstances, make up some type of category. Abstraction of prototypes is thought to occur at the time of exemplar encoding rather than at the time of test retrieval. The structure of the prototype representation takes the same form as that of an exemplar, but it may have some special qualities. For example, the prototype may be more resistant to decay due to forgetting (Elio & Anderson, 1981; Homa et al., 1981). For facial identity, the prototype can be defined as the face that has as its value on each dimension, the mean of all its stored instances. Recognition of probe faces is a function of the distance between the probe and the prototype in psychological space.

Feature-frequency distribution models

The next class of models that represents categories or concepts as a summary are feature-frequency distribution models (e.g. Neumann, 1974; Neumann, 1977). According to feature-frequency distribution models, a category is represented by the frequencies with which feature/dimensional values occur across all seen instances of a category. If faces are represented by their values on *n*-features, then a particular identity could be represented by a frequency distribution of all experienced values across these features. Each identity would have its own distribution. Therefore, frequency increments would only be made to feature values of the distribution matching the identity of the probe face. The probability of a probe face being recognized is a function of the frequencies of the feature values of the probe.

One could think of the pattern represented by the category mode on each dimension to be more or less equivalent to a prototype representation using the mode rather than the mean to represent the category's central tendency. There are some differences between these two types of "prototypes". One is that the modal prototype has no special role or status compared with other items. The modal prototype would not even be stored separately; it would just be reflected in the distribution of features. Another difference is that the prototype is not the only thing stored in this model. All experienced feature values are represented and counted by the summary distribution. Therefore, feature-frequency distribution models retain information about the occurrence of atypical category members and the range of experienced values along each dimension in psychological space.

One key difference between the feature-frequency distribution model and the pure prototype model is that in the prototype model, the features can assume values that have never been experienced. In the feature-frequency distribution model, however, the modal "best representative" pattern will not have any components that have never been experienced before. In fact, only the most frequent values of each feature will represent the modal pattern although the particular combination they form may not have been seen before.

Representing faces as frequencies of particular values may not be sensible except in certain experimental tasks. For example, if the variants of a person’s face are constructed by swapping face features (e.g. eyes, nose, mouth etc.), then this model may be appropriate. If continuous variables are used to represent a face (e.g. by principal components from PCA), then the model must be modified slightly so that frequencies within a range of values can be tallied along each dimension (Neumann, 1977). Using slightly overlapping intervals will allow values that have never been seen before to be treated as the most frequent.

Limitations

There are several shortcomings of prototype-based models. Firstly, categories are represented by prototypes that are context free (Medin, 1989). This leaves the effect of context on categorical judgements (e.g. Roth & Shoben, 1983) and recognition (e.g. Steyvers & Malmberg, 2003) unaccounted for.

Secondly, because the features/dimensions of a prototype are independent and typicality is a function of the distance from the prototypical mean, typicality of an exemplar should be the sum of the typicality of its components. However, this is not the case, and one cannot assume the features to be independent. Prototype models fall short when trying to predict responses to stimuli with correlated features (Medin, 1989; Nosofsky, 1986).

Thirdly, storing only a category's central tendency makes it difficult to form new categories on the fly as the need for a new category arises. Barsalou (1985, 1987) has demonstrated that people can form goal oriented categories such as "foods to eat while on a diet". These categories did not seem to rely on the average prototype to determine category membership. Instead, participants used an "ideal" member as reference (e.g. food with zero calories).

The problems related to prototype models stem more from the way similarity is measured rather than the discarding of the individual instances. Even if all of the individual instances were retained in a prototype model by computing the central tendency at the time of test rather than during encoding, the prototype model would still fail. The reason why most prototype-based models fail in numerous ways is that the similarity between probe and prototype is computed as the sum of independently coded features shared between the two (Medin, 1989).

Hayes-Roth and Hayes-Roth (1977) developed a feature-frequency distribution model that did retain information about the correlations among features. This was achieved by storing not only the individual features but all possible combinations of them as well. For example, if the feature values of a 3 feature stimulus were 1, 2, and 3, frequency counts would be incremented for values 1, 2, and 3 on the first, second, and third feature respectively. In addition to the individual features, the combinations 12_, 1_3, _23, and 123 would also be incremented.

Instance-based Models

The Context Model and GCM

Medin and Schaffer first proposed the context theory of classification in 1978, and it has proved to be very influential as a purely exemplar view of category representation. Nosofsky (1984, 1986, 1991) has further developed this model into the generalized context model (GCM). According to the context theory, categories are represented by storing numerous individual exemplars in memory. Classification decisions are made by comparing the similarity of the probe to each of the stored exemplars. If the similarity of a probe to exemplars in one category exceeds that of other contrast categories, the probe will be classified as belonging to that category.

This approach can also be applied to recognition. In recognition, decisions are made based on the similarity to all exemplars summing across all categories. The summed similarity can be thought of as the overall familiarity produced by the probe. The higher the familiarity of a probe, the higher probability is that it will be recognized. Thus, the same representational structure underlies the process of both categorization and recognition, but the decision rules operating on it are different during the two tasks (Nosofsky, 1991).

In contrast to prototype theory, the context theory assumes that all exemplars are stored, and no summary information is explicitly represented in memory. Although all exemplars are stored, they do not need to be stored veridically. Their encoding is variable and depends to some extent on the nature of the learning task, degree of attention, and participant strategies during learning. Furthermore, the context theory does not propose the complete absence of summary information. Indeed, summary information may very well be explicitly represented in memory. The context theory only maintains that this type of abstract summary information is not used to make recognition or classification judgements (Medin & Schaffer, 1978).

One of the most important theoretical differences between prototype theory and context theory is the assumption of independence of the component features/dimensions of a stimulus. As described earlier, prototype theory assumes all dimensions of a stimulus are independent and are summed to measure summary information such as similarity or typicality. In contrast, context theory assumes that dimensions are multiplicative. The main consequence of interest is that high similarity to a particular exemplar will influence decisions more than average similarity to a number of different exemplars.

Nosofsky (1986) extends the context theory of Medin and Schaffer (1978) from only being able to represent stimulus coded as binary-valued dimensions to being able to work with dimensions having continuous (or >2) values. Nosofsky calls the extended version the generalized context model (GCM). GCM uses a multidimensional scaling approach. A multidimensional scaling solution (MDS) can be used to estimate the psychological distance between two stimuli which is a function of their similarity.

How model works:

First, the distance between the probe face and each exemplar stored in memory is calculated. A weighted Euclidean distance metric is used.[2] Let face *i* be the probe face and face *j* be a stored exemplar:

$$d_{ij} = c[\sum_m w_m |x_{im} - x_{jm}|^2]^{1/2},$$

---

[2] An earlier version of GCM used the Minkowski r-metric.

where x_{im} is the psychological value of face *i* on dimension *m; wm* ($0 \leq wm \leq 1$, $\sum$wm = 1) is the weight put on dimension *m* (weights are free parameters); and *c* is a sensitivity parameter reflecting overall discriminability in the psychological space.

The dimensional values used to compute distance come from a MDS solution derived from similarity ratings that participants made about pairs of faces. The solution arranges the faces into a multidimensional space reflecting their "psychological distance". The weights on the different dimensions can be used to model selective attention to particular dimensions. In effect, the weights act to stretch or shrink psychological space along a dimension (Nosofsky, 1986).

The similarity of face *i* and face *j* is found by converting the distance between them using an exponential decay function (Shepard, 1958, 1987):

$$s_{ij} = \exp(-d_{ij}).$$

Activation of face *j* (a_{ij}) by face *i* is equal to their similarity plus noise. A random normal variable (mean=0) models the random noise that might affect the level of activation of a face stored in memory:

$$a_{ij} = s_{ij} + e_j,$$

Since GCM is designed to model both categorization and recognition, overall activations are calculated by taking each category in turn. The sum of activation produced by a probe face within a single category is referred to as the category's evidence. In a two category case, the evidence for Category 1 given presentation of face *i* is the sum of the activations of all of Category 1 exemplars stored in memory. The evidence for Category 2 is calculated in the same manner.

$$E_{1,i} = \sum_{j \in C1} a_{ij}$$

In a classification task, the evidence difference between the two categories is compared to a criterion and if the difference is large enough a classification is made:

$$E_{1,i} - E_{2,i} > b$$

where the critical amount is *b*, a response bias parameter.

For recognition, the evidence for Category 1 is calculated separately from Category 2 which is also calculated, and the sum of the two values is compared to a criterion. In a face recognition task, each different identity could be considered a category, but I am sticking to the two category example for ease of communication. If the summed activation for both categories exceeds a criterion x_c, an "old" response is made:

$$E_{1,i} + E_{2,i} > x_c$$

Episodic Models of Memory

All episodic models of memory take memory of events as the basis of all processes in memory. Every experienced event lays down its own trace in memory. Thus, long term memory can be seen as a vast collection of episodic traces. Another very important property of episodic models is that retrieval activates multiple events in parallel. The strength of activation of a given memory trace is a function of its similarity to the probe. The encoded dimensions of a stimulus are not assumed to be independent, but their exact relationship varies somewhat among different models. It is more important to grasp the flavour of the episodic approach than to note all of the details and possibilities. That is that there are no stable representations in memory, and it is the *flexibility* of encoding and retrieval that make the processes of memory so dynamic.

Hintzman's Minerva II

Hintzman's Minerva II (1986, 1988) is a good example of an episodic model of memory. To facilitate discussion of this model, it is useful to distinguish primary memory from secondary memory. Primary memory is the representation of the current experience (i.e. the probe). Secondary memory consists of the vast pool of episodic traces representing previously encoded events.

The traces in this vast pool are represented in Minerva II by vectors. Each event is encoded as a unique vector which codes the feature values of each feature/dimension. Like most other instance-based models, encoding the correct feature values on each feature is probabilistic. The representations are quite simple, in that they only code for the presence or absence of features assigning a value of +1 or -1 respectively. A value of 0 is also possible, and it indicates that the feature was not encoded or is irrelevant.

According to Minerva II, when a probe is encountered, primary memory sends a signal of its contents to secondary memory where it activates the pool of stored traces. All stored traces are activated in parallel to some degree. The more similar a trace is to the probe, the more activated it will become. Secondary memory then sends a signal, termed the “echo,” back to primary memory containing the result of the retrieval process.

The echo has two properties that carry different summary information about the retrieved traces. The sum of activations of all traces in secondary memory provides a signal of familiarity termed the *intensity* of the echo. The echo intensity serves as the basis for recognition memory judgments. The other property of the echo is its *content*. This summarizes the actual content of the traces that were retrieved. A given trace will contribute more to the content of the echo the more it is activated by the probe. Thus, the echo is like a composite made from the most similar traces in memory.

How the model works:

Because the vectors only code for presence or absence of a feature, the similarity between the probe and a trace can be computed easily. That is, on every feature, the values will either match, mismatch, or will not be relevant or comparable (if either vector has a zero value that feature is deemed irrelevant). Give matching values a +1, mismatching values a -1, and irrelevant comparisons a 0. Sum these values across all features of the vectors to get the total similarity of the probe to a given trace *i* as in the equation:

$$S_i = \sum_{j=1}^{N} \frac{P_j T_{i,j}}{N_i},$$

where P_j is the value of the j_{th} feature of the probe, $T_{i,j}$ is the value of the j_{th} feature in memory trace *i,* N_i is the number of features for which either P_j or $T_{i,j}$ is nonzero, and N is the number of features in the trace and test item being compared (Hintzman, 1988).

The similarity of a trace to a probe is transformed into the activation of the trace by cubing the similarity:

$$A_i = S_i^{3},$$

where S_i is the similarity of trace *i* to the probe and A_i is the activation of trace *i* in response to the probe (Hintzman, 1988). The sum of activation values across all traces in memory gives the echo intensity as described earlier:

$$I = \sum_{i=1}^{M} A_i,$$

where I is the echo intensity and M is the total number of memory traces (Hintzman, 1988).

The content of the echo is given by a vector of feature values. Imagine traces in secondary memory answering or “echoing” primary memory in response to a probe stimulus. All traces echo back their values on each of the relevant features. The more strongly activated traces will produce a louder echo, thus contributing more to the echo content. The value of the jth feature of the echo content vector can be calculated as:

$$C_j = \sum_{i=1}^{M} A_i T_{i,j} ,$$

Across all features, the echo content represents the information retrieved from secondary memory in response to a particular probe stimulus. Strongly negative or positive values indicate a feature that is shared by many highly activated traces whereas low values indicate that traces were inconsistent in respect to that feature’s value.

Norm Theory

Kahneman and Miller’s (1986) norm theory is another good example of an episodic approach to memory and cognition. Although norm theory focuses more on comparative judgments, the basic principles can easily be applied to a broad range of cognitive tasks. I include a summary of it here because it emphasizes the need for flexible mental representations, and it explains how an event based system of memory can achieve this.

Norm theory assumes that memory consists of a vast pool of episodic traces, and these traces can be rapidly recruited in parallel by a probe stimulus. A *norm* is computed in response to a probe, and it is an aggregate of the recruited episodes. The most important point is that norms are computed ad hoc, so every event brings its own frame of reference to mind. Norms contain information about a stimulus based on similar experiences from the past. If the probe is a category name, a norm might contain useful information about the properties of this category. A face classification task (face/no face) can be used to illustrate the function of norms. To do the task, one must refer to some representation of what a human face ought to look like. A norm can be computed for human faces, but this norm takes as its frame of reference the specific face stimulus being presented as well as the context of deciding whether the stimulus is a face or not a face. Therefore, a slightly different norm would be referenced on each trial. A scrambled face would deviate extremely from a norm for human faces and would likely be judged as not being a face. In contrast, a typical face would match extremely well with its norm and would quickly be judged as a normal looking face.

How the model works:

What most exemplar models would call exemplars or instances, norm theory calls *elements*. In norm theory, however, elements are the representations that are recruited by a probe, and the concept does not generalise to also include the existence of representations in memory independent of retrieval triggered by a probe. A probe recruits a set of elements termed an *evoked set*. Each element in the evoked set can be described by its feature values on different attributes (i.e. dimension). Activation of an element will distribute activation across a range of values. Elements are activated in parallel and to varying degrees. The shape of the activation distribution reflects the generalization gradient across feature values and imprecise assigning of feature values to elements. The area of the distribution associated with each attribute depends on how much the probe activates the element.

Now for each attribute, the activation of each feature value is summed across elements. This produces an aggregate distribution of activation for each attribute which reflects the weighted contribution of each element in the evoked set. These aggregate distributions serve as norms that summarize the evoked set along each attribute.

Any type of stimulus together with its context could serve as a probe. Category references could also be used as probes. These are often important when making comparative judgements. For example, in a laboratory setting, the comparison arising from the question “does this person look more like a male or a female?” The words *male* and *female* would likely serve as probes, each creating its own category norms. The image of the face would also serve as a probe and would perform the useful role of constraining the possible attributes that norms could be constructed for. For example, norms would not be constructed for attributes relating to clothing or height of males and females even though these attributes would often be useful in gender discrimination in everyday life.

Recognition of a particular person’s face also can be achieved through the use of norms. An image of a face could act as a probe and recruit representations of similar faces or of faces seen in a similar context. Norms based on the evoked set of faces would then be compared to the probe face. The comparison is not as simple as in most other models because rather than comparing discrete points in a similarity space, the comparison involves regions of graded activation. However, the norms play the same role as the echo content in Hintzman's Minerva II model (1986). The mode of a norm can be thought of as its prototypical value, but in this case it is computed online.

The Current Work

The experiments presented in the following three chapters investigate the modulating role of different variables on the prototype effect in face recognition. By identifying variables of theoretical relevance and testing their effects, I aimed to further our understanding of the prototype effect. As of yet, the existing literature contains no research that takes as its focus the differing accounts of the prototype effect in face recognition. Experiments were designed to distinguish competing accounts as much as possible. However, as it would be nearly impossible to provide unequivocal evidence in support of a single viewpoint, other variables were explored to further constrain possible models of face recognition. My more general goal, throughout the present work, is to investigate how we learn new people's faces, and in particular, how our memory representations may change as we become more familiar with a face.

The experiments can be grouped into three families according to the type of face stimuli used. One chapter is devoted to each of the following families of experiments that used: (1) faces created from combining greyscale photographs of different facial features, (2) PCA generated faces and morphed faces, and (3) natural photographs of real people. The variables selected for their potential to modulate the prototype effect were examined in experiments spanning across chapters, and thus chapters share related themes and methodologies.

All of the experiments had a study phase followed by a test phase. In the study phase, either two or four exemplars were shown of each identity. In the recognition test, one of the seen before exemplars, the unseen prototype, and an unseen exemplar were shown of each studied identity. Faces of unstudied identities were used as distractors in some of the experiments. Higher recognition of the unseen prototypes compared to the seen exemplars was the criterion chosen to indicate a prototype effect.

Variables Affecting the Prototype Effect in Face Recognition

Familiarity

Learning how the prototype effect in face recognition will fare under conditions of increased familiarity will provide important information about how we represent different people's faces in long term memory.

Different instances (i.e. different images) are more likely to be attributed to the same individual if that individual is familiar to the viewer. If there is a mechanism that averages different instances of the same individual (but not between different identities), then this mechanism will be more likely to lump and average together different instances of a familiar face because those instances are more likely to be treated as belonging to a single identity.

If prototype effects are larger for more familiarized faces than less familiarized faces, this finding would be in accord with accounts that include explicit formation of identity based prototypes. Instance-based models would have more difficulty explaining this pattern of results but cannot be discounted.

If the prototype effect is smaller for more familiar faces, then explanations of previously observed prototype effects for unfamiliar faces need to be re-examined. If indeed there is a blending mechanism operating on different instances of the same individual's face, then it would be difficult to explain why the contribution of this mechanism would decrease if experience with a face is increased.

Context and Identifying Information

The prototype effect has the potential to reveal important clues about the type of representational system underlying our ability to recognize faces. Although the basic effect has been documented in the face recognition literature (Bruce et al., 1991; Cabeza et al., 1999; Cabeza & Kato, 2000), there have not been any serious attempts to isolate its cause. I attempted to do so by devising a series of experiments that predicted different outcomes depending on the theoretical model being considered. Ascribing identity and applying context to the faces will affect how they are processed and therefore influence the strength of the prototype effect. The two major theoretical accounts, prototype and episodic, differ on how context and identity information will affect processing.

I will compare episodic or instance-based models of memory with prototype-based models. The main difference between these two viewpoints is that in episodic models, every experienced event is represented in memory by its own trace, whereas in prototype-based models, an average representation is abstracted during encoding and stored explicitly in memory.

To understand how attaching names to faces might affect the internal representations we form, imagine meeting a pair of identical twins. If you meet them separately, you might never know they are actually different people, and thus never learn to tell them apart. However, given information about their identity, you can learn to distinguish them. To study this experimentally, I created variations on pairs of highly similar faces. In one condition, all these faces were given the same name; in the other, the two sets of variants were given separate names. In the test phase, all faces were presented without names. If the formation of prototype representations normally operates on instances within an identity but not between different identities, then prototype effects are predicted to be lower when a face was learned with two different names as compared to when learned with a single name. An episodic model might abstract a prototype at the time of retrieval, and thus still show a strong prototype effect regardless of whether it was learned with one name or two names.

Essentially the same method was used to apply different associated contexts to faces. In this case, cartoon bodies provided the context. This allowed the manipulation of whether a test face was presented with or without context at the time of retrieval. Predictions vary depending on the details of the model, but if the abstracted prototype representation is stored context free, then the strength of the prototype effect should not be affected by a contextual manipulation. In contrast, an episodic model in which context is encoded in each episodic trace would predict a different pattern of results depending on the manipulated context.

Chapter 2 (the first experimental chapter), presents three experiments using images of face composites as stimuli. Composites were constructed using the PROfit software package which the police use to help witnesses “build” the face of the perpetrator as they remember it (see General Methods for details). This system can produce fairly realistic greyscale images of faces. Faces of differing identities were prepared such that all faces of the same identity shared all of the same facial features (same eyes, same nose etc.), but no two faces of differing identity shared any feature in common. The configuration of features was varied systematically so that there was always a face image with the average identity specific configuration. Following the approach taken by Bruce et al. (1991), the internal configuration of facial features was left intact among faces of the same identity, but the relative positioning of the internal features was manipulated by shifting all of the internal features (as a single unit) up or down on the external face shape.

Chapter 3 used two kinds of face stimuli in separate experiments. Experiments 4 and 5 used faces made by an in-house software program that generates images of faces using PCA (see General Methods for details). The faces are greyscale and have had the hair and ears removed. The image quality is near photographic and substantially higher than PROfit faces. The advantage of using PCA generated faces is that their similarity (in PCA space) can be prespecified and made to order. Later, I will describe how I put this property to use.

Experiment 6 used full colour morphed faces made using Psycho-Morph (Tiddeman, Burt, & Perrett, 2001), a software program similar to the popular Gryphon Morph software. The image quality was extremely high and on the same level as natural photographs (this was largely due to the high resolution of the original photographs). I did not have access to multiple images of the same person's face to morph between, so I used pairs of photographs of different peoples' faces and morphed using tiny increments to produce exemplars that appeared similar enough to belong to the same identity. Using PCA and morphing techniques to probe for prototype effects is important to extend findings using simple configural variants to conditions using more complex subtle changes to the facial features and their configuration alike.

Finally, in Chapter 4, the studied identities were real people. Two full colour, natural images of each identity were shown in the study phase. One of the images was a high quality photograph, and the other was a video still of lesser quality (but still easily identifiable by a familiar viewer). The two images were taken at the same time, and thus by far the most obvious differences in their appearance are the superficial image properties such as brightness, contrast, and graininess. A 50%/50% morph was created for each identity from its two original images. It was important that at least some of the experiments used natural images to ensure any prototype effects found in the preceding chapters were not particular to the artificial image manipulations that were used. Before I can make any claims about how our face recognition system deals with everyday encounters with the same person's face, I need to test for the prototype effect using naturally occurring sources of facial variation, within and between.

The Big Questions

The questions central to this thesis are: how robust is the prototype effect in face recognition? Are there any variables that affect the strength of the prototype effect? In particular, what role does familiarity play? Will the use of context and identity information be able to distinguish between prototype- and exemplar-based accounts of face recognition memory? Can the prototype effect be replicated with other types of stimuli? Will comparable prototype effects be found using real photographs of real people? The studies presented in the next three chapters will address these questions. Their success will be evaluated and discussed at the end of each chapter and in the General Discussion in Chapter 5. The implications of the present work to the larger body of research in face recognition and general memory mechanisms will also be discussed in Chapter 5.

General Method

All of the experiments follow a similar methodology, and this section will cover some general information about the equipment, the preparation of stimuli, and the procedure.

Materials and Apparatus

Equipment

Images were displayed and responses were collected using a Hi-Grade Notino 3600 windows laptop with a 14” monitor at 1024 x 768 resolution. The program E-Studio from the E-Prime software package (version 1.1) was used to present the stimuli and to record any responses made using the laptop keyboard.

Stimuli

Image Manipulation Software

1.) PROfit Windows Version 3.1:

PROfit was used to prepare face stimuli for experiments in Chapter 2. It is a facial composite system designed for use by the police in the area of eyewitness testimony.

Witnesses can work with an operator to build a face that resembles the perpetrator as remembered by the witness.

PROfit uses a large database of greyscale photographs of individual facial features (e.g. eyes, noses, mouths, hairstyles etc.). There are about 200 options per feature depending on the feature. Originally, the options on each feature belonged to different people's faces, but once isolated, the features of different people's faces can be combined to form novel faces. Presently, the database includes features photographed from male faces only.

The system is easy to use. A default face appears on the screen. By clicking on a feature, one can flip through the options for that feature while watching the default face change to display the selection. The configural arrangement of features can also be easily adjusted. Selecting a feature and clicking an arrow button will move the feature a small fixed distance up, down, left, or right relative to the rest of the face.

The resulting composites themselves are fairly realistic faces. Although they are comprised of pieces of photographs, they do not appear as realistic as a photograph of a face normally does. This is partially because the pattern of shading from the original lighting may appear discontinuous especially along the joint between different face parts. The system does allow you to make alterations to the brightness of each feature separately, and this does improve the realism of the face. Similarly, there is an "airbrush" tool that can reduce the noticeability of the feature joints. I used these two types of adjustments frequently whilst preparing the PROfit faces for the experiments in Chapter 2.

2.) PCA Face Generator:

To explain how the generator works, an overview of PCA may be helpful to some readers:

PCA takes a large multivariate set of data and summarizes the patterns in the data by constructing a smaller number of new variables called principal components. The principal components will correlate with the original variables, but they will each reflect independent sources of covariation.

In terms of face images, an image can be expressed as a vector whose components are the intensity/brightness value on each pixel. The number of dimensions in the original image space is the number of pixels per image. PCA can reduce the number of dimensions by forming a new co-ordinate axis through the image space in the direction that maximizes the variability in the data set. Images can be projected onto this new axis and their score recorded. The first axis will always account for the most variance in the set, but subsequent axes will be laid down that account for progressively less and less variance. The result is a set of eigenvectors which are also called *eigenfaces* when the analysis is preformed on a large number of face images. The number of possible eigenfaces is equal to the number of faces, so a relatively small number of them are selected to be the principal components of the space. In theory, the principal components define the space of all possible human faces. Therefore, any point in the space can be expressed as a face image just as any face image can be expressed as a point in PCA space. Thus, the PCA face generator can generate novel face images, and this approach was used to generate stimuli for Experiments 4 and 5.

The principal components used by the in house generator computer program were derived from a set of 300 images of young Caucasian male and female faces. Each face was “marked up” by manually locating 219 points around key features such as eyes, nose, and the face outline, using Psycho-Morph. An average face shape was computed, and each image was morphed to the average shape using Psycho-Morph. All further processing and image generation was carried out in Matlab, with code written by Peter Hancock at Stirling University. The faces were cropped using an oval set to just include the jawline, but exclude the ears and curve across above the eyebrows to exclude hair. The *shape free* face images and the shape parameters for each face were then separately subjected to a Principal Components Analysis. This gives the principal modes of variation in the image domain (eigenfaces) and in the shape domain (eigenshapes). For the purposes of this study, the first 30 eigenfaces and 18 of the first 20 eigenshapes were used. The two removed were numbers 1 and 3, which code for the head nodding and shaking respectively. While attempts were made to get the participants to look directly at the camera, people do vary, resulting in these components carrying much of the variance in shape.

Linear recombination of the eigenfaces and the eigenshapes, in random amounts, specifies a novel face. The specified eigenfaces are added together and then morphed to the shape specified by the eigenshapes to produce the final output image. The generator can also specify images that are contingent on other images. For example, the user could have the generator produce images three at a time. The location of the first face is random. The second and third faces are moved in the opposite direction by a specified amount of distance away from the first face. Thus, the user can control the similarity among the faces by controlling the distances in PCA space. I made use of this feature by creating exemplar images that were equidistant from a prototype image. The twin faces were created this way in Experiment 5 by setting the distances among exemplars of the same face to be very small.

The high level of control over the similarity relations and the speed of generating stimuli are the two major advantages over other methods of image manipulation. The quality of the images is quite good in general, but artefacts and peculiarities are frequent. Trial and error can be used to help avoid image oddities. Even though the PCA space should represent all possible faces, in practice, the total variability in the images produced was somewhat limiting. It would be difficult to produce a set of say 50 different identities that all clearly looked like different people. For experiments only requiring 20 identities or less, the generator can do an adequate job. One final criticism is that similarity is being manipulated within a physical image space, so there is always uncertainty as to how psychologically similar the images are.

3.) Psycho-Morph

To prepare images for morphing, control points must be plotted on the face. Although Psycho-Morph has an auto-delineation option, ultimately, the precise locations of the points must be manually adjusted to produce a high quality morph. Many of the control points are located in important positions on the face like the pupils, the tip of the nose, or along the curve of the eyebrow. An equal number of corresponding control points are plotted onto each image. In addition to morphing, the control points can also be used to fit a mask around the contours of face. In Experiment 6, a black mask was fitted around each face which excluded the hair, ears, and neck.

Psycho-Morph locates the control points of the morph face by moving a specified distance along the vector connecting the control points in the parent images according to the specified degree of morphing. Linear interpolation is used to determine the locations of intervening pixels across the image based on the location of the nearest control point. A fade process then weights the brightness values for each corresponding pixel according to the specified proportion of each parent face in the morph (Wolberg, 1990).

With Psycho-Morph, essentially the same process can be used to take the difference between two parent images and apply that difference to a third face image. Caricatures are generated in this way by taking the difference between a premade average face (average of the population) and a particular person's face and then applying this difference back onto the image of the person's face. The direction and magnitude of the difference can be specified by the user to say, for example, create differing degrees of caricature and anticaricature. This process was used to mimic within face variation in Experiment 6 by applying the differences between two faces to a third face image by the same amount in the positive and negative direction. This leaves the third image as the prototype, or average, of the two newly constructed faces.

4.) Adobe Photoshop 6.0

When small adjustments to the images were required, Adobe Photoshop was used in all experiments unless otherwise stated. Small adjustments included centring faces within the image and resizing images. Changes particular to the different types of stimuli are as follows:

PROfit faces: Hair and ears were removed from the PROfit faces by selecting the to-be-removed area and filling it with the background colour. An inverse rectangular shaped frame usually provided a sufficient fit along the hairline. In cases where the rectangle did not fit well with the shape of the face or hairline, the selection was adjusted accordingly. Importantly, all variants based on the same identity shared the same external features, so once a selection had been made to remove the hair on one variant of a face, that same selection could be reapplied to all the other variants of that identity to standardize the external face shape.

Morphed faces: Some minor alterations were made to images prior to morphing. In particular, the hairline sometimes had to be adjusted slightly if one of the two paired images had hair protruding into the area of the face not excluded by the external feature mask. Such hair was hidden using the clone stamp tool to select a nearby area of skin on the face and “stamp” it down over the hair. Some images needed to be resized and/or centred. Touch-ups were sometimes needed to remove obvious marks like large moles or blemishes. This helped to produce realistic, high quality morphs by reducing artefacts like blurred lines from a stretched blemish.

Natural faces: The faces in the video stills were on a blue background, so this was removed and replaced with a white background. Anything below the top of the neck was removed as well, so the video stills would resemble the photographs. The original blue background had leaked onto the face in some cases, so the blue was stamped over or painted the same colours as the skin tone or hair depending on where the leak was.

Images were resized, centred, and set to the same resolution. The hair of the morphed images needed to be filled in or shaved down a bit to look normal.

General Procedure

Experiments were divided into separate study and test phases. Different identities were presented in the study phase, and more than one exemplar of each identity was studied. A recognition test followed with different versions of each studied identity. The procedure of a typical experiment was as follows:

Study phase: Participants read instructions for the study phase off the monitor, and the experimenter elaborated on them. Participants then raised any questions they might have had.

Images were presented one at a time in the centre of the screen. Presentation was selfpaced, and participants pressed the space bar to view the next face in the series. While they viewed the faces, participants were asked to either remember them or to make a subjective rating about their appearance. The number keys on the laptop keyboard were used to enter ratings. Depending on the experiment, either 2 or 4 exemplars of each identity were presented. The study phase was blocked so that exactly one exemplar of each face was included in each block. This reduced the likelihood that two exemplars of the same face would be shown on consecutive trials. The prototypes never appeared in the study phase. Filler faces were interwoven into the series of study faces. The filler faces were prepared in the same way as the prototypes. In this way, prototypes would not be rejected at test merely because they looked different in some way from anything participants saw in the study phase.

Test Phase: Participants read instructions for the test phase off the monitor, and the experimenter elaborated on them. Again, participants raised any questions they had. The task was either to judge if the image on the screen was one of the images in the study phase, or if an image of the same person had been presented earlier (regardless of whether images were the same or slightly different). Participants rated how confident they were about their decision from 1 to 10 (1=*sure not seen*; 10=*sure seen*).

General format of the test phase was as follows. Again, images were presented one at a time in the centre of the screen, and the number keys on the laptop were used to enter responses. Once a number key was pressed, the next trial proceeded. There were no response time limits. The first six trials of the test phase were practice trials. The data from these trials was not analyzed. For each identity shown in the study phase, there were three different versions of it in the recognition test. One was a seen exemplar, one was the unseen prototype, and one was an unseen exemplar. If the task at test was to recognize the same people shown earlier, then half the test trials presented distractor faces of novel identities.

Chapter 2: The Role of Familiarity in Recognizing Prototypes of Individual Faces

Chapter Introduction

The purpose of the experiments in this chapter was to gain insight into how the identity of faces is represented in memory and how these representations develop with experience. I chose familiarity as a starting point for inquiry into the prototype effect in face recognition and its origins. This seemed an obvious choice because the issue driving all research on prototype effects is how we are able to learn abstract information. We learn based on our experiences, and generally speaking, the more we experience, the more we learn. Despite being fairly poor at recognizing unfamiliar faces, people are surprisingly accurate at recognizing familiar faces even across large transformations in viewpoint, lighting, and expression. How do our representations of faces change as they become more familiar?

Although two past studies, Bruce et al. (1991) and Cabeza et al. (1999), have focussed specifically on learning prototypes of different facial identities, no study to date has examined this at different points in the familiarization process (but see Burton et al., 2005 for a different approach using famous faces). Both Bruce et al. and Cabeza et al. presented participants with a series of exemplars of unfamiliar faces in a study phase followed by a recognition test phase. In the study phase of every experiment, participants either saw four different exemplars of each unfamiliar face or two exemplars twice each. Thus, the degree of learning participants had with the faces was quite limited. If familiar and unfamiliar faces really are processed in different ways, then one might wonder if prototype abstraction could be affected by the amount of experience participants had with faces in the study phase.

My goal was to extend previous research by manipulating how familiar participants were with the faces in the study phase. The simple construction of the face stimuli used by Bruce et al. (1991) appealed to me, so I chose to model my experiments in this chapter after the ones in their study. For example, in their Experiments 7 and 8, Bruce et al. created different exemplars of the same person's face by shifting the internal features (eyes, nose, and mouth) up and down on the face outline. In their bimodal condition, two exemplars of each face were made by shifting the internal features by 9 pixels up (+9) or down (-9) from the starting face (0). Exemplars were presented twice each within a series of study faces. Half of the participants made incidental ratings during the study phase. The other half of participants were warned of the recognition test, and their task in the study phase was to examine the faces and try to remember them. In the test phase, faces were presented one at a time, and participants were asked to rate how sure they were that they had seen that particular exemplar before in the study phase. Participants were more confident of seeing the unseen prototype (0) than the actually seen exemplars (+/-9). Although this was true for participants in the incidental learning group, participants in the intentional group, who were asked to attend to the details of the study faces, did not show a clear prototype effect.

Although Bruce et al.'s (1991) results were intriguing, it cannot be concluded that prototype representations were being formed simply because the prototype effect was observed. More investigation is needed to form a convincing account of what underlies the so called “prototype effect”. It is therefore necessary to move beyond the limits of past experiments that have only used unfamiliar faces under conditions of minimal learning.

The approach taken in my experiments was simple. The procedures and stimuli were kept very similar to those used in Bruce et al.'s (1991) experiments, but the critical difference is that my designs included an initial familiarity training phase prior to the experiment proper. Thus, some faces will have been familiarized in the training phase, whereas other faces will be shown for the first time in the study phase. By using this within subjects approach, recognition performance in the test phase for familiarized vs. unfamiliar faces was compared. The key question of Chapter 1 is does familiarity with the exemplars affect the magnitude of the prototype effect in face recognition?

It was predicted that the prototype effect will be stronger for newly familiarized faces compared to faces that were not part of the initial familiarization training phase. To be clear, increased confidence ratings for prototypes over the seen exemplars should be observed more for the familiarized faces compared to unfamiliar faces.

The above prediction is based on the following rationale. Different exemplars are more likely to be attributed to the same individual if that individual is familiar to the viewer. If there is a mechanism that averages different exemplars of the same individual (but not different individuals), then this mechanism will be more likely to lump and average together different exemplars of a familiar face because those exemplars are more likely to be treated as belonging to a single identity.

Thus, one possible prediction made by a prototype account of face recognition is that more familiar faces will produce stronger prototype effects than less familiar faces. Predictions made by an exemplar account are less clear and likely would depend on the categorical structure of the training stimuli. Despite this, knowing how familiarity affects the prototype effect would be useful for developing formal models of face recognition from either a prototype- or exemplar-based point of view.

Experiment 1

The main purpose of Experiment 1 was to determine whether the magnitude of the prototype effect is affected by the amount of familiarity training participants receive with the studied faces. I also wanted to replicate the prototype effect under conditions of minimal learning as in Bruce et al. (1991) and Cabeza et al. (1999) to ensure the stimuli and procedures I used were capable of producing a prototype effect. The important manipulation was the addition of a familiarity training phase prior to the normal study phase of the experiment. Participants received familiarity training on half of the faces in the study phase.

The stimuli used were PROfit faces and were similar to the faces Bruce et al. (1991) constructed using the Mac-a-Mug system. In order to examine prototype recognition of face identity, a set of variants representing the same individual’s face must be created. This is one of the most difficult aspects of designing experiments in this line of research. The exemplars of a face need to be different enough to be distinguishable but similar enough to be perceived as depicting the same person’s face. Bruce et al. showed that a simple up or down shift of the internal features was capable of producing prototype effects. I borrowed this technique in the present experiment to create identity variants around each starting face.

In the training and study phases, participants learned the faces incidentally in that they were not warned that there would be a subsequent recognition test. Instead, participants rated several attributes regarding the appearance of each face (e.g. attractiveness, honesty etc.). This approach was adopted because Bruce et al. (1991) found it yielded stronger prototype effects compared to when participants were asked to memorize the faces.

Although familiarity was the main focus, the role of similarity among exemplars of the same face was also tested out of additional interest. Past research suggests there is good reason to expect that the similarity of studied exemplars will have an effect on prototype abstraction. For example, Cabeza et al. (1999, Experiment 3) found that when the similarity of exemplars was high, there was a prototype effect, but when the similarity of exemplars was reduced to the point that they no longer resembled the same individual, the prototype effect was absent. Moreover, predictions about the effect of similarity on prototype recognition can be made using the same rationale used to predict the effect of familiarity. That is, if a prototype mechanism operates by averaging exemplars of the same person's face, then more similar exemplars of a face will be more likely to be averaged together into a prototype representation than less similar exemplars would be. Thus, the more similar the exemplars of a face are in the study phase, the higher the prototype recognition will be in the test phase. However, it is harder to make a prediction about how similarity will affect the *strength* of the prototype effect because the studied exemplars would also be predicted to be recognized better when similarity is high. This is because the abstracted prototype will resemble the studied exemplars more whenever variability among exemplars of the same face is low. If one assumes a model that stores only the prototype of an individual's face, then the prototype effect should be strongest when the variability among exemplars of the same face is as high as possible without being so high as to not trigger the averaging mechanism.

Method

Participants

Participants were 24 University of Stirling psychology undergraduates. They received credit toward their degree requirements in the Psychology Honours programme for their participation. Each participant was tested individually and completed the experiment in a single session which lasted approximately 20 min. Participation was voluntary, and at the end of the session, participants were fully debriefed as to the purpose and theoretical basis of the experiment.

Materials and Apparatus

Stimuli.

Face stimuli were made using the software package PROfit by combining photographs of different features to create semi-realistic greyscale images of faces. See the General Methods section at the end of Chapter 1 for more information about PROfit.

Twenty faces of different identities were constructed. No two faces had the same face shape, eyes, eyebrows, nose, mouth, or ears. All faces were males that appeared to be between 16-40 years of age. The faces were centred on a teal background. Image resolution was 10 pixels/cm, and the face portion of the images was approximately 11 cm wide by 17 cm high.

A series of five variants were constructed for each of the twenty identities. Each series consisted of two exemplars that had their internal features shifted down (-2,-4), two exemplars that had their internal features shifted up (+2,+4), and a prototype (+0) face that served as a starting point for the shifting (see Figure 2.1). Features were shifted by fixed amounts corresponding to clicks on the computer mouse. A +2 or -2 exemplar was shifted two clicks up or down respectively on the mouse; a +4 or -4 exemplar was four clicks up or down respectively.

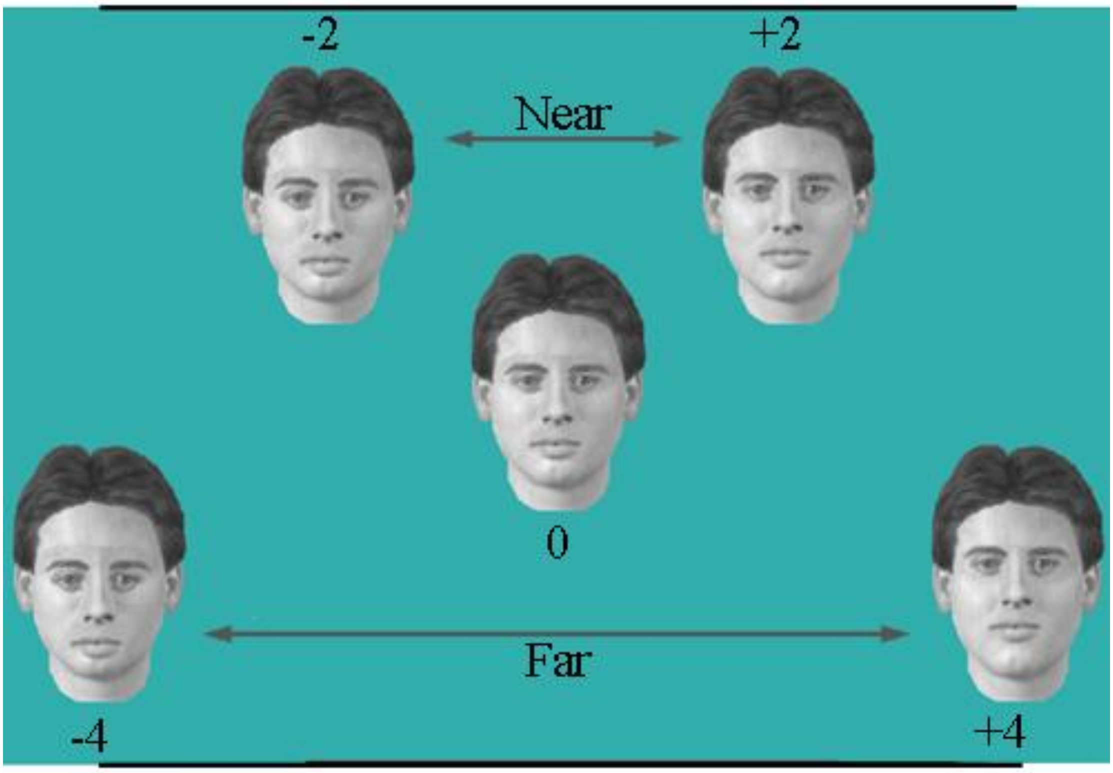


Figure 2.1. Example stimuli from Experiments 1, 2, and 3. Shows four exemplars of the same face made by shifting the internal features up and down on the face outline by a fixed distance. The central face was the starting face and served as the prototype of the set of exemplars. The uppermost and lowermost faces show a pair of seen exemplars in the near and far condition respectively.

Equipment.

All aspects of the experiment were conducted using a Hi-Grade Notino3600 windows lap top with a 14” monitor at 1024 x 768 resolution. The program E-Studio from the

E-Prime software package (version 1.1) was used to control the presentation of stimuli and to record responses made using the laptop keyboard.

Design

A 2 (Familiarity) x 2 (Distance) x 3 (Exemplar Type) within-subjects design was used to compare prototype effects across two levels of familiarity. The two levels of familiarity were familiarized and unfamiliar. Familiarized faces were shown in both the training and study phases whereas unfamiliar faces were shown only in the study phase. Distance refers to the distance between the position of the internal features of the two seen exemplars for a given identity shown in training and study. Manipulating the internal feature distance between the two exemplars affects their similarity. This factor was included out of additional interest. The two levels of distance were near and far. Near faces were shown as +2 and -2, but far faces were shown as +4 and -4 in the training and study phases. The three levels of exemplar type were prototype, seen, and unseen. Prototype faces had features placed in the position that corresponded to the average of two exemplars shown of a given identity in training and study. Seen faces were one of the two exemplars of an identity shown in training or study. Unseen faces were one of the two non-average exemplars of an identity that were never shown in the training or study. For example, if +2 and -2 were shown in the study phase, either -4 or +4 was shown as the unseen face in the test phase.

Half of the 20 face identities were randomly assigned to be used as familiarized faces; the other half, as unfamiliar faces. Familiarized faces were presented in an initial training phase as well as a separate study phase, whereas unfamiliar faces were not presented until the study phase of the experiment. Half of both the familiarized and unfamiliar face identities were assigned to the *near* condition; the other half, to the *far* condition. Exemplars of near faces were more similar, or “near”, to one another and were shown as +2 and -2 versions in the training and study. Far faces were shown as +4 and -4 versions.

In addition to the 20 critical identities, six filler identities were created in a similar manner without adjusting the configuration from the starting position. Thus, no additional variants were made on these identities; just the zero version was used. The fillers were used so that participants would not notice that there was anything “strange” about the shifted faces that were seen in the study phase and reject the prototype at test merely on the basis that it looked more normal than anything seen in the study phase.

Recognition memory performance was measured by recognition confidence ratings collected during the test phase. A 10 point Likert-type scale (1= *definitely never seen before,* 10= *definitely seen before*) was used to assess how sure participants were that each exemplar had been shown previously in the study or training phases. Importantly, judgments were made about the particular image rather than the person’s face in general. Prototype effects were calculated by subtracting ratings of seen faces from those of prototype faces.

Programming and Counterbalancing.

The particular identities allocated to the familiarized, unfamiliar, near and far conditions were rotated, so that across participants all faces served in all conditions of the experiment. The series of faces prepared for the training and study phases were divided into two halves. The first half contained one exemplar of each identity; the second half contained the other exemplar. This was done to reduce the chance that two exemplars of the same identity would be shown consecutively within the series. Each half of the training and study series contained an equal number of plus (+2 or +4) and minus (-2 or -4) faces. Order of presentation of the two halves was randomized as was the order of exemplars within each half. Three of the six filler items were randomly assigned to each half of the study phase as well.

The series of faces for the test phase was prepared by selecting three exemplars of each of the 20 identities. The prototype (0) face was always one of the three of each identity. The remaining two were either -2 and +4, OR +2 and -4. Half of the identities at test used -2, +4, and half used +2, -4, and the particular identities used were counterbalanced across participants. Order of presentation during the test phase was randomized for each participant.

Procedure

The experiment was divided into three separate phases: the training phase, the study phase, and the test phase.

1. Training.

In the training phase, all 10 of the identities selected to be familiarized were presented one at a time for 4 s. Two exemplars of each identity (+2/-2, or +4/-4) were presented four times each over four blocks. For each block, participants were asked to rate each instance on a different attribute: intelligence, confidence, honesty, and attractiveness respectively, as part of an incidental learning task. All images were presented for 4 s each. After 4 s, the face was replaced with a white screen containing a prompt message. The message reminded participants of what they were rating and instructed them to enter a response. Participants could not respond until the prompt screen was displayed. The next trial began immediately after a response key was pressed.

2. Study.

In the study phase, all 10 of the identities selected to be unfamiliar were presented in addition to the same 10 familiarized face identities presented in the training phase. Again, two exemplars (+2/-2, +4-4) of each identity were presented. Also included in the series were six filler faces. Each exemplar was presented only once, and participants tried to guess its perceived age. All faces were presented one at a time for 4 s each, and responses were entered when a prompt screen replaced the face as in the training phase.

3. Test phase.

In the test phase, 10 familiarized and 10 unfamiliar face identities were presented one at a time. For each familiarized and unfamiliar face, one seen and one unseen exemplar were presented (e.g. -2/+4 OR +2/-4) as well as the unseen prototype (0) for each identity. Thus in total, 60 images were presented as test items (20 identities x 3 exemplar types). Participants were asked to judge how sure they were that each image was *exactly* the same as one they had seen in the study phase by rating each test face on a 10 point Likert-type scale (1=*definitely never seen before*, 10=*definitely seen before*) by pushing the appropriate number key on the keyboard. Faces remained on the screen until a key was pressed to enter a response. There was an intertrial interval (ITI) of 500 ms during which a blank white screen was displayed. Participants were informed that no entirely new people's faces would be presented in the test phase. Rather, all faces would belong to people shown in the study phase, but some of the images would be identical and some slightly different to the particular images presented in the study phase.

Results

Main Analyses

Mean ratings given to the test faces in each of the experimental conditions are shown in Figure 2.2. A repeated measures ANOVA was conducted to analyze whether the factors familiarity, distance, and exemplar type had any influence on recognition ratings. The type I error rate was set at .05 for all analyses in all experiments.

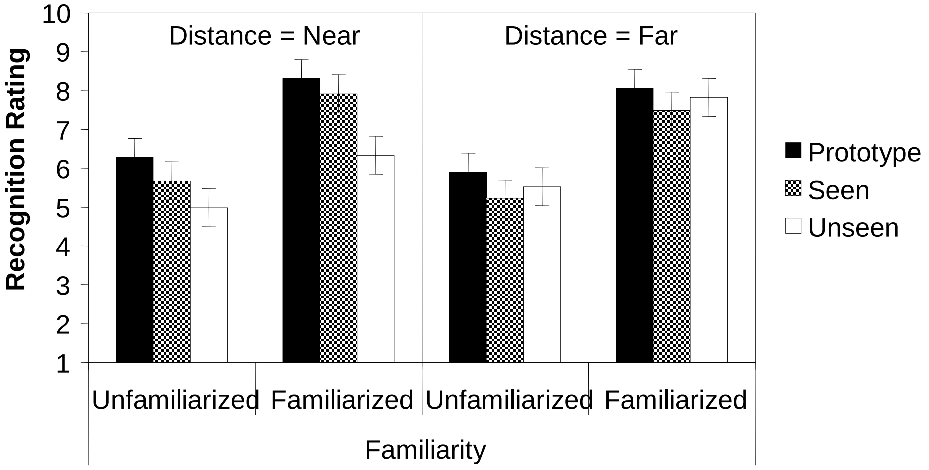


Figure 2.2. Mean recognition confidence ratings (1=*definitely never seen before*, 10=*definitely seen before*) of prototype, seen, and unseen exemplar types across levels of familiarity and distance between studied exemplars in Experiment 1. Error bars represent 95% within-subject confidence intervals (Loftus & Masson, 1994).

There was a main effect of familiarity on ratings, *F*(1,23)= 62.01, *MSE*= 4.91, *p*<.001. As one would expect, familiarized faces (*M*=7.65, *SD*=1.23) were rated significantly higher than unfamiliar faces (*M*=5.60, *SD*=1.48). There was no main effect of distance, *F* < 1. Near faces were rated equally as high as far faces. The main effect of exemplar type was significant, *F*(2,46)= 15.02, *MSE*= 1.52, *p*<.001. Mean confidence ratings (with standard deviations in parentheses) for prototype, seen, and unseen exemplars were 7.12 (1.21), 6.57 (1.37), and 6.17 (1.34) respectively. Post hoc paired-samples *t* tests using a Bonferroni adjustment for multiple comparisons confirmed that prototypes were rated significantly higher than seen exemplars which were rated significantly higher than unseen exemplars (see Table A1 in Appendix A for exact statistics).

There were two significant interactions. The first was a Distance x Exemplar Type interaction, *F*(2,46)= 15.62, *MSE*= 1.01, *p*<.001. As can be seen in Figure 2.3, although prototypes were rated highest at both levels of distance, seen exemplars were rated higher than unseen exemplars for near faces, but the opposite was observed for far faces. This may seem surprising, but as will be discussed later, it is likely because unseen fars were closer to the prototype than seen fars. Post hoc paired-samples *t* tests (on all possible pair-wise comparisons) detected two significant differences among exemplar types in the near condition but none in the far condition after Bonferroni adjustment. In the near condition, the unseen exemplar was rated significantly lower than the prototype, *t*(23)=4.85, *SD*=1.65, *p*=.001, and the seen exemplar, *t*(23)=4.25, *SD*=1.31, *p*=.004. Post hoc comparisons also confirmed that unseen exemplars were rated significantly higher in the far condition compared to the near condition, *t*(23)=3.70, *SD*=1.35, *p*=.018.

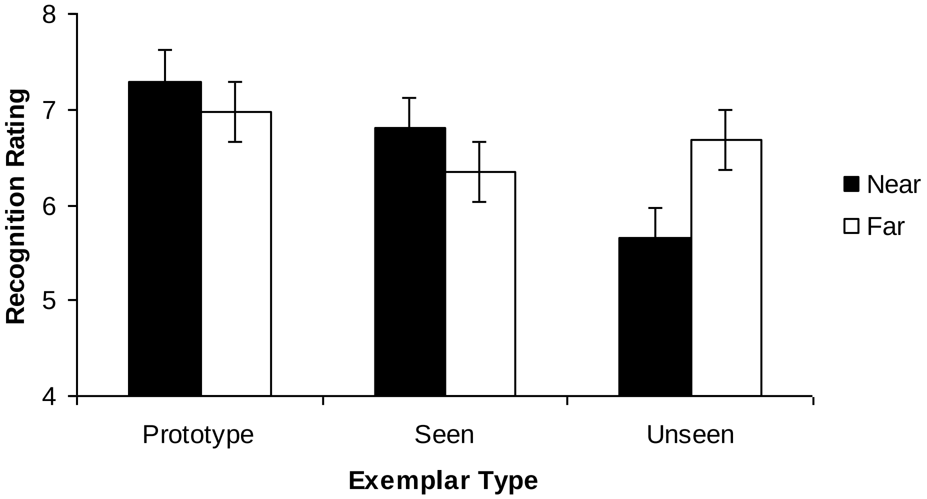


Figure 2.3. Mean recognition confidence ratings of prototype, seen, and unseen exemplar types across levels distance (near and far) between studied exemplars in Experiment 1. Error bars represent 95% within-subject confidence intervals.

The second interaction was the three-way among familiarity, distance, and exemplar type, *F*(2,46)= 3.32, *MSE*= 0.46, *p*=.045. This interaction was due to a smaller rise in recognition ratings for familiarized over unfamiliar faces specifically when a face was unseen and near. In other words, the familiarity training raised ratings by approximately equal amounts for all exemplar types across both levels of distances with the only exception being unseen near faces which received less of an increase. Post hoc Bonferroni paired-samples *t* tests (on all possible pair-wise comparisons) confirmed that ratings were significantly higher after familiarization training in all conditions except when the face was unseen and near. All interactions not mentioned were non-significant (*p*> .05).

Analysis of Prototype Effects

An additional repeated measures ANOVA was conducted to compare just the prototypes with seen exemplars (ignoring the unseen condition) across all levels of familiarity and distance. This analysis directly tests the experimental hypothesis that familiarity will increase the strength of the prototype effect (defined here as prototype – seen). Furthermore, if any of the interactions revealed by the main ANOVA were modulating the prototype effect, dropping the unseen condition from the analysis will provide a more sensitive and direct means of assessing this possibility.

There was a main effect of familiarity in the expected direction, *F*(1,23)= 63.21, *MSE*= 3.59, *p*<.001. There was a main effect of distance, *F*(1,23)=7.75, *MSE*=0.92, *p*=.01. Near faces (*M*=7.05, *SD*=1.21) were rated higher than far faces (*M*=6.66, *SD*=1.31). The main effect of exemplar type was also significant, *F*(1,23)= 15.02, *MSE*= 1.52, *p*<.01. Prototypes were rated higher than seen exemplars. Most importantly, there were no significant interactions (all *p*'s> .4). The strength of the prototype effect was unaffected by familiarization or distance (i.e. similarity) of exemplars in the learning phase as manipulated in this experiment.

Discussion

Results of Experiment 1 revealed an advantage for recognizing the average picture of a given person's face over an actually seen picture of that same person. This prototype effect was observed both when participants had only minimal exposure to a person's face and when they received addition familiarity training beforehand. The prototype effect was also demonstrated at both levels of exemplar similarity (near and far). However, the magnitude of the prototype effect was not influenced by either familiarity or the distance between studied exemplars. Recognition ratings of prototypes received just as much of a boost from the familiarity training as the actually seen exemplars did.

Thus, the prediction that the strength of the prototype effect would be stronger for newly familiarized faces compared to the unfamiliar faces was not supported. This prediction was based on two key assumptions. The first was that different exemplars are more likely to be attributed to the same individual if that individual is familiar to the viewer. The second was that there is a mechanism that averages information across exemplars belonging to the same identity but not between different identities. Why then did the results fail to show an effect of familiarization training on magnitude of the prototype effect? Perhaps the two exemplars of each identity looked so similar to one another that they would be treated as the same individual even without any previous exposure or familiarization. Familiarity might have greatest impact on the prototype effect when exemplars are maximally dissimilar while still being similar enough to be treated as the same person's face. The lack of effect of distance on the strength of the prototype effect is inconsistent with this idea. However, the unseen exemplars of far faces were within the range of the seen exemplars whereas the unseen exemplars of near faces were outside the range of the seen exemplars. This makes it difficult to compare prototype effects between near and far faces because participants responded quite differently to the seen and unseen exemplars depending on whether the exemplars shown in the study phase were near or far.

It is important to consider the possibility that identities are not stored as prototypes produced by a mechanism that averages exemplar representations of the same person's face. Since the prediction that familiarity would increase the prototype effect was based on the assumption that prototypes were abstracted during the study phase, this may be why the prediction failed. Could the lack of effect of familiarity be explained by any other perspectives on the nature of memory representation? The simplest conception of human memory is that we store and retrieve information about each exemplar separately (i.e. a nearest neighbour model). This type of simple model would not predict additional familiarity to increase prototype effects. In fact, it would make the opposite prediction. Namely, that the more exposure one has to the individual exemplars, the less likely one is to believe having seen the unseen prototype compared to one of the actually seen exemplars. However, the simple model would predict actually seen exemplars to be better recognized than an unseen prototype at all levels of familiarity. Bruce et al. (1991) found some evidence that unintentional learning during the study phase yielded stronger prototype effects than intentional learning. Perhaps consciously trying to remember and distinguish each exemplar reduces the likelihood of exemplars being lumped together and being incorporated into a single average representation. It would be interesting to know whether participants are even capable of selectively recognizing the exemplars that were actually studied. To examine this, Experiment 2 was conducted which was the same as Experiment 1 except that participants were told to try to remember exactly the way each picture of each face looked while they completed the familiarity training and study phases.

Experiment 2

The goal of Experiment 2 was to test whether participants would still give higher recognition ratings to the unseen prototype compared to a seen exemplar even after an extended learning phase requiring intentional memorization of each exemplar. If participants really are capable of selectively recognizing the specific seen exemplars, then they should give higher recognition ratings to seen exemplars than to prototypes. On the other hand, prototype abstraction may occur whenever two exemplars of the same face are shown despite the efforts of participants to store and retrieve exemplars separately.

To relate back to the rationale proposed in Experiment 1, if the mind stores facial information pertaining to identity by averaging exemplars of the same face, then intentionally trying to remember each exemplar as a separate entity will lead to reduced prototype effects (compared to incidental learning) to the extent that intention is capable of interfering with the averaging process. With additional learning, participants should be better able to differentiate the exemplars and build separate representations for each of them.

Experiment 1 found that additional learning did not help participants to reject the unseen prototype as having been one of the pictures seen before. However, if participants were using the additional learning trials to try to distinguish between exemplars of the same face, then familiarization might affect the strength of the prototype effect. In Experiment 2, by using an intentional learning task and providing an additional training phase, we aimed to give participants a much better chance of selectively recognizing the actually seen exemplars than the incidental task in Experiment 1 provided.

Method

Participants

Participants were 24 University of Victoria undergraduates enrolled in an introductory psychology course. Participants received extra credit points in their class for their participation. Participants were tested in a single session lasting approximately 25 min.

Materials and Apparatus

Stimuli.

Stimuli were the same as in Experiment 1 except 4 more identity sets were constructed thus upping the total number of different face identities to 24.

Programming and Counterbalancing.

The same procedures were used as in Experiment 1 except that the order of the prototype, seen, and unseen exemplar of each identity at test was counterbalanced across participants, so it would be possible to check if recognition ratings to faces were influenced by earlier variants of those faces presented earlier in the test phase.

Equipment.

Same as in Experiment 1

Design

The design was the same as in Experiment 1.

Procedure

Each session was conducted as in Experiment 1 except participants were informed at the start of the experiment that there would be a subsequent recognition memory test for the faces they were about to see. The structure of the identity series and the nature of the recognition rating task were fully explained beforehand. Participants were shown an example of an identity series that was not shown in the actual experiment. Also, this time there were 12 and 24 different identities presented in training and study phases respectively. The test phase contained a total of 72 trials (24 different identities x 3 different versions).

Results

Main Analyses

Mean ratings given to the test faces in each of the experimental conditions are shown in Figure 2.4. Recognition ratings were analyzed using a repeated measures ANOVA which included the factors familiarity, distance, and exemplar type.

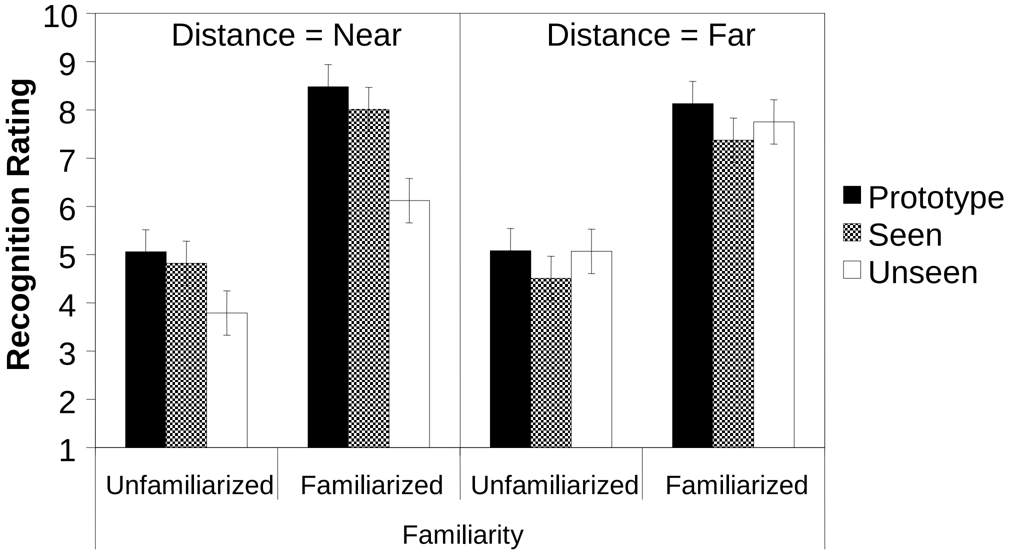


Figure 2.4. Mean recognition confidence ratings (1=*definitely never seen before*, 10=*definitely seen before*) of prototype, seen, and unseen exemplar types across levels of familiarity and distance between studied exemplars in Experiment 2. Error bars represent 95% within-subject confidence intervals.

There was a main effect of familiarity on ratings, *F*(1,23)= 175.98, *MSE*= 3.50, *p*<.001. As expected, participants gave higher recognition ratings to familiarized faces (*M*=7.64, *SD*=1.09) compared to unfamiliar faces (*M*=4.72, *SD*=0.82). The main effect of exemplar type was significant, *F*(2,46)= 21.71, *MSE*= 1.12, *p*<.001. Means (with standard deviations in parentheses) for prototype, seen, and unseen exemplars were 6.69 (0.74), 6.18 (0.78), and 5.68 (1.14) respectively. Bonferroni paired-samples *t* tests indicated that prototypes were rated higher than seen exemplars which were rated higher than unseen exemplars (see Table A2 in Appendix A for exact statistics). There was a nonsignificant trend of distance, *p*=.10. Overall, faces studied as a pair of far exemplars were rated slightly higher than those studied as a pair of near exemplars.

Two interactions reached statistical significance. The Familiarity x Exemplar Type interaction that was only a non-significant trend in Experiment 1 was significant in the current experiment, *F*(2,46)= 3.40, *MSE*= 0.82, *p*=.02. As can be seen in Figure 2.5, familiarity training increased ratings to prototypes and seen exemplars by similar amounts, but had less of an impact on unseen exemplars. Bonferroni paired-samples *t* tests on the difference scores between levels of familiarity (familiarized – unfamiliar) revealed that the familiarity difference was greater for prototypes than unseen exemplars, *t*(23)= 3.64, *SD*= 1.21, *p*=.007. The difference between levels of familiarity was also slightly greater for seen faces compared to unseen faces although this difference did not approach significance, *p*=.195. Prototypes and seen exemplars did not differ from one another in terms of the difference between levels of familiarity.

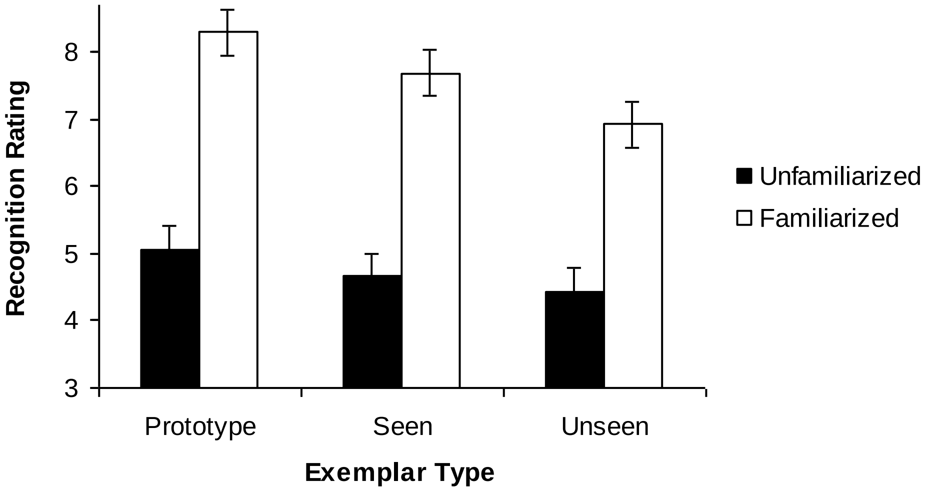


Figure 2.5. Mean recognition confidence ratings of prototype, seen, and unseen exemplars in Experiment 2 comparing faces that had been part of the familiarization training to those that had not. Error bars represent 95% within-subject confidence intervals.

There was also an interaction between distance and exemplar type, *F*(2,46)= 22.01, *MSE*= 1.17, *p*<.001. Like Experiment 1, the general pattern was that unseen exemplars were the lowest rated type of exemplar in the near condition, but seen exemplars were the lowest rated in the far condition. Bonferroni post hoc paired-samples *t* tests confirmed that unseen exemplars were rated lower than prototypes and seen exemplars in the near condition (see Table A3 in Appendix A for exact statistics). However, the differences among exemplar types in the far condition were not significant after adjustment for multiple comparisons. The Bonferroni comparisons also indicated that unseen exemplars were rated lower in the near condition than the far condition, *t*(23)=4.87, *SD*=1.47, *p*=.001.

Analysis of Prototype Effects

As in Experiment 1, an additional repeated measures ANOVA was ran to further examine the effects of familiarity and distance on the strength of the prototype effect (ignoring the unseen condition). The main effect of familiarity was significant and in the expected direction, *F*(1,23)= 179.46, *MSE*= 2.62, *p*<.001. The main effect of distance just missed significance, *F*(1,23)=4.07, *MSE*=1.18, *p*=.056. Near faces (*M*=6.59, *SD*=0.68) tended to be rated higher than far faces (*M*=6.27, *SD*=0.88). The main effect of exemplar type was significant, *F*(1,23)= 15.37, *MSE*= 0.82, *p*<.001. Prototypes were rated higher than seen exemplars. Importantly, no interactions approached significance (all *p*'s> .2). Thus, the two-way interactions of exemplar type with familiarity and with distance in the main analysis were primarily driven by ratings made to unseen exemplars and had no reliable influence on the ratings of prototypes relative to seen exemplars.

Discussion

In Experiment 2, participants were unable to reject the unseen prototype even if they tried to memorize each of the exemplar images in the study phase in exact detail. Furthermore, even after a familiarization training, they still could not learn to reject the prototype face. However, the strength of the prototype effect was unaffected by whether the face was part of the familiarization training or not.

In fact, just as in Experiment 1, participants were actually worse at correctly rejecting the unseen prototype as one of the seen exemplars from the study phase after familiarization training on that identity. Familiarization training had the effect of raising recognition ratings to all versions of a given identity (i.e. prototype, seen exemplar, unseen exemplar). This is perhaps surprising considering participants were made fully aware of the way exemplars of the same identity would vary (internal features shifted up or down), and they were also warned a prototype exemplar might be shown later during the recognition test. One might argue that participants were simply unable to distinguish the prototype from one of its seen exemplars because the difference between the images was too subtle to be detectable. However, the recognition rating data suggest otherwise. Participants reported being *more* confident of having studied the prototype image compared to an actually seen exemplar, so it was not the case that participants were insensitive to the differences produced by this method of image manipulation.

Even the unseen exemplar was more confidently rated as having been seen before after familiarization. Although familiarization raised recognition ratings more for prototypes and seen exemplars compared to unseen (near) exemplars, I did not expect additional learning from the familiarization training to make participants worse at correctly rejecting unseen exemplars falling outside the range of variation of the studied pair of exemplars of the same identity.

Since intentional learning still produced prototype effects comparable to those in Experiment 1, it is possible that if an averaging mechanism was underlying the observed pattern of results, this mechanism may be driven largely by the visual similarity of exemplars. Put simply, the system will average together any two exemplars that look similar even if the viewer attempts to retain a distinct representation for each. This may not be the case if exemplars are learned by a different method like categorization training with feedback. Obviously categorical influences must play some role in an abstraction mechanism for prototype formation to be a plausible account of face recognition. Otherwise, it would not be possible to eventually learn to tell identical twins apart. This raises two important questions leading from the results of Experiments 1 and 2. First, if participants are given categorical training with feedback to help them build separate representations of two similar exemplars, will they still recognize the prototype more confidently? This question will be addressed later in Chapter 3. Second, was the degree of learning in the familiarization training sufficient to produce an observable change in the magnitude of the prototype effect? This question was the focus of the next experiment.

Experiment 3

Experiment 3 had a straightforward purpose. If the familiarization training in Experiments 1 and 2 had no effect on the prototype effect, then was this because it was not a strong enough manipulation of familiarity? Even though, in the previous experiments, the number of exposures to familiarized faces was 4 times that of the unfamiliar faces, participants still may have treated them similarly to the unfamiliar faces. Perhaps learning in the familiarization phase could still have been considered minimal. In Experiment 3, the amount of familiarization training that participants received was doubled.

Method

Participants

Participants were 24 University of Victoria undergraduates enrolled in an introductory psychology course. Participants received extra credit points in their class for their participation. It took participants approximately 40 min to complete the experiment.

Materials and Apparatus

Stimuli.

Same as in Experiment 2

Equipment.

Same as in Experiments 1 and 2

Design

Design was the same as in Experiment 2. However, to be clear, note that the exemplars in the familiarization training were presented eight times each rather than four times each as in Experiment 2. Also, the current experiment used an incidental learning procedure as in Experiment 1 rather than an intentional learning procedure as in Experiment 2.

Procedure

The procedure was kept almost identical to that in Experiment 1. There were only two differences. The familiarization training was doubled and four more identities were added to the stimulus set (two familiarized and two unfamiliar).

In addition to rating each instance on intelligence, confidence, honesty, and attractiveness, participants also rated how responsible, playful, kind, and sociable they thought each person in the training phase looked. Thus, the familiarization training was extended from four blocks of trials in Experiment 1 to eight blocks (Experiment 3). As before, each identity was depicted by exactly two different exemplars, and each of these exemplar pairs was only presented once per block.

As a reminder, the task at test was to rate the familiarity of each image of each face based on how sure you are that the same exact image had been presented in the study phase. The rating was from 1-10 (1=*sure new*, 10=*sure old*). One small detail differed from Experiment 1 in the way the instructions at test were explained to participants. During the test instructions of Experiment 3, participants were shown an example face as a +2 and -2 exemplar. Next, they saw all five variants of that identity on the same screen, in order, from -4 to +4. They were told that if they actually had seen this person's face in the study phase (which they had not), in the test phase, they should only give high ratings to exemplars which were identical to the ones seen in the study phase (+2 and -2 in this example). If a -4, 0, or +4 exemplar is shown in the next phase, and you are absolutely certain you saw this person's face before, you should, despite this, rate the face low if you think the particular image of this person looks slightly different than before. The example was explained to participants one more time assuming they had seen the +4 and -4 exemplars in the study phase. The example face was added to the test instructions to make them aware of how slight the differences might be among exemplars of the same face.

A summary of the number of presentations in each phase may be helpful since the number of identities was increased from 20 (Experiment 1) to 24 (Experiment 3). In the training, 12 identities were shown as exemplar pairs. Each member of the pair was presented eight times each throughout the training. The study phase contained the 12 identities from the training phase plus 12 unfamiliar identities. Exemplar pairs were presented once each in the study phase. In the test phase, 24 identities were presented. Each identity was presented once as the unseen prototype, once as the seen exemplar, and once as the unseen exemplar.

Results

Main Analyses

Mean ratings given to the test faces in each of the experimental conditions are shown in Figure 2.6. A repeated measures ANOVA was conducted to analyze whether the factors familiarity, distance, and exemplar type had any influence on recognition ratings.

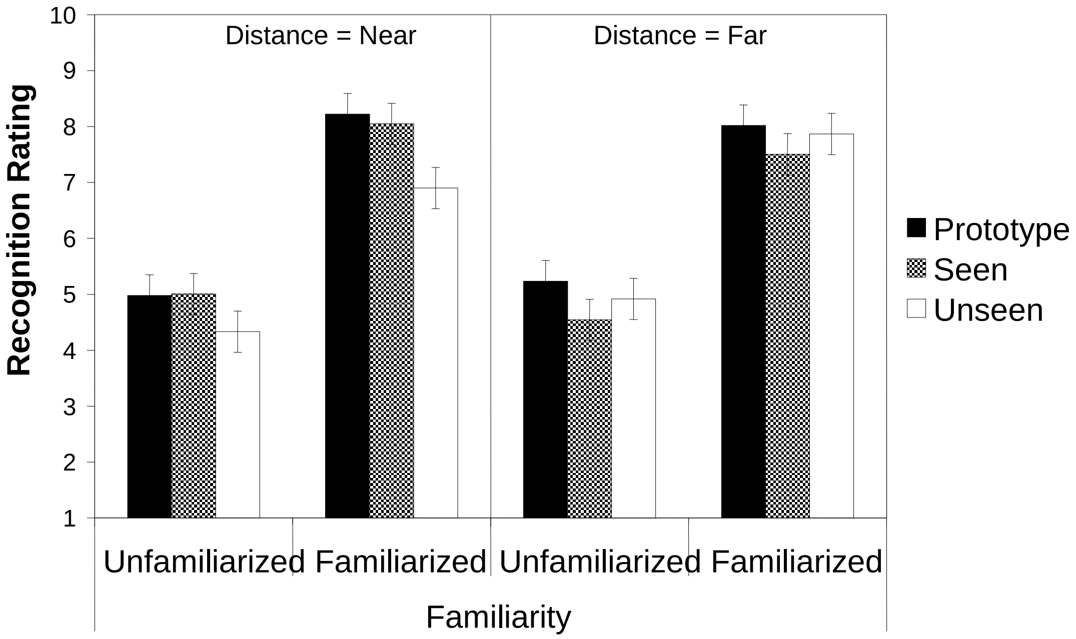


Figure 2.6. Mean recognition confidence ratings (1=*definitely never seen before*, 10=*definitely seen before*) of prototype, seen, and unseen exemplar types across levels of familiarity and distance between studied exemplars in Experiment 3. Error bars represent 95% within-subject confidence intervals.

There was a main effect of familiarity on ratings, *F*(1,23)= 218.41, *MSE*= 2.82, *p*<.001. As expected, familiarized faces were rated significantly higher than unfamiliar faces. There was no main effect of distance, *F* > 1. Near faces were rated as high as far faces. There was a main effect of exemplar type, *F*(2,46)= 17.02, *MSE*= 0.52, *p*<.001. Mean confidence ratings (with standard deviations in parentheses) for prototype, seen, and unseen exemplars were 6.61 (0.94), 6.28 (1.13), and 6.01 (1.25) respectively. Post hoc Bonferroni comparisons revealed that prototypes were rated higher than seen exemplars, *t*(23)=3.20, *SD*= 0.52, *p*= .012. Seen exemplars were, in turn, rated higher than unseen exemplars, *t*(23)= 2.62, *SD*= 0.51, *p*=.046.

There was one significant interaction. Distance interacted with exemplar type, *F*(2,46)= 15.18, *MSE*= 0.65, *p*<.001. As in Experiments 1 and 2, prototype faces were rated highest at both levels of distance. However, seen faces were rated higher than unseen faces when they were near, but the reverse was true when they were far. Bonferroni post hoc comparisons found that ratings to prototypes did not differ across levels of distance, *t*>1; but seen exemplars were rated significantly higher in the near condition than the far condition, *t*(23)=2.73, *SD*=0.90, *p*=.036; and unseen exemplars were rated significant higher in the far condition than in the near condition, *t*(23)=3.54, *SD*=1.07, *p*=.006.

There was also a trend towards the three-way interaction, Familiarity x Distance x Type, *p*=.074. As in Experiments 1 and 2, this was mainly due to the unseen near faces receiving less of a recognition ratings boost compared to all other conditions.

Analysis of Prototype Effects

As in Experiments 1 and 2, an additional repeated measures ANOVA was ran which further examined the effects of familiarity and distance on the strength of the prototype effect (ignoring the unseen condition). The main effect of familiarity was significant and in the expected direction, $F(1,23)= 214.50$, $MSE= 2.03$, $p<.001$. The main effect of exemplar type was significant, $F(1,23)= 10.25$, $MSE= 0.54$, $p<.01$. The main effect of distance was observed only as a trend, $F(1,23)=2.51$, $MSE=1.08$, $p=.13$. However, the main effects of distance and exemplar type were qualified by a two-way interaction, $F(1,23)= 6.56$, $MSE= 0.52$, $p=.02$. Although prototypes were rated higher than seen exemplars overall, this difference was clearer for far faces than for near faces. Bonferroni post hoc paired samples t-tests revealed that the prototype effect was significant in the far condition, $t(23)=3.83$, $SD=0.77$, $p=.005$, but not in the near condition.

Discussion

The results of this experiment were highly consistent with those of Experiments 1 and 2. Overall, participants were more confident about having seen the unseen prototype in the earlier study phase compared to the exemplars they really did see in the study phase. The magnitude of the prototype effect was not influenced by the amount of familiarization training in the learning phase. However, similarity (i.e. distance) of exemplar pairs in the learning phase did influence the strength of the prototype effect, and this will be returned to later in the discussion.

The main purpose of this experiment was to probe for differences in the strength of the prototype effect due to the level of familiarity training in the learning phase. Since the null effect of familiarity on the prototype effect found in Experiments 1 and 2 left open the possibility that the familiarity manipulation had not been strong enough, the amount of learning trials in the initial familiarity training phase was doubled in the current experiment. By the end of the learning phase, each exemplar of a familiarized identity was presented nine times vs. only once for exemplars of an unfamiliar identity. Although the familiarization training was effective at raising recognition confidence ratings, the increase in ratings was approximately equal for both prototype and seen exemplars. Therefore, even when the familiarity training was doubled (relative to Experiments 1 and 2), it still made no difference to the magnitude of the prototype effect.

Unlike Experiments 1 and 2, the results of Experiment 3 indicated stronger prototype effects for faces learned in the far condition compared to the near condition. The simplest explanation of this result is that seen near exemplars (+/-2) are recognized more confidently because they are more similar to the prototype (0) than seen far exemplars (+/-4). Assuming a prototype model, participants would store an average representation of each identity during the learning phase which would correspond to the prototype (0) stimuli shown in the test phase. Thus, being more similar to the stored prototype, the near (+/-2) faces were recognized more than the far (+/-4), but since the average value was 0 in both learning conditions, near and far prototypes (0) were recognized equally well. What is left unexplained by this account is the reason why Experiments 1 and 2 did not show the same pattern of results.

Further examination of the pattern of means suggests that distance had the general effect of increasing recognition ratings of prototype and seen exemplars in the near condition relative to the far condition. The exception to this pattern is unfamiliar prototypes which, in Experiment 3, showed a trend for lower ratings in the near condition relative to the far condition. This trend may provide a clue because in Experiment 2, unfamiliar prototypes also did not show the basic near over far pattern. In Experiments 2 and 3, participants were shown example stimuli and informed how the seen exemplars would differ from the prototype and unseen exemplars (in Experiment 3, learning was incidental, but participants were fully informed about the nature of the test phase). Being aware of how subtle the differences were among exemplars of the same face may have affected participants' confidence or perhaps how they used the ratings scale. In fact, it is possible that the effect of distance on the prototype effect may be due to floor effects caused by the task instructions. Confidence of unfamiliar faces was at floor in all conditions except for +/-4 test faces which were rated with low confidence as having not been seen. This account is consistent with the data across all three experiments. Confidence ratings were higher in Experiment 1; thus, the prototype effect was not modulated by distance.

The overall pattern of results of Experiments 1, 2, and 3 could be summarized as the 0 (unshifted) images rated highest, followed by +2/-2 and +4/-4 respectively. This pattern is what would be expected if participants were matching test items to stored prototype representations of each different face identity and not retaining the individual exemplars in memory.

Chapter Discussion

The experiments in the current chapter were designed to investigate whether the prototype effect in face recognition is affected by the level of familiarity participants have with the different face identities. Experiments 1-3 tested this idea by giving participants preexposure to half of the identities in an initial familiarization training phase before the study phase proper began. In the study phase proper, the other half of the identities which had not been presented in the familiarization training were presented in addition to those that had just been familiarized. At test, participants rated how sure they were that the image on the screen was identical to one of the studied images. For each of the identities in the study phase, one seen exemplar, one unseen exemplar, and the unseen prototype were presented in the test phase. Across three experiments, the level of familiarization did not alter how confident participants were that the unseen prototype had been presented in the study phase relative to one of the actually seen exemplars. Thus, the prediction, based on a prototype account of memory, that the strength of the prototype effect would increase after additional familiarization was not supported.

The main finding of Chapter 2 was that the prototype effect in face recognition is obtainable under conditions of both minimal and extended learning of identity exemplars. Previous studies (Bruce et al., 1991; Cabeza et al., 1999) have found the prototype effect in face recognition under conditions of minimal learning, but none had tested whether the effect would still be present at higher levels of learning. Experiments 1-3 replicated the prototype effect found at low levels of learning and extended this finding to higher levels of learning. So although it could not be concluded for certain whether learning or familiarity can influence the magnitude of the prototype effect, there was strong evidence that increased learning would not cause the prototype effect to disappear.

Why were the predicted results not obtained? It was reasoned that familiarity would increase prototype effects because different exemplars are more likely to be attributed to the same individual if the viewer is familiar with that person's face. If there is a mechanism that averages different exemplars of the same individual, then this mechanism will be more likely to combine different exemplars of a familiar face compared to a less familiar face because those exemplars are more likely to be perceived as the same identity. Perhaps this prediction failed because identity information was highly consistent among exemplars of the same face. Thus, an additional training phase did not increase the likelihood that participants would treat the two exemplars as belonging to the same person's face because it was obvious even on the first presentation.

It is difficult to ascertain how well the pattern of results found in Chapter 2 generalizes to other circumstances. The current findings may have been dependent on the way the identity variants were constructed and the task at test. Using an internal feature shift manipulation to create different exemplars based on the same face identity had a few notable consequences. First, the external features of the face were held constant among exemplars of the same identity. Since external features, especially hairstyle, are important to recognizing unfamiliar faces (Bonner, Burton, & Bruce, 2003), prototypes, seen, and unseen exemplars all contained a lot of unaltered information that would support recognition. Even the internal features, and the spatial relations among them, were identical across all exemplars (including prototypes) of the same identity. It was only the vertical positioning of the internal features, moved as a unit, on the face shape that differentiated same-face exemplars. If the mechanism underlying the prototype effect relies on a signal that two exemplars are the same person's face, then it may have been important that the facial features of different exemplars of the same identity were identical. Perhaps this made it easier for prototype formation to occur or for both studied exemplars to be retrieved together regardless of which image of a particular face was presented in the test phase. It is possible that additional learning in the training phase may produce a different pattern of results if exemplars of the same face vary subtly and smoothly in both texture and shape information.

Because images of the same identity all had the same internal and external features, all test images will probably engender a feeling of familiarity to some extent. Furthermore, participants were told that there would be no new people's faces in the test phase, only new images of the same people seen in the study phase. If all test images belonging to the same identity looked so similar to one another, how could participants be sure of which images were exactly the same as in the study phase and which ones were not? The results indicate that they could not. Participants confidently rejected neither the unseen prototype nor even the unseen exemplar. The sense of general familiarity participants got from the images shown at test might have influenced participants' ratings of confidence at least to some extent. Participants could not rely on successful retrieval of a memory of a prior occurrence of an exemplar as an indication that the test image was indeed one of the studied images because there was no way to know that the test image was truly identical or just too similar to be distinguished. Thus, they fell back on a subjective evaluation of how familiar each face seemed to them. As more average looking faces tend to be judged higher in terms of their subjective familiarity (Bartlett, Hurry, & Thorley, 1984), to the extent that the prototypes were perceived as more typical, they may have been at an advantage in this type of task. Bruce et al. (1991) pointed out that the default unshifted zero position of the facial features may give an appearance that is more typical within faces of the general population. Thus, it is possible that the prototype effects observed in Experiments 1-3 were partially due to the tendency of prototype faces to be slightly more typical looking.

However, there is evidence that the prototype effect in face recognition cannot be explained merely as a consequence of face typicality. Bruce et al. (1991, Experiment 4) attempted to rule out the confound of typicality by shifting the internal features of the faces down from the default position (0) in steps of 3 pixels. In the study phase, participants saw the -3, -6, -12, and -15 shifted exemplars of each face identity. At test, participants made a forced choice between the unseen -9 and the unseen 0 versions belonging to the same person's face. Participants chose the -9 prototype 82% of the time. In two other experiments (Experiments 7 and 8), Bruce et al. also ruled out the possibility that participants were responding to the average configuration of all faces in the experimental set. They presented three types of identity sets. In the study phase, the "up" set showed +3, +6, +12, and +15, and the "down" set showed -3, -6, -12, and -15. The third type of set was the "bimodal" set. It showed +9, +9, -9, -9. The prototype configuration of entire set of faces was still 0; however, only faces in the bimodal condition also had an identity prototype of 0. Faces in the up and down conditions had identity prototypes of +9 and -9 respectively. Each of the identities shown in the study phase was presented in the test phase as a -9, 0, and +9 at some point in the test series. In general, participants rated +9, -9, and 0 faces highest in the up, down, and bimodal conditions respectively. Thus, Bruce et al. showed evidence that participants were storing and accessing configural information specific to each of the studied identities rather than relying on typicality or a global configural prototype to make their ratings. Furthermore, it was the prototype of each identity that participants responded most strongly to rather than the most typical looking face or the average configuration of faces in the experimental set. The results of their experiments provide persuasive evidence that the prototype effect in face recognition is observable even when the unseen prototype is not the most typical exemplar of an identity.

It is possible that typicality may have a partial influence on participants' confidence ratings even if it is not the primary determinant. In another study, not reported in the present work, I tested whether identities centred around the default unshifted (0) face would produce larger prototype effects than identities centred around prototypes that had been shifted up or down from the default position. Half of the identities were left in the "normal" configuration as they were in Experiments 1-3. The other half were made "extreme" by shifting all the variants in each identity set either up or down by four units of distance. Thus, in the test phase, the internal facial features of half of the unseen prototypes were in the zero position, and half were in the +/-4 position. The results showed nothing. The only exception was a main effect of familiarity (used the same familiarization procedure as Experiment 1) which was unsurprising. No prototype effects were found for identities in either the normal or extreme conditions. However, this experiment was also designed to test whether an identity-based face recognition task would produce a prototype effect. Therefore, there were a number of differences between this unreported experiment and Experiments 1-3. The task at test was different. Rather than an image recognition confidence rating, a face/identity recognition confidence rating was used instead. Although it was not expected to do so (see Cabeza et al, 1999), judging the prior occurrence of an identity, rather than an image, may have affected the pattern of results. While the role of the type of task at test will not be discussed any further in this chapter, the lack of prototype effects in this experiment may provide an important clue in interpreting the results of later experiments (Chapter 4) using a similar recognition task, and this point will be returned to in the General Discussion (Chapter 5). Also, in order to avoid ceiling effects, the hair was cropped out of each image in the test phase. Cropping may have affected the perception of the internal feature positioning since it shortened the distance from the internal features to the top of the face/head (cropping often removed a small part of the top of the forehead as well as the hair). The nature of the recognition task also required a number of distractor faces to be used as well. Experiments 1-3 never included any novel identities in the test phase. Because the normal faces did not replicate the prototype effects found in Experiments 1-3, the role of typicality in the earlier experiments could not be interpreted from the results of the experiment that was not reported. Furthermore, since all typicality effects were null results anyway, the experiment was not reported as it did not provide any useful information. However, in Chapter 3, experiments will be reported that used stimuli in which the prototype and seen exemplar did not systematically differ in terms of face typicality.

One noteworthy finding from Experiments 1-3 was that not only were participants unable to correctly reject the unseen exemplar with even a moderate degree of confidence, but after familiarization training on a face, they were actually worse at doing so. Although familiarization did increase participants' ability to discriminate seen from unseen exemplars (for unseen exemplars outside the range of the seen exemplars), participants nevertheless became *more* confident that an unseen exemplar was one of the studied exemplars if the studied exemplars were shown more times during the learning phase compared to less times. This suggests that participants were unable to ignore the familiarity of the identity even if the face was presented as an unseen exemplar.

Since results of Experiments 1-3 all showed an interaction between the type of exemplar shown at test and the distance (of the internal feature position) between the pair of same-face exemplars learned at study, it is worthwhile to discuss what was causing this pattern to emerge. First of all, remember that a near face was studied as +2 and -2 and a far face was studied as +4 and -4. The unseen exemplar shown at test was always taken from the pair not seen in the study phase; thus, +/-2 for far faces and +/-4 for near faces. Now the pattern was that seen exemplars were rated higher than unseen exemplars in the near condition, but the opposite was true in the far condition. Prototypes were rated higher than seen and unseen exemplars at both levels of distance. Hence, the pattern of results could be expressed as: confidence ratings were highest in response to the average prototype of the two studied exemplars, and ratings decreased as similarity of the test image to the prototype decreased. This finding is highly consistent with a pure prototype account of face recognition. According to this account, only the prototype of an identity is stored in memory, and recognition is achieved by matching the probe exemplar to the different prototype representations. Recognition performance should increase as similarity between the probe and the closest stored prototype representation increases.

Alternatively, the higher similarity of the two studied near exemplars may have led to a higher net activation level because the seen exemplar would be a near perfect match to the stored trace of its presentation in the study phase, but also a moderate match to the other near exemplar. All things being equal, in both distance conditions, the seen exemplar highly activates its own trace in memory; and thus, the only difference is how much additional activation will be received from the other studied exemplar of the same face. Hence, a seen exemplar was rated higher if it was learned as a pair of near, rather than far, exemplars because it was more similar to, and thus more likely to recruit, the other studied exemplar of the same identity. This is how an episodic model might account for the interaction between distance and exemplar type.

One final remark will be made about the outcome of the experiments presented in this chapter. It is clear that the familiarization training did increase recognition of the seen exemplars at least when measured by recognition confidence ratings. It did not, however, increase how well the seen exemplars could be distinguished from other unseen exemplars most notably the prototype. Because the prototype was rated *higher* than the seen exemplars, the possibility is ruled out that recognition ratings are driven by the degree of match between the probe image and the nearest <u>exemplar</u> stored in memory. Likewise, there was no solid evidence that participants “built up” separate representations for each of the studied exemplars in the familiarization training. That is at least in the sense that the representations of exemplar pairs seemed to generalize to unseen exemplars (including prototypes) equally well regardless of the level of familiarity despite the task being to recognize the precise image. Thus, whether the combining happens during encoding or during retrieval, the traces of the two studied exemplars seem to be used together during the recognition process. In conclusion, there was no evidence to suggest that the familiarization training either encouraged or discouraged different exemplars of the same identity to be stored or retrieved separately. Perhaps this is why it did not modulate the prototype effect. If categorical learning could improve participants’ ability to discriminate exemplars, would the prototype effect still be observed? This is one of the questions explored in the next chapter.

Chapter 3: The Role of Context and Identity Information in Prototype Formation

Chapter Introduction

The purpose of Chapter 3 was to identify and test an experimental manipulation which would predict different outcomes depending on the theoretical model being considered. Episodic, or instance-based, models of memory were contrasted with prototype-based models. As discussed in Chapter 1, the main difference between these two viewpoints is that in episodic models every experienced event is represented in memory by its own trace, whereas in prototype-based models, an average representation is abstracted during encoding and stored explicitly in memory. Thus, a recognition advantage for a person's average face could be explained by the summed activation across episodic traces (e.g. Hintzman, 1986) or by its close correspondence to that face's prototype representation in memory.

Differential support for either an episodic or prototype account of memory is notoriously difficult to obtain from experimental results. Predictions based on these two accounts are almost always identical. Although, it might be impossible to ever provide unequivocal evidence for or against prototype formation, obtaining a result that poses a serious challenge to either theoretical viewpoint is crucial to developing plausible models of memory. To this end, in the experiments of this chapter, the context and identity information associated with a face during learning were manipulated. Ascribing identity and applying context to the faces will affect how they are processed and therefore influence the strength of the prototype effect. This will place constraints on how a plausible model of memory might operate.

To understand how attaching names to faces might affect the internal representations we form, imagine meeting a pair of identical twins. If you meet them separately, you might never know they are actually different people and thus never learn to tell them apart. However, given information about their identity, you can learn to distinguish them. Until then, according to a prototype account, the faces of both twins might be represented by a single prototype face in memory. To identify each twin accurately, one must build up a separate prototype representation for each of them.

In contrast, an episodic account maintains that a memory trace is laid down for every experienced event, or episode, with either twin. These traces will become activated if one of the twins is encountered. If the viewer does not know they have been seeing twins, memory traces of both twins will likely be highly activated when recognizing either twin alone. But if you are trying to learn to tell the twins apart, information of the identity of the twin will be stored as part of the episode of viewing their face. This identity information will become activated when trying to recognize this twin in the future. The basic idea is that viewing twin A should activate traces of twin A more than traces of twin B because twin A traces are, by and large, more similar to the current experience (i.e. the probe) with twin A. For the sake of simplicity, let us assume the names of the twins are the only nonface identity related information relevant to the task of learning to tell the twins' faces apart. Since activation of a memory trace activates all *relevant* information stored in the trace, the associated name of the face will become available as well. Since traces with twin A's name will be more highly activated and thus contribute more to the retrieved content, the face will be identified as twin A. Although to the extent that identity information is not important to the retrieval task, a probe image (e.g. an image of Twin A) may not selectively activate traces containing the same identity information whilst suppressing activation to traces containing conflicting identity information. For example imagine flipping through a photo album of a family with identical twin boys whom you are at least moderately familiar with. Several photos were of the twin's most recent birthday party. One photo showed Twin A's face as he blew out the candles on his cake. On your next visit to the family's home, you notice a framed picture which you do not remember from your last visit. You wonder if you have seen it before, so you go over to get a good look at the photograph it contains. It was a photo of Twin B's face as he blew out his birthday candles, and you falsely recognize it as one of the photos you saw in the album. Although you may have been able to correctly identify the twin in either photograph, when the task was image recognition, you failed because the probe image was being compared to how similar it was to previously seen photographs. Whereas, if the task was to verify the identity of the twin ("is this Twin A"), activation could be effectively restricted to traces of faces containing Twin A identity related information, and information specific to the particular image would become less important.

To emphasize the differences between the prototype and episodic accounts, remember that the episodic account maintains that every experience is encoded and stored in memory, whereas the prototype account maintains that the central tendency alone is sufficient to represent a concept or category. Furthermore, while an episodic model is capable of abstracting the central tendency of a category's members, this abstraction does not occur until the time of retrieval. Abstraction in a prototype model always occurs at the time of encoding whenever an exemplar of that category is experienced. At the time of retrieval, an episodic model will activate all stored memory traces in parallel. The strength of activation of each trace is dependent on its similarity to the stimulus probe. In contrast, a prototype model would never involve mass parallel activation because each of the representations stored in memory already reflects a number of past experiences, and the sole purpose of the abstraction mechanism is to differentiate faces on the basis of their identity.

According to a prototype account, all learned categories are represented by their own prototype representation. As applied to the case of face recognition, proponents of this account assert that if the system is operating as it should, a prototype representation should be abstracted across exemplars of the same person's face but not across exemplars of different people's faces (Bruce, 1994; Cabeza et al., 1999). The assumption here is that facial identity is the relevant level of categorization for face recognition to operate on. Although there may be other perceptual categories of faces represented by prototypes (e.g. gender, emotion, race etc.), they do not play a role in recognizing the identity of a person's face. A standard prototype account would not allow for a representation of, say, John's face to be composed of a prototype of John's face looking happy and another prototype for John's face looking sad. Although happy John and sad John are both valid categories, a prototyping system will only abstract a single "John" face, and this representation incorporates all instances of John's face regardless of emotional expression or any other contextual factor. Having a prototypical happy John face stored in memory might potentially aid recognition of John's face when he happens to be smiling and in a good mood. However, the task of face recognition is not to recognize and discriminate happy John from sad John; it is to recognize and distinguish any instance of John from any other person's face. Since the human face recognition system has evolved to discriminate between different identities and generalize across instances of the same identity, identity is the predetermined level of categorization that abstraction operates on (although see Bruce

1994 for a discussion about the necessity of multiple prototypes representing different viewpoints of the same face in developing models of face recognition based on image averaging).

In contrast, categories do not exist in memory outside of a context according to an episodic account. Category knowledge comes into being at the time of retrieval and is constructed from the pool of episodic traces stored in memory. The recognition probe event provides the context for retrieval. This approach maintains that there is no need to specify in advance which categories will be helpful to represent in memory for later use. If all experienced events are stored in memory, then category abstraction can occur anytime that category knowledge is needed. After the abstracted information has been retrieved and used for some immediate purpose, the individual episodic traces are retained in memory and can be used again and again for many different purposes. If the task is to recognize the face, an image of John’s face will activate traces of episodes with similar looking faces primarily those of past experiences with John’s face. Although all traces in memory are presumed to be activated to some extent, the degree of activation of all but the most similar traces is thought to be negligible. An aggregate representation is constructed reflecting the shared attributes of the highly activated traces. This aggregate then functions as a prototype representation for comparison with the probe stimulus. However, unlike a prototype, a representation of John’s face in an episodic model of memory will be transient in the sense that it will be a little different each time it is retrieved because it must be constructed anew. The precise image of John will affect how activated each trace becomes, and thus the ease of recognition. Many other factors besides similarity to the probe stimulus affect how strongly a trace is activated. Recency, task demands, and contextual information are three of the most important.

The experiments in this chapter explored how contextual information associated with a face might influence how it is learned, represented, and recognized. The experiments were designed to make contrasting predictions depending on whether a prototype or episodic account of memory is assumed. One key question raised in this chapter is whether identifying information like a person's name plays a special role in guiding the abstraction process, or whether it functions the same as other contextual information.

Experiment 4

Experiment 4 tested whether contextual information has any effect on the strength of the prototype effect in face recognition. Contextual information (as the term is used here) could refer to any type of information presented with a face that is not considered to be a face or part of a face.[3] For example, the butcher's shop and the counter he stands behind provide a context for recognizing the butcher's face as does the smock he always wears while working. Understanding the effect of context is an important step toward understanding how we represent faces because the role of contextual information in the face recognition process differs between prototype and episodic accounts of memory.

---

[3] Although it is possible that a face could serve as a type of associated context for another face if the two faces are seen together, I restricted the type of contexts under consideration to only include nonface contexts. If a face is associated with another face it raises the concern that information from the two faces may be miscombined during encoding or retrieval especially if the two faces are similar (see Busey & Tunnicliff, 1999) or presented simultaneously (see Reinitz & Hannigan, 2001). See the memory conjunction errors section in Chapter 1 for a discussion relating to this issue.

According to a prototype account of face recognition, an exemplar of a face is recognized by matching it to prototype representations stored in memory. If the exemplar is similar enough to one of the stored prototypes, it will be recognized. Note that this is purely a face to face comparison, and the derived similarity value does take into consideration how similar the current (i.e. probed) context is to contexts of previous encounters with that face. Thus, in a prototype model of face recognition, identity prototypes are stored context-free.

In contrast, in an episodic approach, context is important to face recognition. Context is part of the event information stored in each episodic memory trace. In the case of face recognition, each trace contains information about a face plus the context the face was seen in. Context is also present in the retrieval probe. Thus, the probe-trace similarity of both the face and the context determines the activation level of each trace (although similarity of the face information would be weighted more heavily than contextual information in determining the overall match). Also, if a person's face has been learned in a fairly consistent context, the studied context can be retrieved from memory in response to a probe that does not contain the studied context. Of course, seeing a face outside its normal context may make it slightly more difficult to recognize. A close match between the face part of a trace and probe will produce enough activation to lead to successful recognition most of the time. It is important to keep in mind that the saliency of an encoded context varies, and the effect of context is particularly dependent on the saliency of contextual information in the probe.

In this experiment, contexts were applied to the faces during the learning phase as well as in the recognition test phase. The applied contexts were drawings of cartoon bodies. The bodies were much smaller than the face images which gave the overall image a resemblance to a style of caricature art in which the subject's head is disproportionately large compared to his body. The design is very similar to the basic design used by the experiments in Chapter 2. This time, the participants learned four different face exemplars from each identity set. Half the participants learned all four face exemplars of a given identity with the same body. The other half of the participants learned two face exemplars with one body and the other two exemplars from the same identity set with another body.

This design also manipulated whether a test face was presented with or without context at the time of retrieval. Predictions vary depending on the details of the model, but if the abstracted prototype representation is stored context free, then the strength of the prototype effect should not be affected by a contextual manipulation. In contrast, an episodic model in which context is encoded in each episodic trace would predict a different pattern of results depending on the manipulated context. In particular, episodic models would predict that prototype effects would be weaker when a face is learned in different contexts and one of those contexts is reinstated at the time of recognition compared to if no context was presented at test or if the face had always been learned with the same context. Reinstating the context at test increases the amount of information contained in the probe that will be used to retrieve the corresponding information from similar traces stored in memory. A trace that matches closely to the probe in terms of both the face and the context will be activated more when the context is well represented in the probe than when the probe contains very little contextual information. The higher the total activation produced by all traces in memory, the more powerful the signal of familiarity will be in response to the probe.

Thus, when a seen exemplar is presented at test with its context reinstated, it should produce more activation than if context had not been reinstated. Furthermore, recognition of a prototype probe, which relies on similarity to traces of multiple exemplars, will not benefit as much from context reinstatement if the face was learned with two different contexts because traces with a context that conflicts with the probe will be inhibited.

Method

Participants

Participants were 48 University of Victoria undergraduates enrolled in an introductory psychology course. Twenty-four participated in the consistent context group and 24 participated in the inconsistent context group. Participants received extra credit points toward their course grade as compensation for participating. Each participant was tested individually and completed the experiment in a single session lasting approximately 25 min. Participation was voluntary, and at the end of the session, participants were fully debriefed as to the purpose and theoretical basis of the experiment.

Materials and Apparatus

Stimuli.

The stimuli consisted of photographs of human faces and cartoon drawings of human bodies wearing clothes. Displays were constructed by positioning the face just above the body centred on a white background. Faces were 10 cm wide and 13 cm high with an image resolution of 28 pixels/cm.

All face stimuli were greyscale computer generated photographs of young adult faces. All face photographs were cropped within an oval frame that preserved internal facial features but blocked hair, ears, and neck. An in house Principal Component Analysis (PCA) generator program was used to generate the face photos to specification (See General Methods in Chapter 1 for a description of how the generator operates). This method was used to generate 12 sets of face photographs; each set contained a collection of different variants around a central prototype. The apparent gender of the generated faces was constrained by using components only from females for half of the face sets and only from males for the remaining sets.

To generate each face set, an initial random face was generated by picking component values at random from within the normal range of the population of faces used to produce the model. Despite this, faces were occasionally produced that looked abnormal, due to an unusual combination of parameters. In that case, the program was simply rerun to generate a new set. The variations, described in more detail below, were generated by adding random amounts to the image and shape parameter sets to give a direction of change, and then constraining the distance, as measured within the parameter space, to be fixed, to a value found by trial and error to give suitably different images.

In the end, the generator produced a set of eight exemplars for each different face identity, collectively termed an *identity set*. One exemplar, termed the *prototype* (P), was the average of all the other exemplars combined. Two more exemplars, termed the *local averages* (L), were generated equal distances from P along a trajectory in opposite directions. Another pair of exemplars, termed the *seen instances* (S), was generated in a similar fashion around each L. Thus, each of the two L's were the average of two S's, and P was the average of the two L's and at the same time was the average of the four S's. Finally, two more exemplars were generated in opposite directions along a trajectory orthogonal to the first trajectory. One of these two was selected to serve as an *unseen instance* (U), which was not the average of any of the other exemplars just described. Even though only one U was included in the final set of stimuli, two were generated, so that if there were any oddities or artefacts in the appearance of the first face but not the second, the second face could be used as a replacement. Figure 3.1 illustrates the relationships among the eight different exemplars for each identity set.

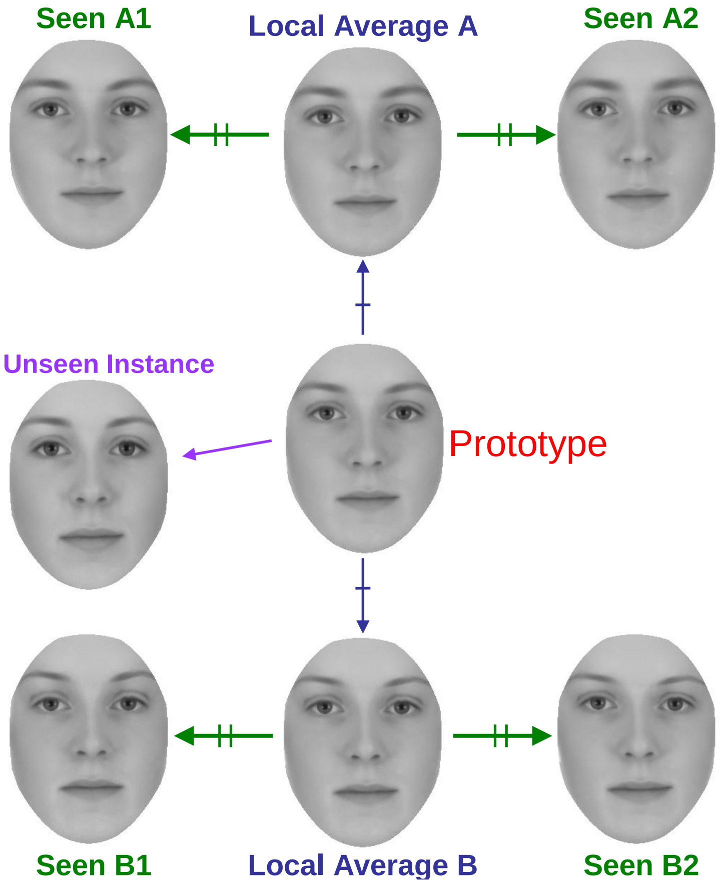


Figure 3.1. Relationships among the different exemplar types within a face identity set. Arrows of the same colour represent changes of equal distance along a trajectory in PCA space. The prototype is the global average of all exemplars in the set excluding the unseen instance. The four seen instances shown in the study phase are labelled as A1, A2, B1, and B2. Local averages A and B are the averages of the A and B pairs of seen instances respectively.

Context pictures consisted of scanned images of cartoon bodies taken from children's activity books. Figure 3.2 shows two examples of a context body paired with a face. Twelve bodies were taken from the book Tracing Fun: Funny Faces and 12 more from the book How to Draw Cartoons. All context bodies were black and white line drawings of human bodies wearing clothes from about the neck down. Some of the cartoon bodies were wearing casual clothes, but many were wearing uniform or costume type clothes such as an American football uniform or a laboratory coat. The scanned images were altered using Adobe Photoshop 6.0 to remove all heads, unwanted accessories, and imperfections. Small changes needed to be made to a few of the images in order to create a set of half plausibly male and half plausibly female bodies. Thus, half of the context bodies were assigned to be male and half to be female based on how "male" or "female" the body and clothes looked at my discretion.

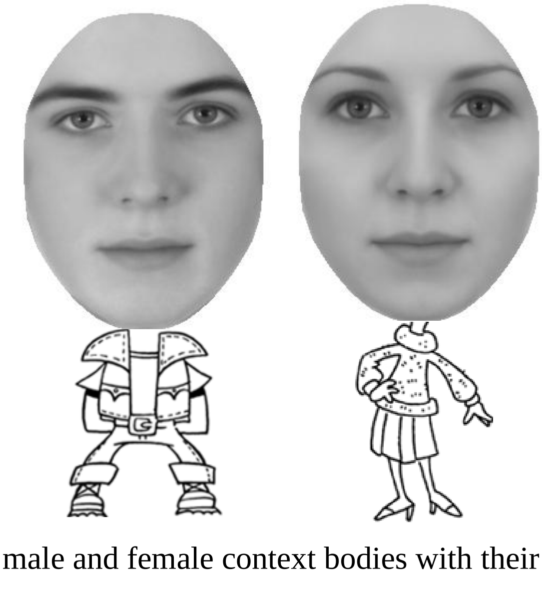


Figure 3.2. Example male and female context bodies with their faces from Experiment 4.

Equipment.

All aspects of the experiment were conducted using a Hi-Grade Notino3600 windows lap top with a 14" monitor at 1024 x 768 resolution. The program E-Studio from the

E-Prime software package (version 1.1) was used to control the presentation of stimuli and to record responses made using the laptop keyboard.

Design

The three independent variables in this experiment manipulated the context the face was shown with in the study phase, the context the face was shown with in the test phase, and the type of face exemplar shown in the test phase. The context of the face in study was manipulated between groups, while the context and type of face shown at test were both within-subject variables.

Participants were randomly assigned to one of two study context groups, either the consistent study context or the inconsistent study context group. In the consistent study context group, participants saw all exemplars within each identity set with the same context body, context A, during the study phase. In contrast, participants in the inconsistent study context group saw half (one pair) of the face exemplars with context body A, and the other half with context body B. Put simply, in the consistent study context group, the same person's face was always presented in the same context, but in the inconsistent study context group, a person's face was shown in two different contexts- half the time in one, half the time in the other.

In the test phase, four different types of face exemplars were shown from each identity set. The four levels of face exemplar type were Prototype (P), Local Average (L), Seen Instance (S), and Unseen Instance (U). Remember that P was the average of the four S faces and also the average of the two possible L faces. It is also important to note that no P, L, or U items had been seen before in the study phase. Only S instances were identical to face photographs that were actually shown before in the study phase.

There were three levels of test context: test context same, test context different, and test context none. Test context same trials presented a face exemplar with the same context body it had been shown with before. Test context different trials presented a face exemplar with a different context body than it had been shown with before. Finally, during test context none trials, the face exemplar was presented alone on the screen with no body to serve as a context. In the test context different condition, the different context was always the body of a face from a different identity set shown in the study phase. This meant that the context bodies presented in the test phase had all been seen before in the study phase, but only half of them were seen with the same face.

Taken all together, the experiment used a 2 (Study Context) x 4 (Exemplar Type) x 3 (Test Context) mixed factorial design with a total of 24 conditions. Each participant contributed data to half of the conditions according to which context study group they were assigned to.

Recognition memory performance was measured by recognition confidence ratings collected during the test phase. A 9 point Likert-type scale (1=*definitely never seen before*, 9=*definitely seen before*) was used to assess how sure participants were that each exemplar had been shown previously in the study or training phases. As in the other experiments reported thus far, judgments were made about the particular image rather than the identity of the person’s face in general.

Counterbalancing.

Each identity set was randomly paired with two context bodies (Contexts A and B) with the constraint that all three were of the same gender. Gender of the face was always kept in equal numbers across all grouping of items for counterbalancing or other similar purposes described below.

To reduce the likelihood of presenting exemplars of the same identity set consecutively in the study phase, four sublists were constructed such that each one contained exactly one of the four S faces from each identity set. The order of the sublists within the series of study faces as a whole was random for each run of the series, as was the order of the S faces within each of the four sublists.

In the test phase, the P face, one of the L, one of the S, and the U face were presented from each identity set. Since there was only one P and one U face in each identity set, all participants saw the same ones. Because there were two L and four S faces, one of each type was presented at random for each identity set and for each participant. Similarly, when Contexts A and B were both shown in the study phase, and the same context was to be presented in a test trial, either Context A or B was presented at random to each participant for P and U faces since there were two possible context bodies to choose from. Order of the P, L, and S faces was counterbalanced such that the first, second, or third picture shown within an identity set was equally often a P, L, or S face. The counterbalancing groups of faces were rotated across participants so that every possible ordering of P, L, and S could be shown of every identity set. The list of test items was divided into three blocks: one block for the 1st, 2nd, and 3rd presentation of exemplars from the same identity set. Since U faces acted more or less as distractor items, the order of their presentation was only controlled such that there was an equal number of U faces presented in each of the three blocks, and groups of U faces were rotated around the three blocks across participants. Test context was also counterbalanced such that each block of trials and each type of face (P, S, L, and U) contained an equal number of same, different, and no contexts. Again, test context was counterbalanced across participants so that within each block every identity set could be shown in all possible combinations of face exemplar type and context.

Procedure

Participants were randomly assigned to be in either the consistent or inconsistent study context group. Participants sat at a comfortable distance, centred in front of a laptop. The following procedure was completed twice through by each participant. The values reported below for both number of trials and item types are the totals across both runs of the experiments. Six identity sets were used for the first run and a different six for the second run.

Study phase.

Participants were told that they would see a series of pictures of faces with cartoon bodies on the screen and that they were to rate each picture from 1(*lowest*) to 9(*highest*) on different personality and perceptual characteristics. Ratings could be based on information from both the face and clothes, but ratings could not be based solely on the clothes alone. Participants were warned that they should try to remember how each picture looked because there would be a facial recognition test following the ratings task. It was suggested to them that trying to remember the clothes might help them recognize the pictures of the faces later because some of the faces would be wearing the same clothes in the recognition test. It was pointed out that no two different people would be shown wearing the same outfit in the study phase. Then an example identity set was shown to the participants in a figure/diagram to give the participants an idea as to what the stimuli at study and at test would look like.

In the study phase, a series of face exemplars (S faces) with a context cartoon body were shown sequentially on the screen. The series consisted of 4 exemplars (S faces) from each of the 12 different identity sets and its associated context body. The series was repeated five times, each time in a random order. For each run of the series participants rated a different characteristic about the appearance of each face-body display. The order of the first four ratings (intelligence, kindness, attractiveness, and confidence) was randomized for each participant. The fifth rating was always about how well a particular face was matched with a particular outfit. Before the start of each run of the series, a message appeared on the screen instructing participants as to what characteristic to rate. A reminder of the values of the 1-9 scale for each attribute appeared with labels at the bottom of the screen. Stimuli remained on the screen until participants entered their response rating by pressing the appropriate number key on the laptop’s keyboard. Trials advanced automatically with each typed response. No data were collected during the study phase.

Test phase.

Participants were asked to rate how well they could recognize each photograph of each face from 1(*definitely NOT seen in study phase*) to 9(*definitely seen in study phase*). The test phase followed the same procedure as the study phase except for the following differences. Participants were asked to base their rating only on the face part of the stimulus display rather than the clothes as well. Participants were told that sometimes the face would be shown in the same clothes it was wearing before, but other times it would be wearing a different person's clothes or have no body at all. Also, participants were warned that their ratings should be based on the *exact* photograph of the face, rather than the face in general, and that all faces they were about to see were either identical to or only just slightly different from ones shown in the study phase (no entirely new people's faces).

Among the series of 48 faces shown at test, exactly four different exemplars (1 S, 1 L, 1 P, and 1 U) were presented from each of the 12 identity set. The computer recorded the recognition confidence ratings entered on each trial.

Results

Mean ratings given to the test faces in each of the experimental conditions are shown in Figure 3.3.

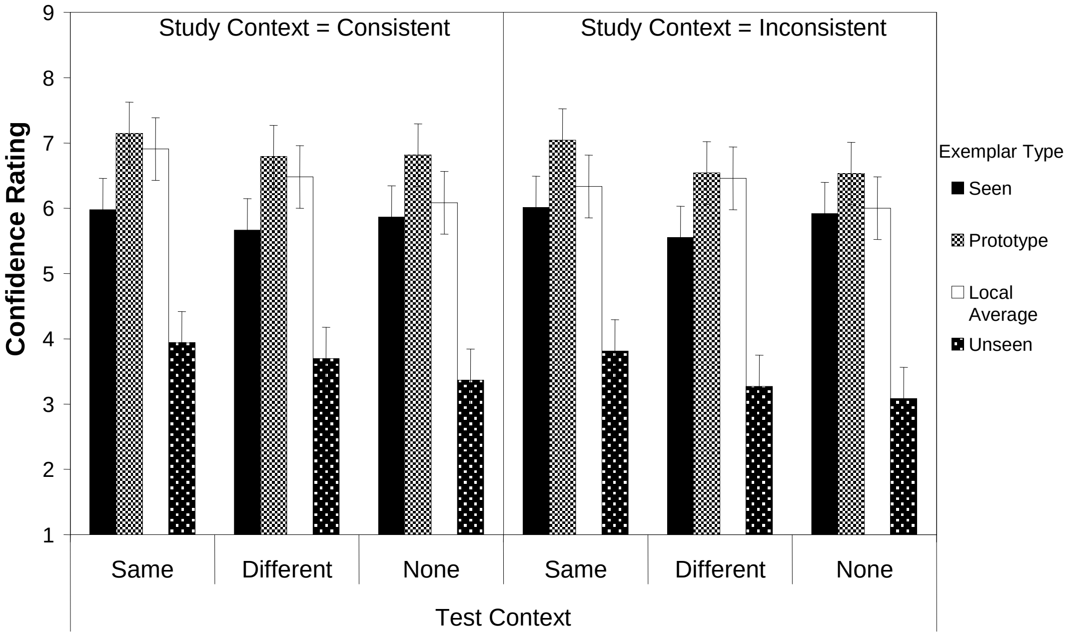


Figure 3.3. Mean recognition confidence rating (1=*definitely never seen before*, 9=*definitely seen before*) to face images in Experiment 4 as a function of exemplar type, context consistency in the study phase, and context congruency at test. The context for each test face was either the same as or different to its studied context, or absent. Error bars represent the 95% within-subject confidence interval and are appropriate for comparing patterns of means across conditions of exemplar type and test context (Loftus & Masson, 1994; Masson & Loftus, 2003).

A 2 (Study Context) x 4 (Exemplar Type) x 3 (Test Context) mixed factorial ANOVA was conducted to analyze the influence of these factors on recognition ratings. Study context was a between subjects measure. Exemplar type and test context were both repeated measures.

There was no main effect of study context on recognition ratings, *F*(1,46)= 1.62, *MSE*= 2.94, *p*= .21. Although faces studied with a consistent context (*M*= 5.73, *SD*=.44) were rated slightly higher than those studied with inconsistent contexts (*M*= 5.55, *SD*=.54), this difference was not significant.

There was a main effect of exemplar type on recognition ratings, *F*(3,138)=174.48, *MSE*= 1.76, *p*< .001. In order from highest recognition confidence to lowest was the prototype (*M*= 6.81, *SD*=.70), the local average (*M*= 6.38, *SD*=.79), a seen instance (*M*= 5.83, *SD*=.80), and an unseen instance (*M*= 3.53, *SD*=.99). This pattern of means is displayed graphically in the left panel of Figure 3.4. Bonferroni post hoc paired-samples *t* tests revealed that all possible pairwise comparisons were significant (see Table B1 in Appendix B for exact statistics).

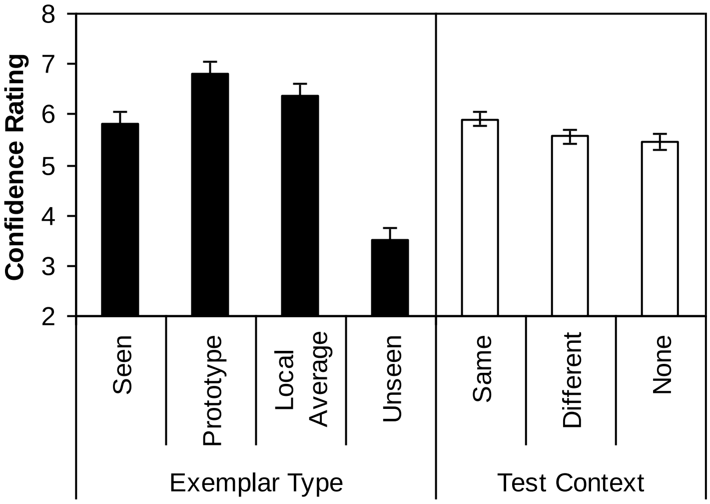


Figure 3.4. Mean recognition confidence ratings to face images in Experiment 4 varying in exemplar type (left panel) and test context (right panel). Error bars represent the 95% within-subject confidence interval constructed separately for each panel.

There was also a main effect of test context, *F*(2,92)= 10.26, *MSE*=0.99, *p*< .001. As can be seen in the right panel of Figure 3.4, when the same context (*M*= 5.90, *SD*=.67) was presented along with a face in both the study phase and the test phase, participants rated their confidence higher than when a different context (*M*= 5.56, *SD*=.66) or no context (*M*= 5.46, *SD*=.59) was provided at test. Bonferroni post hoc comparisons found that same test context faces were rated significantly higher than both different and no context, but different and no test context faces were rated equally (see Table B2 in Appendix B for exact statistics). None of the interactions were significant.

Discussion

The manipulation of context within neither the study phase nor the test phase had an effect on the strength of the prototype effect. Whether exemplars from the same identity were all studied with the same context body or two different ones made little difference to recognition confidence ratings at test. Reinstating the studied context at test increased confidence ratings overall. However, context reinstatement did not have a differential effect on images according to their exemplar type (prototype, local average, seen instance, or unseen instance).

Perhaps the most surprising finding was that confidence ratings showed such high sensitivity to the exemplar type of the test image. The difference in appearance among exemplars of the same face was often barely noticeable, and yet participants clearly rated their confidence highest to prototypes.

An overall prototype effect was observed, but its magnitude was consistent across all levels of the other variables. This was consistent with the prediction based on a prototype account of memory in which each face identity is represented by its own prototype stored independent of any associated context. The results were inconsistent with the prediction based on an episodic account of memory that context would weaken the prototype effect when the same identity was learned in different contexts, and one of these contexts was reinstated at test. Context reinstatement should increase activation to seen exemplars more than prototypes when exemplars of the same identity were learned with two different contexts. Since the prototype effect was not affected by the context presented at study or at test, then either context does not affect how we store and retrieve our internal representations of faces or context was not salient enough in the current manipulation to produce an observable change in magnitude.

Experiment 5

In Experiment 5, participants learned to associate a name with each of the face exemplars in the study phase. The purpose of the name learning task was to encourage participants to treat all exemplars with the same name as belonging to the same person’s face and to treat all exemplars with different names as belonging to different peoples’ faces.

Learning to combine instances of faces sharing the same identity is of utmost importance to the success of a prototyping mechanism. Therefore, learning identity related information in the form of face-name pairs could potentially guide the abstraction of identity-based facial prototypes. Therefore, based on a prototype account, it was predicted that prototype effects will be stronger for faces learned with the same name than faces learned with two different names.

An episodic model might abstract a prototype at the time of retrieval, and thus still show a strong prototype effect regardless of whether it was learned with one name or two names. In Experiment 4, it was predicted, based on an episodic model, that context would have an effect on the strength of the prototype effect when the context learned at study was reinstated at test. In this experiment, names could be thought of as a type of associated context information. However, no modulation of the prototype effect was predicted in the current experiment because names were not presented in the test phase, and the task at test was to recognize the specific image shown in the study phase. Therefore, name information was of little relevance to the task at test, so it should not alter the strength of the prototype effect.

Method

Participants

Participants were 48 University of Victoria undergraduates enrolled in an introductory psychology course. Twenty-four participated in the same name and 24 participated in the different name versions of the experiment. All other details regarding participation were the same as in Experiment 4. The only exception is that this experiment took approximately 30 min to complete rather than 25 min as in Experiment 4.

Materials and Apparatus

Stimuli.

The stimuli consisted of photographs of human faces and first names. Displays for the study phase were constructed such that the face and name were centred horizontally on a white background, but the face was moved upwards slightly to leave space for the name to be presented at the bottom of the display. No names were presented on the screen during the test phase.

The face stimuli were constructed using the same method as in Experiment 4. Therefore, 12 identity sets (6 male and 6 female) of different people's faces were generated, each consisting of 8 different exemplars (1 P, 2 L, 4 S, and 1 U). The only difference between face stimuli in this experiment from those in Experiment 4 was that in general, exemplars within each identity set in the current experiment were more similar to one another. Exemplars were made more similar to look like they could plausibly depict the same individual.

Names were arbitrarily chosen from a booklet of baby names. A total of 24 names were selected. Half of the names were male names and half were female names.

Names considered to be unisex were not selected. All names were two or three syllables in length. No two names shared the same initial letter.

Equipment.

Same as Experiment 4

Design

The two independent variables in this experiment manipulated whether a person's face was learned with one name or two names and also the type of face exemplar shown in the test phase. The number of names learned for each face at study was a between-subjects variable, while the type of face exemplar shown at test was a within-subjects variable.

Participants were randomly assigned to one of two name groups, either the same or different name group. In the same name group, participants saw all four exemplars within each identity set with the same name during the study phase. In contrast, participants in the different name group saw half (one pair) of the face exemplars with one name, and half with another name.

In the test phase, five different types of face exemplars were shown. The five levels of exemplar type were Prototype (P), Local Average (L), Seen Instance (S), Unseen Instance (U), and entirely new Foil faces (F). Foil faces were not included as a level of exemplar type in the subsequent analysis as they were only presented for task demand purposes. Faces belonging to the first four levels (P, L, S, and U faces) were part of the same identity sets that S faces were taken from in the study phase. As a reminder, no P, L, and U items had been seen before in the study phase. Only S faces were identical to one of the photographs actually shown before in the study phase.

Taken together, this resulted in a 2 (Name) x 4 (Exemplar Type) mixed factorial design with a total of eight conditions. Each participant contributed data to half of the conditions according to whether they were in the same or different name group. Recognition confidence was measured using the same confidence scale as in Experiment 4.

Counterbalancing.

Each identity set was randomly paired with two names (names A and B) with the constraint that all three were of the same gender. Like Experiment 4, the study series of names and faces was divided into four sublists, and the order within and among sublists was randomized.

In the test phase, the P face, one of the L, one of the S, and the U face were presented from each identity set. Selection of the individual items was made on the same basis as Experiment 4. The order of the P, L, and S faces was also counterbalanced as in Experiment 4, as was the presentation of U faces. Since this experiment also included entirely new faces as distractors, these faces were divided into three groups. One group was presented in each of the three blocks. Across participants, the three groups of new faces were presented in every possible order.

Procedure

As in Experiment 4, the procedure was actually completed twice through by each participant. The values reported below for both number of trials and item types are the totals across both runs of the experiment. Six identity sets were used for the first run and a different six for the second run.

Study phase.

Participants were told that they would see a series of pictures of faces, and the name of the person in each picture would be presented below the face. The task was to learn the names of the different people's faces. Participants were informed whether there would be two or four pictures of each different person's face. All pictures of faces sharing the same name were to be treated as being different pictures of the same person. Accordingly, all pictures with different names were to be treated as different people. Participants in the different name group were informed that there would be pairs of very similar looking brothers or sisters within the series of faces that might be difficult to tell apart. All participants were warned that after the learning parts of the experiment there would be a recognition test of the specific images they studied in the learning parts.

The study phase was divided into four rounds. The series of 48 study phase faces was presented once during each of the four rounds to give a total of 192 study trials. Items remained on the screen until participants pressed the space bar to advance to the next display. In Round 1, participants passively viewed each face in the series with its name one at a time on the screen. Round 2 also did not require participants to enter a response, and it had three displays per face item in the series. First, a face was presented alone on the screen, and participants were asked to guess its name silently. Once the space bar was pressed, the first letter of the name of the face appeared under the face to act as a memory cue. Finally, the third display showed the completed name below the face. In Round 3, the task was to name the face by entering the first letter of the name on the keyboard. Round 3 had two displays per face item in the series. The first display presented the face alone, the second, displayed the face and feedback about the response. The first display in Round 3 remained on the screen until participants entered a naming response which triggered the onset of the second display. Feedback appeared on the screen indicating if the preceding response was correct or incorrect in blue or red letters respectively. Feedback always included the complete correct name of the face. Round 4 was a repeat of the Round 3 procedure. In summary, Round 1 introduced the faces with their names, Round 2 helped with learning the names as well as the faces, and Rounds 3 and 4 targeted face naming and provided feedback to assist with fine grained discrimination.

Test phase.

In the test phase, participants were asked to rate how well they could recognize each photograph of each face from 1(*definitely NOT seen in study phase*) to 9(*definitely seen in study phase*). Participants were warned that their ratings should be based on the *exact* photograph of the face, rather than the face in general, and that there would be new, but very similar, pictures of the people they had just studied within the upcoming series.

There were a total of 60 faces in the test phase series. For each of the 12 identity sets studied 1 P, 1 L, 1S, and 1 U face was shown. Twelve F (Foil) faces that were not part of any identity set in the study phase were also included in the series. Faces were presented one at a time and remained on the screen until a number key was pressed on the keyboard indicating the recognition rating.

Results

Study Phase

In Rounds 3 and 4, participants responded with the correct name on 79% of the trials if they were in the same name group and on 52% of the trials if they were in the different name group. Although all letters of the alphabet were possible responses, participants in the same name group learned three names at a time of each gender. Participants in the different name group learned six names at a time per gender. Based on these numbers chance responding would lie at about 33% and 17% for the same and different name groups respectively. However, it is likely participants in the different name group could narrow their responses down so that they were effectively guessing between two names from the same identity set on many of the trials.

Test Phase

Mean ratings given to the test faces in each of the experimental conditions are shown in Figure 3.5. A 2 (Name) x 4 (Exemplar Type) mixed factorial ANOVA was conducted to analyze the influence of these variables on recognition ratings. Name was a between-subjects IV, and exemplar type was a within-subjects IV.

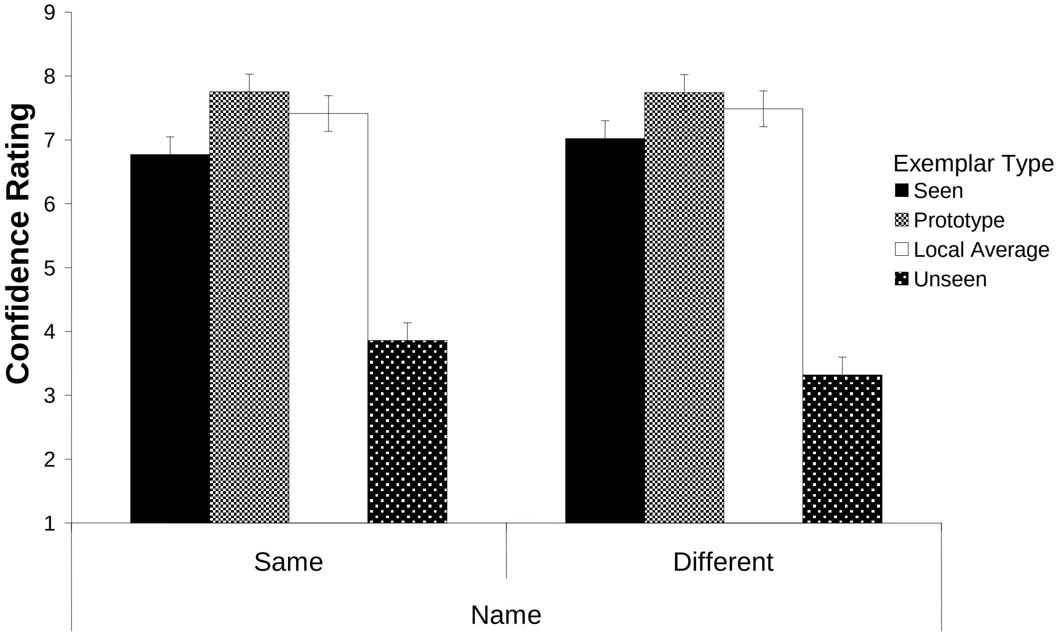


Figure 3.5. Mean recognition confidence rating (1=*definitely never seen before*, 9=*definitely seen before*) to face images in Experiment 5 as a function of exemplar type at test and whether exemplars of the same face were learned with the same name or two different names. Error bars represent the 95% within-subject confidence interval and are appropriate for comparing patterns of means across conditions of exemplar type.

There was no main effect of name on recognition ratings, $F(1,46)= .12$, $MSE= 1.26$, $p= .73$. Faces studied with the same name ($M= 6.45$, $SD=.53$) were rated equally as high as those studied with two different names ($M= 6.40$, $SD=.59$).

There was a main effect of exemplar type on recognition ratings, $F(3,138)= 356.43$, $MSE= .50$, $p<.001$. In order from highest recognition confidence to lowest was the prototype (P; $M= 7.74$, $SD= .69$), the local average (L; $M= 7.45$, $SD= .84$), a seen instance (S; $M= 6.89$, $SD= .85$), and an unseen instance (U; $M= 3.59$, $SD= .93$). Bonferroni post hoc paired-samples *t* tests revealed that the differences in all possible pairwise comparisons were significant.

The interaction between name and exemplar type was also significant, *F*(3,138)=2.78, *MSE*=0.50, *p*=.045. The pattern of means suggest that the interaction was driven by an effect of name on the confidence ratings at test for U faces but not for the other exemplar types. When faces were learned with the same name, confidence ratings of U faces were higher than when the faces had been learned with two names. However, this difference did not reach significance in Bonferroni post hoc comparisons which compared the difference between learning with the same name vs. different names across each level of exemplar type.

Also, participants rated distractor faces in the same name group (*M*=1.70, *SD*=.64) slightly higher than participants in the different name group (*M*=1.5, *SD*=.56). This difference was not significant according to an independent samples *t* test, *t*(46)=1.16, *p*=.25.

Discussion

As in Experiment 4, there was a very clear overall prototype effect. However, the naming manipulation did not affect the strength of the prototype effect. Whether the task in the study phase encouraged participants to treat the exemplars of a set as the same person or as two different people made no difference to how confident participants were that the unseen prototype was one of the images in the study phase.

The results of this experiment, taken on their own, were actually more consistent with an episodic account of memory rather than a prototype account. The prediction based on an episodic account was that naming in the learning phase should not influence image recognition in the test phase because names were irrelevant to the task at test. The task at test was simply to judge (i.e. rate one’s confidence) whether the test image was identical to one of the images studied earlier. Since the test faces were presented without names, and the task did not require names to be recalled, the probe and hence the retrieved information should contain very little, if any, name information. The information encoded in probe determines what information will be entered into the probe-to-trace similarity comparison. Thus, although traces of the seen exemplars from the study phase may have contained information about the associated name, this name information would not have affected the similarity between the probe and a given trace.

It was perhaps surprising that the naming manipulation did not have any effect on the strength of the prototype effect because a prototype account of face recognition predicts that naming should have an effect. If learning to associate different names with the two pairs of twin exemplars did indeed help participants to build separate prototype representations for each twin, one would expect the local averages to be rated highest at test for participants in the different name condition. Even if participants could not perfectly discriminate between the twins, one would expect at least some drop in ratings to the overall prototype from increased sensitivity to image changes especially those along the trajectory that discriminated the twins. Thus, the results suggest that either different identities are not represented by their own facial prototype, participants did not receive enough learning trials to produce an observable change in the strength of the prototype effect, or that the naming task does not in principle affect prototype abstraction regardless of learning level (at least with the stimuli used in this experiment).

In fact, participants in the different name group were only responding with the correct name (indicated by the first letter) on about half of the trials, so there is reason for the concern that participants did not receive enough discrimination training. Therefore, before abandoning the investigation into how learning name-face associations influences the way we store and retrieve internal representations of identity in faces, Experiment 5 was repeated using a different type of face stimuli and an increased degree of learning in the study phase.

Experiment 6

Experiment 6 is a follow up to Experiment 5 again using the name learning approach. Through learning to associate face exemplars with the same name or two different names, I aimed to manipulate how participants perceived identity in those faces. Although the naming task in Experiment 5 had no significant effect on the prototype effect, it is possible that the high similarity of exemplars of the same identity set made the name learning task ineffective. To investigate this possibility, Experiment 6 tested the prototype effect at three levels of exemplar similarity. This manipulated the similarity between the two named pairs of exemplars within a single identity set. When the pairs were very similar, it would be like seeing two photographs of each twin of a pair of identical twins. When the pairs were very dissimilar, they resembled pairs of photographs of two different people. It was predicted that the naming manipulation would affect the prototype effect most when exemplars were of moderate similarity. The reasoning was that exemplar pairs that were very similar might be too difficult to learn to discriminate, and faces that were very dissimilar might always be perceived as two different people regardless of whether they shared the same name or not.

This is an important follow up because if the different named faces are too similar to be discriminated, then both prototype and episodic accounts of memory would predict no effect of the naming manipulation. An episodic account would have predicted no effect of naming regardless of similarity. A prototype account would interpret poor twin discrimination as a failure to build separate prototype representations for each twin. Thus, no matter if the twins' faces were learned as one person or two, exemplars at test would still be recognized by matching to a single prototype. Because the point of manipulating identity information was to make contrasting predictions of prototype and episodic accounts, it was important to vary the similarity of the named exemplar pairs (i.e. the twins) to ensure it was not preventing us from observing an influence of the naming manipulation on the strength of the prototype effect.

Rather than PCA generated faces, Experiment 6 used full colour morphed images. These images are very realistic and rich in individuating information which might further increase the effectiveness of the name learning task.

Method

Participants

Forty-eight undergraduate students enrolled in an introductory psychology course at the University of Victoria participated in this study. Each participant was tested individually in sessions lasting approximately 30 min. At the end of each session, participants were fully debriefed and awarded partial course credit. Data from all 48 students were collected and analysed.

Materials and Apparatus

Stimuli.

Twelve face sets were constructed. The end products were full colour photographic quality images depicting faces of young adults. Images were approximately 9.5 cm wide and 13.5 cm high (29 pixels/cm resolution). They were displayed centred on a black background. Each face set consisted of a number of exemplars which could depict a single named identity (in the same name condition) or a pairing of two named identities (in the different name condition). Three versions of each face set were prepared which differed by the degree of similarity (low, medium, or high) between the two paired identities (if assuming the different name condition).

Regardless of the level of similarity, face sets were made by selecting two photographs of different people of the same gender and morphing them together by different proportions to create two morphed images that served as the local average images in the test phase (see Figure 3.6). Remember that in the different name condition, these local averages depict two different, although possibly very similar, people's faces.

Around each local average, two more variants were created giving a total of four training exemplars (two pairs) in each face set. These were created using Psycho Morph software which can take the difference between two images and apply it to a third image transforming it to emphasize the differences between the first two images (see General Methods section in Chapter 1 for more information about Psycho Morph). Two more photographs of different people of the same gender were used to serve as the first two images in the procedure just described. Whenever possible, large differences in skin tone were avoided when selecting these two photographs because the differences between them created the within variability for each pair in a face set.

Each local average was morphed by a certain amount equally in both the positive and negative direction of the difference between the first and second photographs (see Figure 3.7). All similarity levels (low, medium, and high) of the same face set used the same two “variation images” and morphed their local averages by the same amounts (physically the same amounts). In contrast, variants from completely different face sets all used different variation images, and the amount of morphing used to create the variants was individually determined for each face set.

Although only four exemplars in each face set were used in the study phase, a fifth and sixth were constructed for use in the test phase only. The fifth was the prototype, and fell at the midway point between the two local averages. The sixth was an out of range unseen exemplar, and it was made by creating a morph that fell two steps away from one of the local averages in the opposite direction of the other local average (if one imagines a morphed continuum divided into 10% steps between the original pair of faces chosen for each face set). Of course, the two local averages were also part of the face set, and they too were only shown in the test phase.

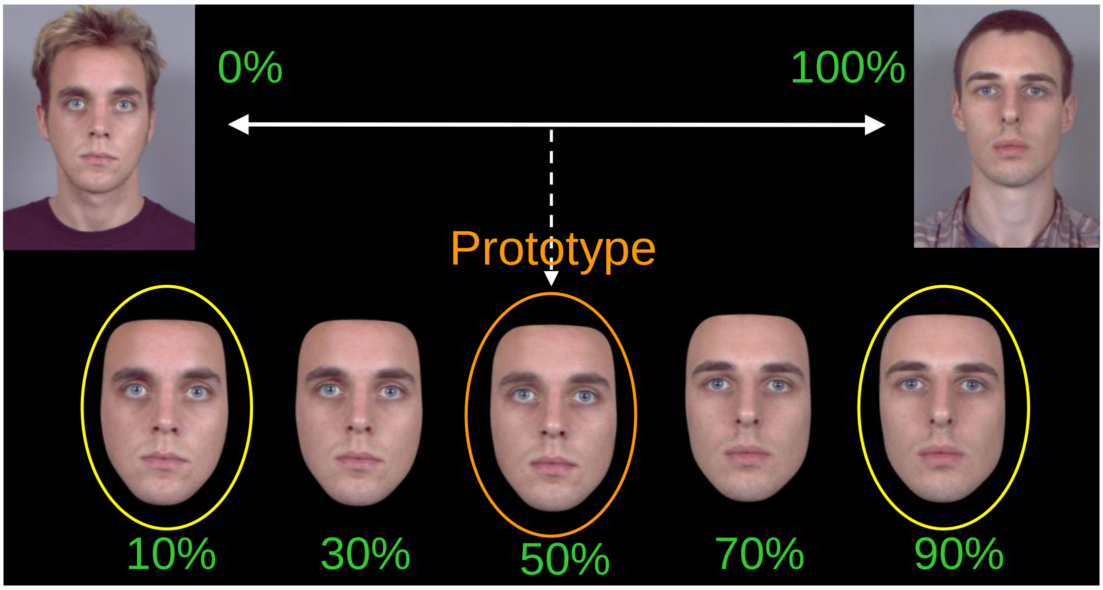


Figure 3.6. Example stimuli from Experiment 6 depicting the construction of the global and local prototypes. A full morphed continuum between two different faces was created using increments of 10% (only a subset is shown above). For each level of similarity (low, medium, and high), a pair of morphs was selected as the two local averages. If, for example, the faces circled in yellow were selected as the local averages for the low similarity condition, the global prototype would be the 50% morph because it is exactly half way between the two local averages.

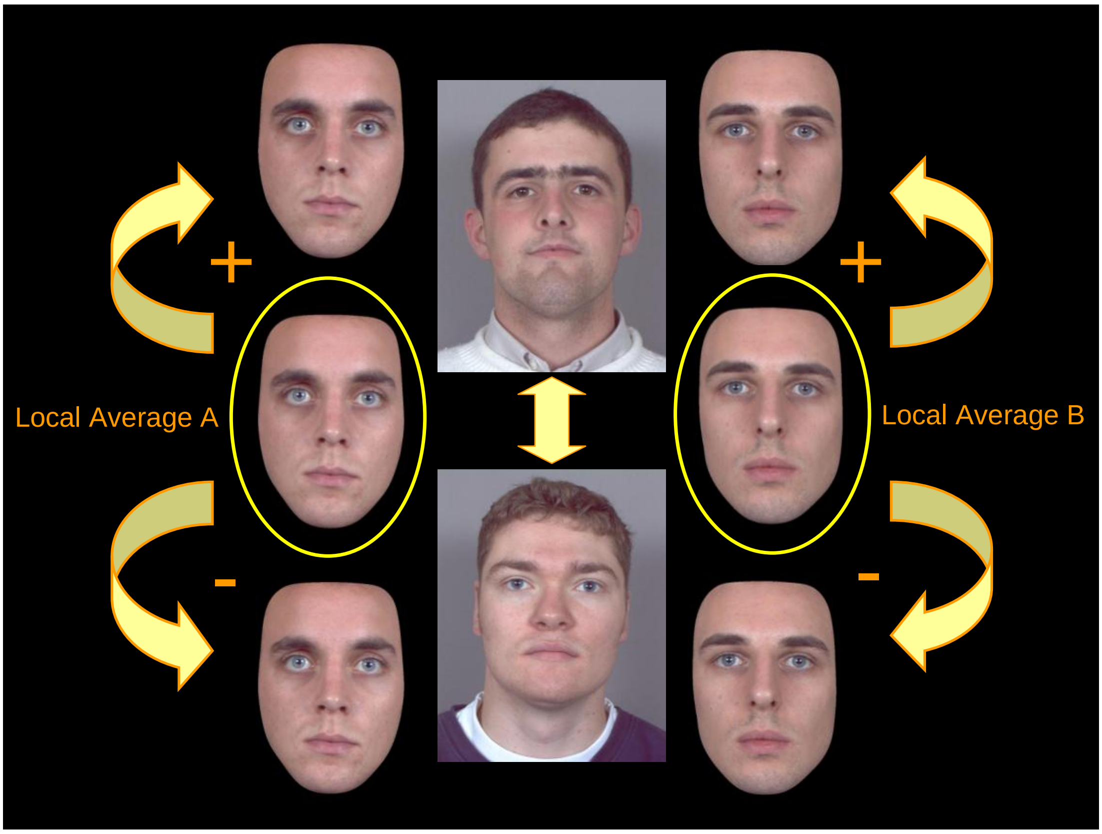


Figure 3.7. Illustration of the manipulation used in Experiment 6 to generate within-identity variation. Circled in yellow are the 10% and 90% morphs from Figure 3.6 shown here as a pair of low similarity local averages. Above and below the local averages are the exemplars presented in the study phase. These exemplars were made by taking the difference between the two "variation" images in the middle column and applying this difference to the local averages. Local averages were transformed in both the positive and negative direction which amounts to exaggerating the difference toward the opposite variation image.

Pilot testing.

Every time the morphing transform was used, the amount of morphing needed to be specified. Pilot testing was conducted to get a rough idea of how much morphing the stimuli needed to undergo in order to achieve the desired effect (to appear as images of the same person or different people). Participants were shown pairs of images that had undergone different degrees of morphing and were asked to classify them as either the same person but different photos, different people, or totally identical photos. One thing that became clear from doing the pilot study was that the perceptual effect of the morphing procedure depended on the similarity of the particular photos being morphed rather than just the quantitative value the software used to warp and blend the images.

Equipment.

Same as Experiment 4

Procedure

Participants were randomly assigned to 1 of the 48 counterbalancing conditions. The experiment was divided into separate study and test phases. Before participants began the study phase of the experiment, the procedures for both the study and test phases were outlined. Detailed instructions for the test phase were delivered immediately after completion of the study phase.

At the start of the session, participants were told they would be learning the names of several faces. They were warned about the image recognition test that would follow the learning phase. Participants were told the following: “All pictures with the same name are the same person. Some faces will have undergone extreme makeovers, so they will look very different from one picture to the next. On the other hand, some faces will look almost the same as another face with a different name. In this case, the different faces are probably siblings or twins.”

Since the entire experimental procedure was repeated twice, only the procedure for the first run through will be described. The second run through was identical to the first, but it used a different set of faces.

Study phase.

The run through consisted of three sub-blocks. In all sub-blocks, faces were presented one at a time against a black background in the centre of the computer screen. There were no time constraints placed on the participants. The screen remained unchanged until the space bar was pressed to advance. Order of presentation of the trials was randomized within each block for each participant.

In the first sub-block, all the faces in the study set were presented (6 sets x 4 exemplars= 24 images) with their corresponding names printed below. Participants were instructed to simply view the faces with their names and try to remember them. In the second sub-block, the study set was again presented only once, but this time the face appeared on the screen initially without a name, and participants tried to guess the name without actually responding. Pressing the space bar brought up the first letter of the name to provide a cue if the name could not be recalled from the face alone. Pressing the space bar once again brought the entire name into view. The third press of the space bar initiated the presentation of the next face in the series. The third sub-block presented the study set four times through (6 sets x 4 exemplars x 4 repetitions= 96 images). Initially, each face was presented alone on the screen. Participants pressed a key on the keyboard corresponding to the first letter of the name of that face. The key press triggered feedback to appear on the screen indicating if the response was correct or incorrect. If incorrect, the feedback displayed the correct response below.

Test phase.

In the test phase, participants completed a recognition confidence task. Recognition confidence was measured using a Likert-type rating scale from 1-9 (1=*definitely NOT* *seen in study phase*; 9=*definitely seen in study phase*). Participants were asked to make their ratings about how sure they were that they saw the particular *image* before rather than the person’s face in general. They were also informed that all images they were about to see would be of the same people they had just been learning, so there would be no new people’s faces. Participants were told that a rating of 1 meant that they were 100% sure they had never seen the test image before in the study phase. In contrast, rating of 9 meant that they were 100% sure they had seen that exact image before in the study phase no change in the image whatsoever. Participants pressed the number keys 1-9 on the keyboard to enter their responses. It was pointed out that 5, because it was the central value of the scale, was equivalent to having zero confidence.

As pointed out earlier, the entire procedure was conducted twice each time with a different set of faces. To avoid redundancy, only the first test phase procedure will be described. A series of 18 images were presented one at a time on a black background in the centre of the computer monitor. A black line with the numbers 1-9 below it was displayed below the image. This line was a reminder of the confidence scale. It was labelled “definitely not seen” and “definitely seen” beside the 1 and the 9 on the scale respectively. When participants pressed a number key to indicate their response, there was a 1 s ITI in which a black screen was displayed. Immediately following this short pause, the next image in the series automatically appeared on the screen. For each of the six faces learned in each study phase, three versions, a prototype, a local average, and an unseen exemplar, were presented during the test phase.

Design

The experiment had three within-subjects independent variables and one dependent variable. The three independent variables manipulated whether a face identity set was learned with one name or two names, the similarity of exemplar pairs within an identity set, and also the type of face exemplar shown in the test phase. Overall, the structure of the experiment followed a 2 (Name) x 3 (Similarity) x 3 (Type) fully within subjects design.

The two levels of name were different and same. In the same name condition, participants learned to associate all four studied exemplars within a given identity set with the same name in the study phase of the experiment. In the different name condition, participants learned to associate one exemplar pair within an identity set with one name and the other exemplar pair with another name. The three levels of similarity were labelled low, medium, and high. These labels referred to the level of similarity between the two pairs of studied exemplars from the same identity set. For example, high similarity in the different name condition would be like seeing twins because the images of the two twins were extremely similar. The three levels of exemplar type were prototype, local average, and unseen exemplar. Prototype refers to the average of all four exemplars of a given face identity set regardless of whether that set had been learned with one name or two names. Local average refers to the average of a pair of studied exemplars from the same face identity set. In the case of learning with different names, a local average, in theory, would serve as a prototype of a particular named person’s face because it would be the average of the two exemplars associated with that name in the study phase. Unseen exemplars were similar in appearance to at least one pair of studied exemplars, but they were not presented in the study phase. Importantly, unseen exemplars were outside the range of variation exhibited by studied exemplars from the same identity set. Note that all of the exemplar types shown in the test phase were technically unseen images since none of them were identical to one of the images presented in the study phase. However, the label unseen exemplar has been given to unseen images that did not correspond to the average of the seen exemplars in all our previous experiments, so for the sake of consistency, it was used again here.

The dependent variable was image recognition confidence. It was measured using a 9 point Likert-type scale (1=*definitely NOT seen in study phase*; 9=*definitely seen in study phase*).

Counterbalancing.

In total, participants received training on 12 different face sets. All participants saw exemplars from the same 12 face sets. The only difference in the facial images used among participants was the particular exemplars chosen from each face set. Three variations in exemplar sets existed for each face set. In other words, the exemplars of each face set differed depending on whether the face was to be shown in the low, medium, or high similarity condition for any given participant. Counterbalancing procedures ensured that, across participants, each face set was seen an equal number of times at each level of similarity and that each participant learned an equal number of face sets at each level of similarity.

The experiment was designed so that the face stimuli set had three different versions allowing each face set to be presented at each level of similarity (low, medium, or high) across the three versions. To balance the levels of name within each version, half of the face sets at each level of similarity were assigned to have one name (same), the other half, two names (different). To balance across participants, three more counterbalancing versions were made that duplicated the first three except for the assignment of same or different name was reversed. Thus, any difference in participant responses between the same and different named faces could only be attributable to the effect of the name learning manipulation since there would be no difference in the actual face stimuli being presented.

The 12 face sets were broken into two groups of 6, so the experimental procedure could be conducted twice through, once with each set. The only other important counterbalancing in the learning phase that has not been described thus far is that the order of the two experimental blocks was reversed. When block order was reversed, some small changes were made to reduce the impact of stimuli specific effects. One thing that was changed was the particular name selected from the two possible names assigned to each face set. For example, if the names assigned to Face Set 1 were Jerry and Mark, and Jerry was chosen as the name to be used in the same name condition, the reversed block order chose the other name, Mark, for use in the same name condition.

Results

Study Phase

Table 3.1 shows mean accuracy of participant performance in the name learning study phase of the experiment. Only the third sub-block performance is shown in this table as neither of the first two sub-blocks generated data.

	Similarity		
Name	Low	Medium	High
Different	0.85 (0.36)	0.73 (0.44)	0.63 (0.48)
Same	0.97 (0.16)	0.95 (0.21)	0.98 (0.12)

Table 3.1. Mean accuracy of naming faces in the study phase of Experiment 6 (standard deviations in parentheses).

Table 3.1 suggests that participants had little trouble distinguishing faces from different face sets because they almost always responded with the correct name to faces in the same name condition. They did, however, have some difficulty distinguishing between the exemplar pairs within a face set as would be required when naming faces in the different name condition. In this condition, as one would expect, the smaller the difference between exemplar pairs of a given face set, the less accurate participants were at naming them. Indeed, many participants were unable to correctly name the "twin" faces in the different name, high similarity condition more than 50% of the time. Keep in mind that the accuracy predicted by chance performance is ambiguous because all letters of the alphabet were possible responses, but only nine letters corresponded to names used in the study phase. Also consider that in the different name condition, if participants were guessing only between the two names associated with exemplars from the same identity set, they would be predicted to be correct 50% of the time due to chance alone.

Test Phase

Means of all experimental conditions are shown in Figure 3.8. A 2 (Name) x 3 (Similarity) x 3 (Exemplar Type) repeated measures ANOVA was conducted on confidence ratings to detect any differences due to exemplar type, learning a face with one name or two, and the similarity among exemplar pairs of the same face in the study phase.

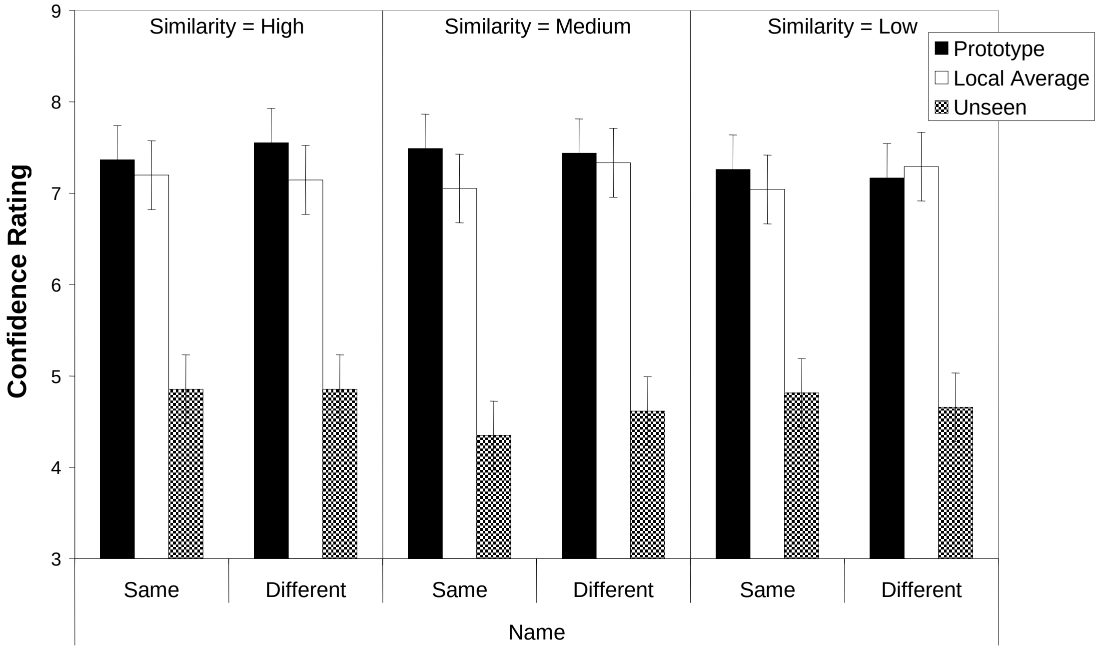


Figure 3.8. Mean recognition confidence rating (1=*definitely never seen before*, 9=*definitely seen before*) to face images in Experiment 6 as a function of exemplar type at test (prototype, local average, and unseen), similarity of same-face exemplar pairs, and whether same-face exemplar pairs were learned with the same name or two different names. Error bars represent 95% within-subject confidence intervals.

There was no main effect for name, *F*< 1. Faces from identity sets learned consistently with the same name (*M*= 6.38, *SD*=1.07) were rated equally as high as faces from identity sets learned with two different names (*M*= 6.45, *SD*=0.96). There was no main effect of the similarity between pairs of exemplars from the same identity set, *F*< 1. Overall, confidence ratings to test faces were approximately equal regardless of whether their identities had been learned from a set of high, medium, or low similarity exemplars. In order, mean confidence ratings for the low, medium, and high conditions were 6.37, 6.38, 6.50 (*SD*s= 1.13, 1.02, and 1.12 respectively).

The main effect of exemplar type was significant, *F*(2,94)=224.95, *MSE*=2.87, *p*<.001. Ratings to prototypes (*M*= 7.38, *SD*=.94) and local averages (*M*= 7.18, *SD*=1.03) were clearly higher than to unseen exemplars (*M*= 4.69, *SD*=1.37). This was confirmed using Bonferroni post hoc paired-samples *t* tests. Prototypes were also rated slightly higher than local averages. This difference failed to reach significance but only after a Bonferroni adjustment for multiple comparisons was made (see Table B3 in Appendix B for exact statistics).

None of the interaction effects were significant, all *F*s< 1.1.

Discussion

Of all the main effects and interactions, none turned out significant except a main effect of exemplar type. The prototype and the local average were both rated higher than the unseen exemplar. There was a borderline overall prototype effect (prototype over local average). This prototype effect was smaller than the ones observed in Experiments 4 and 5, but it was significant when distractors (i.e. unseen exemplars) were excluded from the data analysis. However, considering half of the faces in this experiment were studied with the same name and half with different names, it is perhaps surprising that any sort of overall prototype effect was found without any qualifying interactions (particularly Name x Exemplar Type). It is possible that the method of face-name learning used in this experiment was insufficient to promote categorical learning of the differently named faces. Perhaps if a different method had been employed, then a weakening or absence of the prototype effect might have been observed for the different named faces. That aside, the present results seem to suggest a more or less automatic process of image averaging that may rely mainly on similarity between one picture and the next without much heed from other contextual cues or identity information.

It may be worth noting that the low similarity condition was the only condition that suggested that the naming manipulation was affecting the strength of the prototype effect. The only trend in the reverse direction of a prototype effect was found in the different name, low similarity condition. For faces learned with two different names, participants showed better discrimination of exemplar pairs in the low than in the medium or high condition in the naming task in the study phase. Because of the pilot testing, we know that exemplar pairs in the low condition are judged as belonging to two different individuals. It is interesting, therefore, that despite appearing to be two different people, in the same name condition, low similarity faces showed a trend toward a prototype effect. Perhaps less learning is required to provoke participants to treat two different people's faces as belonging to a single person than to provoke participants to treat extremely similar faces (e.g. twins) as belonging to two different people. This idea raises the question of whether learning a face set with two different names could lower the prototype effect or if learning a face set with a single name could increase the prototype effect, or both. Since this experiment found no significant effect of name, it is also possible that the manipulation would have no effect. Even still, it would be worthwhile to test for prototype effects using the low similarity stimuli in a non-naming learning task such as those used in the experiments in Chapter 2. If two different people's faces give rise to the prototype effect, then what implications would this have for our understanding of the origins of the prototype effect in face recognition and for the mechanisms involved in recognition memory in general? This issue is considered in the General Discussion in Chapter 5.

Experiment 7

Given that the stimuli used in Experiments 4 and 5 were so similar, it was surprising that they produced such large overall prototype effects. Even though there was not anything obviously different in the construction or appearance of the prototype, by virtue of being the average of a set, it may have tended to look more typical of faces in the general population than the seen exemplars (and also the local averages). Therefore, Experiment 7 was designed to test whether the prototype stimuli were rated highly in Experiment 4 and 5 because they were the average of a face identity set, or if they were rated highly due to some unintended property of the stimuli. If ratings given to the prototype stimuli were influenced by some unintended property, then they should be rated highly (i.e. higher than seen exemplars) even when they are not the average of the seen exemplars shown in the study phase.

The current experiment used identity sets taken from Experiments 4 and 5. The recognition test phase remained essentially the same as well. The crucial difference was that this time, participants studied only one seen exemplar from each identity set rather than all four. Thus, the prototype of the identity set was no longer the average of the seen exemplars shown in the study phase. Likewise, the local average was no longer the average of one of the pairs of seen exemplars. In fact, none of the exemplars shown at test truly represented the prototypical image of an identity. Therefore, it was predicted that if the prototype effect in face recognition relies on the prototype stimulus being the central tendency (mean) of a set of varying experiences with a single individual’s face, then the “pseudo prototype” stimuli in this experiment should not be recognized higher than the seen exemplar because they were not the central tendency of the studied exemplars.

Method

Participants

Participants were 32 visitors to the University of Stirling. The experiment was conducted during the university's Open Day event where prospective students and their friends and family members can get information about the different programmes and services the university has to offer. Adults visiting the information table of the Psychology Department were asked if they would like to participate in a short experiment about face recognition. Those who were interested were tested individually at one end of the table. The experiment took about 3-5 min to complete. At the end of the session participants were debriefed and invited to ask questions. Participants were approximately16 – 65 years of age.

Materials and Apparatus

Stimuli.

A subset of face identity sets from Experiments 4 and 5 were used. Six identity sets from each of Experiments 4 and 5 were selected for use in the current experiment.

Equipment.

Same as in Experiment 4

Design

The only independent variable in this experiment was the type of exemplar seen at test. Although participants only studied one seen exemplar per identity set, they were tested on four in the subsequent recognition test phase. The four levels of exemplar type were prototype, local average, seen exemplar, and unseen exemplar. Importantly, these labels only refer to the exemplar type of an image as it was used previously in Experiments 4 or 5. Therefore, the current usage of these labels only has meaning in the context of Experiments 4 and 5. For each of the identity sets used in this experiment, the prototype, the local average, the seen exemplar, and the unseen exemplar were the images that served in Experiments 4 or 5 as the prototype, the local average, the seen exemplar, and the unseen exemplar respectively. Beyond that, the only other similarities are that the seen exemplar was indeed the image studied in the study phase and all other exemplar types were only presented during the test phase. Note that the prototype and local average did not actually correspond to the average of the studied exemplars, as in this case there was only one exemplar studied of each identity.

The dependent variable was recognition confidence and was measured by a rating using a Likert type scale from 1-9 (1=*definitely NOT seen in study phase*; 9=*definitely seen in study phase*). The task was to rate how confidently you believe the particular image, rather than the person's face in general, was presented in the study phase.

Counterbalancing.

Although there were a total of 12 identity sets used in this experiment, each participant only saw half of them. The 12 sets were broken down into 2 groups of 6 sets each. Half the participants completed the experiment using the first group of stimuli, and the other half completed it using the second group of stimuli. Also, within each identity set used, the particular seen exemplar shown in the study phase was counterbalanced across participants. Since there were four potential seen exemplars in each set, 4 counterbalancing groups were formed on that basis. Taken together, this made a total of 8 counterbalancing conditions: 2 (stimuli group) x 4 (seen exemplar). However, after 16 participants had been tested, the local average was switched from the more distal (i.e. less similar) local average to the more proximal one so that data could be obtained for both types of images. Ideally, this would have been a within-subjects factor, but this detail was overlooked in the initial design. Similarly, since there were two local averages, the number of counterbalancing conditions doubled so that half of the time the local average more proximal (i.e. more similar) to the seen exemplar was presented and half of the time the local average more distal to the seen exemplar was presented. The order of items presented in both the study and test phases was randomized for each participant.

Procedure

Study phase.

A series of six images of faces was presented, and participants were asked to memorize each picture exactly as shown. Participants were warned before the study phase began that after they had seen the faces, they would be asked to recognize the identical picture of each face from a series of similar pictures of the same person. An example of a seen exemplar (not use elsewhere in the experiment) was presented on the screen followed by a display of the exemplar just shown plus the three other seen exemplars in that identity set. The seen exemplar that was just shown had a red box around it to indicate that it was the target exemplar. It was explained to participants that although the four images may appear to look like the same person, the images themselves are all slightly different from one another, and that they should keep this in mind while they are studying the images. Another example of a set of four seen exemplars was shown at the bottom of the display. The examples were given just to give participants some idea as to how much variability they should expect within a face identity and between different people's faces.

Each participant saw one seen instance per face set from six different face sets. Faces were presented one at a time. The study phase was self paced in that the face stimuli remained on the screen until participants pressed the space bar when they felt ready to move onto the next face. There was an ITI of 500 ms during which a white screen was shown.

Test phase.

A series of 24 faces were presented, and participants rated their recognition confidence for each one. The ratings were made using a scale of 1-9 (1=*definitely NOT seen in study phase*; 9=*definitely seen in study phase*). Like the previous experiments in this chapter, the rating was made about the prior occurrence of the *exact* image rather than the person's face in general. Participants were warned again that some images would look very similar to ones they saw in earlier, but if there was even a slight difference between how the image appeared at test from the study phase, they should consider it having *not* been seen in the study phase.

Four versions of each identity shown in the study phase were rated in the test phase: the prototype, the local average, the seen exemplar, and the unseen exemplar.

Just as in the study phase, the test phase was self paced, and images were presented one at a time. There was an ITI of 500 ms during which a white screen was presented on the monitor.

Results

Table 3.2 provides the means and standard deviations of confidence ratings at each level of exemplar type. The general pattern indicates that prototypes were rated similarly to the actually seen exemplars. Local averages were rated somewhat lower than seen exemplars and prototypes, but these three conditions were all rated higher than unseen exemplars. Note, however, that confidence was low in all conditions especially for prototypes, local averages, and seen exemplars (a rating of 5= zero confidence).

Exemplar Type	Mean	Std. Deviation
Seen Exemplar	5.65	1.17
Prototype	5.67	1.19
Local Average	4.97	1.36
Unseen Exemplar	3.87	1.37

Table 3.2. Mean recognition confidence rating (1=*definitely never seen before*, 9=*definitely seen before*) to face images in Experiment 7 as a function of exemplar type.

A repeated measures ANOVA conducted across the four levels of exemplar type confirmed that the type of exemplar had a significant main effect on confidence ratings, $F(3,93) = 22.30$, $MSE = 1.03$, $p < .001$. Bonferroni paired-samples t tests (on all possible pairwise comparisons) revealed that participants rated the unseen exemplars lower than all other conditions, and the local averages lower than the seen exemplars (see Table B4 in Appendix B for exact statistical values). There was also a strong trend for prototypes being rated higher than local averages, but this difference did not remain significant after the Bonferroni adjustment for multiple comparisons.

Although they will not be reported in full, additional analyses were conducted to compare confidence ratings made by participants who saw either the distal or proximal local averages in the test phase (as noted in the counterbalancing section, half the participants rated only distal, and half rated only proximal local averages). Note that since only one seen instance was studied, the distal local average refers to the local average of the opposite pair of instances. To briefly summarize the main findings, whether participants rated the distal or proximal local average, it made no difference to their confidence ratings to the seen, prototype, and unseen exemplar types. However, there was a significant difference between ratings to the distal (*M*=4.37, *SD*=1.25) and proximal (*M*=5.56, *SD*= 1.24) local averages. Also, comparing ratings among exemplar types made by participants in the distal and proximal groups separately found that distal local averages were rated significantly lower than prototype and seen exemplars but not differently to unseen exemplars. Proximal local averages, on the other hand, were rated no differently than prototype and seen exemplars, but they were rated significantly higher than unseen exemplars.

Discussion

Results of Experiment 7 showed that when prototype stimuli from Experiments 4 and 5 were reduced to the status of unseen, yet very similar, exemplars, they were not recognized more confidently than the actually seen exemplars. This result stands in contrast to those of Experiments 4 and 5 which showed highly reliable prototype effects. In fact, the current experiment found no difference among confidence ratings to seen exemplars, prototypes, or the proximal local average. Results of Experiments 4 and 5 showed a clear ordering of prototype, local average, seen exemplar, and unseen exemplar from highest to lowest mean rating. However, the results of this experiment did converge with those of Experiments 4 and 5 in that the unseen exemplars were rated substantially lower than all other exemplar types (except the distal local average). This last result is the easiest to explain because the unseen exemplar was the least similar to the seen exemplar.

It was hypothesized that if prototype items had some unknown characteristic (other than being the average of an identity) that caused participants to rate them higher than other exemplar types, then those same prototype items would be rated highest at test regardless of whether they actually were the prototype of a studied identity or not. This hypothesis was not supported since the prototype items, no longer true prototypes, were given nearly identical ratings to seen exemplars. Therefore, there is more than just mere typicality driving the prototype effects observed thus far.

On the other hand, the pattern of results did not exactly fit what one might have predicted on the basis of exemplar generalization either. In the case where only one exemplar is studied for each identity, all else being equal, recognition should be optimal when the identical exemplar is presented and should decrease as similarity to the seen exemplar decreases. Thus, the most straightforward prediction made by this account would have been that seen exemplars were rated higher than all other exemplar types. Since this was not the case, one could still assume that, under this account, participants were not sensitive to the subtle image differences among exemplar types. However, the prototype effects found in the previous experiments demonstrate that in fact participants' confidence ratings are sensitive to small image changes.

The easiest framework that accommodates the current results with the previous results is an exemplar based framework assuming parallel activation of exemplar traces. When only one exemplar of an identity is studied, the similarity of prototype stimuli is high enough to produce activation levels comparable to seen exemplars (as measured by ratings). However, when two or more exemplars of an identity are studied, the prototype, by virtue of being similar to all seen exemplars highly activates all of their traces in memory, while seen exemplars only highly activate their own trace (although they may activate traces of other seen exemplars of the same identity, but not to the same degree as their own trace, or if the prototype was the probe). Therefore, a prototype can be recognized via generalization from one similar studied exemplar, but recognition confidence increases when traces of two or more similar studied exemplars are available to support recognition of the prototype.

Chapter Discussion

The experiments presented in this chapter were designed to test predictions about the prototype effect based on prototype and episodic accounts of memory. This was accomplished by manipulating the role of context and identity information in the learning and recognition of previously unfamiliar faces. If faces of different identities are represented by prototype representations, then it was predicted that pairing exemplars with the same or different names at study would affect the strength of the prototype effect but contextual information would not. In contrast, if faces are represented in a distributed fashion across a pool of episodic traces, then it was predicted that learning inconsistent contextual information would affect the strength of the prototype effect when context was reinstated at test but pairing exemplars with the same or different names at study would not.

The tests of context and identity information were designed to be interpreted together, so that neither account would only predict null results. However, as it turned out, neither the context nor the naming manipulation had a modulating effect on the prototype effect. Thus, the overall pattern taken across Experiments 4-6 is not fully consistent with the predictions made by either the episodic account or the prototype account. Although one might consider the pattern results as partially consistent with both accounts, since *only* null effects were observed, it does not make for a persuasive argument.

Why did the observed results fail to match the predicted pattern of results of both prototype and episodic accounts of memory? The easiest explanation is that one or both of the manipulations of context or identity information was not salient enough. If participants are not required to associate the assigned context with the face, they may be very unlikely to do so. Thus, the contextual content of both stored traces and retrieval probes alike may have been insufficient to produce an observable change in the prototype effect. Likewise, perhaps participants needed more learning trials on the naming task. Maybe high levels of naming performance can only be obtained after seeing the “twins” simultaneously on the screen for comparison.

There is good reason to believe that if participants had been trained in the naming task until a certain performance criterion had been reached, results of the recognition ratings at test would have shown a decrease in the prototype effect. Stevenage (1998) reported learned categorical perception of identical twins after training participants to name a large number of different photographs of each twin. That is, she found that participants rated two images of the same twin as appearing more similar after name training compared to their initial pre-ratings of the same images before the name training. Conversely, she also found that when comparing an image of twin A with one of twin B, participants rated the similarity between the two images lower after name training compared to before name training.

The implication of Stevenage's (1998) findings is that if participants learn the twins well enough in the naming task to produce categorical perception, then both the prototype account and the episodic account of memory would predict decreased prototype effects at test (in the different name small variation condition). If the perceptual similarity of the exemplars is actually decreasing through learning, then separate prototype representations would be more likely to form. Likewise, decreased exemplar similarity would also lead to decreased similarity to a prototype probe image, so a prototype presented in the test phase would not activate the traces of the stored exemplars as highly as it would have in the absence of categorical perception. In Experiments 5 and 6, it was unlikely that learning was sufficient to cause categorical perception of the "twin-like" faces. But we were not investigating the role of categorical perception in recognizing identity prototypes. If categorical perception had occurred, prototype and episodic accounts of the experimental results would have become even more difficult to disambiguate. The influence of categorical perception on the strength of the prototype effect likely goes above and beyond influences stemming from the nature of the face representations themselves. A reorganization of a multidimensional face space to reflect category learning can account for categorical perception without making any additional assumptions regarding the prototypical or episodic nature of representations encoded in the space.

Concerns over the level of learning aside, it is very interesting that the basic prototype effect was observed consistently despite contextual and categorical influences in Experiments 4-6. Particularly in Experiments 4 and 5, the differences in images of the same identity were extremely subtle, and yet participants showed a clear increase in confidence ratings to the unseen prototypes over the actually seen exemplars.

Experiment 7 tested recognition confidence of the same stimuli used in Experiments 4 and 5 but only presented one exemplar in the study phase for each different identity. The important finding from this experiment was that the prototype (as in formerly a prototype in Experiments 4 and 5) was not rated any differently than the actually seen exemplar. Thus, the prototype images were easily confusable with the seen exemplars. But why were prototypes recognized even *more* confidently than the seen exemplars? Results of Experiment 7 indicate that the effect does depend on the prototype being the central tendency of the studied exemplars, or at least that the prototype of a series of same-face exemplars is not rated higher in confidence than the seen exemplars through some process of generalization from one of the stored exemplars.

Overall, the present chapter's results seem to suggest a more or less automatic process of image averaging that may rely primarily on similarity between one image and the next without much heed from other contextual cues or identity information. In this way, the findings from this chapter are consistent with those from Chapter 2 which found that manipulations of familiarity, intentional learning instructions, and similarity of same-face exemplars had no observable effect on the strength of the prototype effect in face recognition. The prototype effect in face recognition seems to be a robust effect that is difficult to destroy (or strengthen) using a variety of experimental manipulations and types of stimuli.

Chapter 4: Investigating Prototype Effects Using Natural Sources of Within-Identity Face Variation

Chapter Introduction

No original images of real people's faces were used in any part of the experiments in Chapter 3. Thus, the prototypes and their constituent exemplars depicted artificial identities. One cannot exclude the possibility that the variation within exemplars of the same identity was not representative of natural variation. These criticisms also apply to all of the experiments in Chapter 2 as well. Therefore, in the experiments of the current chapter, real photographs of real people's faces were used to provide more realistic within-identity variation of a person's face. The key question was: will the prototype effect be observed when the exemplars are natural, unaltered images of real people's faces? This was the original question I started out with in the first experiment of this chapter. Based on its results, follow-up experiments were conducted manipulating the degree of learning and name learning. The introduction to Experiment 8 will discuss the inspiration for this chapter in more depth.

Experiment 8

In this experiment, one photograph and one video still of each target person's face were presented in the study phase. Then, in the test phase, the photograph, the video still, the average morph of the photograph and still, and some new distractor faces were presented for recognition.

At test, the question was "did you see this person before (in the experiment)? OLD/NEW." The initial reason this task was chosen was to shift the focus of recognition away from the fine details of the images. There was a concern that the differences in image quality may make it easy to reject the prototype morphs as having been seen before in the experiment if an image recognition task was used. Asking “did you see this person before?” has higher ecological validity than the question, “did you see this exact image before?” because it mirrors the task of everyday face recognition more closely. Understanding how we are able to recognize the people we know in everyday life and why we sometimes make mistakes is the fundamental goal of face recognition research.

Using a standard OLD/NEW forced choice task would also be a rigorous test of the prototype effect in face recognition. Picture-to-picture recognition performance is known to be quite high. Therefore, one might expect participants to have a very high rate of OLD responses to the identical images shown in the test phase. To use real images of real people and find that a previously unseen average morph of the two images has an even higher hit rate than the actually seen images would be a very convincing demonstration that the average variant of a person’s face is the most recognizable one.

This experiment does not make any contrasting predictions regarding prototype and exemplar accounts of memory. However, if recognition occurs through generalization from a stored exemplar, then unseen morphs could never be better recognized than previously seen exemplars which match their stored representation near perfectly. Therefore, if the prototype effect is observed, it is good evidence that the two studied images of the same person’s face are being recognized as one. This could occur by means of an averaging process acting on the two studied images at the time of encoding, or by parallel activation of the stored representations of the two studied images at the time of retrieval.

The key question of this experiment was will the prototype effects observed previously using artificial sources of within-face variation be observed when the same-face exemplars are unaltered images of real people? It was predicted that a prototype effect would be observed because the sources of within-face variation should be more realistic when learning a face from unaltered images. Therefore, if anything, one might predict even stronger prototype effects using natural images than artificially generated ones.

Method

Participants

Forty-eight undergraduate students enrolled in an introductory course at the University of Victoria participated in this study. Each participant was tested individually in a single session lasting approximately 15 min. At the end of each session, participants were fully debriefed and awarded partial course credit.

Materials and Apparatus

<u>Stimuli.</u>

Images were taken from the U.K. Home Office Police Information Technology Organization (PITO) database. This collection of photographs and video stills of police officers was used previously by Bruce et al. (1999) in a lineup task. Male police officers posed separately for a photograph and a short video clip. For each officer, both the photograph and the video clip were taken at the same time, so their appearance in most respects was very similar. Notably, the hair styles of the officers would have been almost identical at the time of the video and the photograph. The original stimuli set contained a pair of images of each officer. One was a photograph and the other was a video still taken from the clip. I arbitrarily selected the image pairs of 63 officers avoiding any pairs that contained an image where the angle of the head was not directly forward facing or where teeth were visible. An example of a pair of images for one of the selected officers is shown in Figure 4.1.

Figure 4.1. Example of stimuli used in Experiments 8, 9, and 10. On the left is a relatively high quality photograph of the same officer shown on the right in a relatively low quality video still.

Both the photos and the video stills were full colour and showed the face and neck. Expression of the face was neutral. The main difference in the appearance of the photographs compared to the video stills was the quality of the image. The photographs (now termed “High” or “High quality”) were of higher image quality and thus appeared crisper and showed more detail of the texture of the faces compared to the video stills (which will now termed be “Low” or “Low quality”). The lower quality video stills often appeared brighter than their corresponding photographs. This sometimes made the person’s hair and skin tone look fairer in the video still compared to the photograph. There also appeared to be some slight differences between the High and Low quality images as to how the light fell on the face, and Low quality images tended to have more shadowed regions than the High quality images.

The High and Low quality images of each officer were blended, or morphed, together using the Psycho-Morph software package. All 63 "Morphed" images were the result of blending the High and Low image of 63 officers together by equal (50%/50%) proportions. Morphing was achieved by manually plotting 175 landmarks on each face (see Chapter 1 General Methods for more details).

The High, Low, and Morphed images were all cropped to remove the background around the face. Hair, ears, chin, and upper neck remained in the image. All images were sized to 275 x 375 pixels at a resolution of 72 dpi. When presented on the monitor, the face portion of the image was approximately 11 cm wide by 16 cm high. Faces were centred on a white background. See Figure 4.2 for an example of a High, Low, and Morphed image of one officer's face.

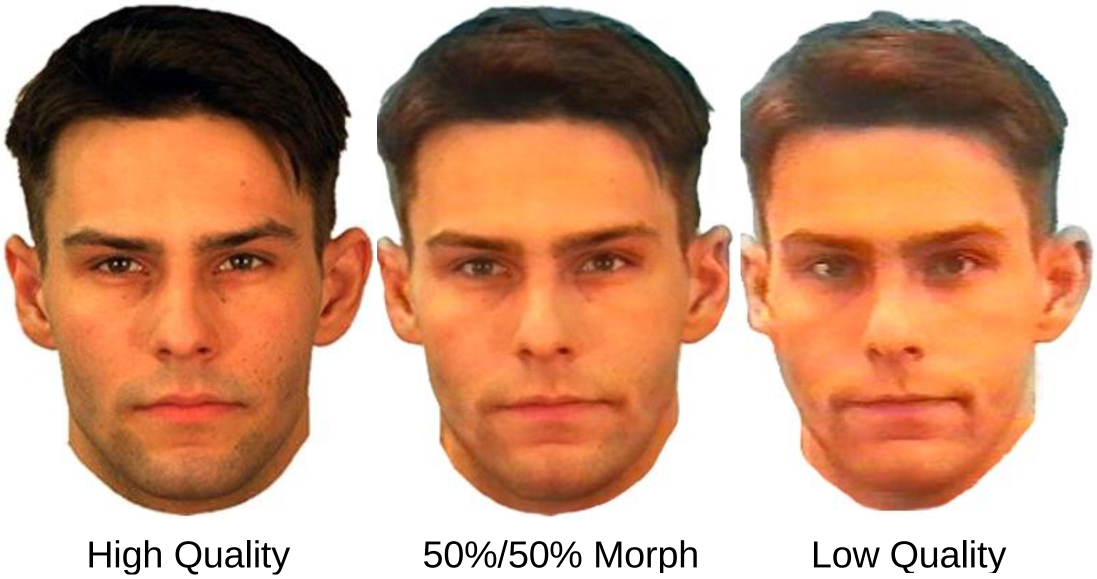


Figure 4.2. Example stimuli from Experiments 8, 9 and 10 showing the high, morph, and low quality images of the same officer's face.

Apparatus.

All aspects of the experiment were conducted using a Hi-Grade Notino3600 windows lap top with a 14” monitor at 1024 x 768 resolution. The program E-Studio from the E-Prime software package (version 1.1) was used to control the presentation of stimuli and to record responses made using the laptop keyboard.

Design

The structure of this experiment probed whether participants remembered more strongly the High, Low, or Morphed image of a person’s face having seen different versions earlier.

The independent variable manipulated in this experiment was the type of image shown in the test phase. More specifically, image type at test was manipulated within-subjects and had three levels: High quality, Low quality, and Morphed. As a reminder, High quality stimuli were photographs taken with a camera, and Low quality images were video stills captured from short video clips. Morphed images had been morphed, or blended, halfway between the High and the Low images for each officer’s face.

Recognition memory was compared across test image types to test the experimental hypothesis that if faces are stored and retrieved in such a way that enhances recognition of the average appearance of a particular person’s face, then after having studied a High and Low quality image of each officer’s face, participants will recognize the unstudied Morphed images better than the actually studied High or Low quality images It is important to note that Morphed images of target faces were never shown in the study phase.

There were two dependent variables that measured face recognition performance: accuracy and confidence. Accuracy was measured as the proportion of trials participants responded “OLD”. Recognition confidence was measured by a three choice classification that participants made after each OLD/NEW response in the test phase. The three choices were 1 (*random guess*), 2 (*pretty sure*), or 3 (*totally certain*). Confidence was recoded to become a 6 point scale by recoding 1, 2, and 3 as 3, 2, 1 respectively for trials on which a NEW response was made and recoding as 4, 5, 6 respectively for trials on which an OLD response was made. Thus, for the purpose of analysis, 1 was certain the face had not been seen, and 6 was certain the face had been seen.

Counterbalancing.

The faces assigned to be targets and distractors were counterbalanced. Since each target was presented in three different versions at test, faces were divided into four groups for the purpose of counterbalancing. One group would serve as the target list, while the other three would be High quality, Low quality, and Morphed distractor groups. The only other counterbalancing balanced the order of the High quality, the Low quality, and Morphed image types during the test phase. In the end, there were 24 counterbalancing conditions which worked out to two faces and two participants tested per counterbalancing condition.

Procedure

Participants were randomly assigned to counterbalancing conditions. The experiment was divided into separate study and test phases. Instructions were given for each phase separately immediately before that phase began. Instructions for the test phase were delivered immediately after completion of the study phase.

Study phase.

In the study phase, participants were shown a series of 30 images of faces and asked to rate the typicality of each face from 1 to 7 on a Likert type scale with 1 being *low typicality* and 7 being *high typicality*. Participants entered their responses by pushing the appropriate number on the laptop keyboard. It was mentioned to participants that although all of the images in the upcoming series were different from one another, some images may be of the same person's face. Participants were also warned to study each face closely as their recognition memory of those faces would be tested in the next phase of the experiment. Thus, learning in the study phase was intentional, but the typicality task was included to encourage participants to engage more with the faces by giving them something specific to do while they viewed the faces. No data was recorded or analyzed for this part of the experiment.

The study phase was self-paced. Images remained on the screen until participants pressed a number key on the keyboard to indicate their typicality rating for the current face. A blank white screen was then displayed for 100 ms before the next face in the series was presented.

Each participant saw 6 of the target officers' faces in the study phase (1 High + 1 Low = 2 images/officer; 2 images x 6 officers = 12 target images). A set of 12 morphed images of filler officers' faces were presented once each in the study phase. They were used as filler faces to ensure participants would not reject the Morphed images in the test phase simply because they had not seen any morphed images in the study phase. Six more fillers (3 High and 3 Low) were also added to the study set. The same set of fillers faces was used for all participants.

Test phase.

In the test phase, a new series of face images was shown. Half of these were seen in the previous study phase (targets); half had not been shown in the study phase (distractors). Participants were asked to decide, as quickly as possible, if each face, presented one at a time, in the series was OLD (this person was seen before at some point in experiment) or NEW (never been shown before in experiment). They were instructed to respond as quickly as possible by pressing either the "*Z*" or "*L*" key on the keyboard to make an OLD or NEW response respectively (response keys were swapped for half of the participants). It was emphasized that an OLD response should be given as long as the person's face was shown before even if the particular image of that face was not shown before. Participants were told that they would see the same "old" faces from the study phase more than once in the test phase, and some of those images would be identical to those shown in the study phase and some would be different images taken of the same person. The "different" images would likely differ in image quality compared to the image(s) shown of the same person earlier in the experiment. Since the same officers' faces were being shown more than once in the test phase, participants were told that even if they did not recognize the person's face the first time it was presented, they still might recognize a different picture of that person later on in the test phase. In that circumstance, it may be difficult to decide if that person had actually been presented in the original study phase. Recognition may simply be due to the presentation of that same face earlier in the test phase. Because responses were to be made as quickly as possible, participants were told to simply respond “OLD” to any face they recognized from any point earlier in the experiment regardless of whether they remembered seeing it in the study phase specifically or not.

In addition to making an OLD/NEW judgement, participants were asked to classify how confident they were that the OLD/NEW response they just gave was correct. After each OLD/NEW response was entered, participants classified their confidence as 1 (*random guess*), 2 (*pretty sure*), or 3 (*totally certain*). Participants classified their confidence aloud, and the experimenter recorded the appropriate number using a pen and paper. Participants were under no time constraints when classifying their confidence.

The format of each trial was as follows. A face image appeared on the screen. The image remained on the screen until the participant entered an OLD/NEW response on the keyboard. The key press triggered a prompt instructing participants to classify their confidence as 1, 2, or 3. Descriptions of each classification number were provided on the screen. The participant then spoke their numbered response aloud to the experimenter. Participants pressed the space bar at this point to advance to the next trial.

The test phase consisted of 78 trials. The first 6 were practice trials. The remaining 72 trials were the test trials. Practice and test trials were identical in format, but stimuli for the practice trials were taken from a separate subset of faces. All participants received the same practice trials although the order of the trials was randomized per participant. Test trials followed without interruption after the practice trials. Test trials were also presented in a random order; the only constraint due to counterbalancing the order of presentation of test image type for each target officer's face.

Of the six practice trials, 3 were OLD (1 Morphed, 1 High, and 1 Low) and 3 were NEW (1 Morphed, 1 High, and 1 Low). The three OLD faces were faces from the morphed filler set shown in the study phase. The three NEW faces were also taken from the police set, but were never used in any other part of the experiment.

Images of 6 officers were used as critical targets and 6 more as filler targets in the test trials. OLD is the correct response to a target face. Each target was shown one time in three different versions (1 Morphed, 1 Low, and 1 High) throughout the series. Thus, there were a total of 36 target test trials. In the remaining 36 trials, a face that had never been presented before was presented. On these distractor trials the correct response would be NEW. There were 12 High, 12 Low, and 12 Morphed distractors presented throughout the test series.

Results

Three repeated measures ANOVAs were conducted to probe for recognition differences due to the type of image shown in the test phase. Separate ANOVAs were conducted on mean scores of the confidence ratings, the proportion of OLD responses, and on the *d*' of the OLD/NEW responses. It is important to consider false alarm rates so that comparisons of recognition accuracy can be made. Utilizing *d*' as a measure of performance is particularly informative in this experiment because one might expect differences in old/new discrimination using stimuli differing in image quality. An alpha level of .05 was used for all statistical tests.

All test trials.

Although it was not one of the measures of recognition of primary interest, reaction time data were analyzed to compare the speed of OLD/NEW responses across the different quality test images. Mean reaction times (in ms) on correct response trials were similar for Morph, High, and Low quality target images and were, in order, 1372, 1370, and 1350 (*SD*s= 460, 399, 383 respectively). A repeated measures ANOVA on reaction times for the three image types confirmed that these differences were not significant.

Table 4.1 reports the means and standard deviations of participants' confidence ratings and the proportion of OLD responses to the Morph, High, and Low quality image types. Overall, there was no difference in recognition confidence ratings of Morph, High quality, and Low quality target images, $F(2, 94)=0.60$, $p=.55$. Likewise, when they were presented as distractors, Morph, High, and Low quality images also showed no differences in confidence ratings, $F(2,94)=0.96$, $p=.39$.

Analyses of OLD/NEW responses produced results consistent with those of confidence ratings just described. The proportion of target trials that participants responded OLD to did not differ across image types, $F(2, 94)=0.82$, $p=.45$. Distractor responses also did not differ across image types, $F(2,94)=0.30$, $p=.74$. Unsurprisingly, *d*' scores among Morphed (*M*=2.44, *SD*=1.25), High (*M*=2.53, *SD*=0.95), and Low (*M*=2.54, *SD*=1.06) quality images were not significantly different, $F(2,94)=0.15$, $p=.86$.

		Confidence Rating		Proportion OLD	
	Image Type	Mean	Std. Deviation	Mean	Std. Deviation
Target	Morph	5.26	.60	.87	.15
	High	5.30	.66	.89	.14
	Low	5.37	.53	.90	.12
Distractor	Morph	2.50	.68	.22	.17
	High	2.40	.58	.21	.14
	Low	2.55	.59	.23	.16

Table 4.1. Mean confidence rating (1= *sure unseen*, 6= *sure seen*) and proportion of old responses (with standard deviation) to targets and distractors of different image types across all test trials in Experiment 8.

In summary, analysis of recognition across all trials of the experiment did not reveal any differences among Morphed, High, and Low quality test images. However, participants' ability to recognize a given officer may have been affected by prior presentations of the same officer's face earlier in the test phase. Because analyzing data across all test trials may obscure important patterns of results found only on the very first presentation of an officer's face, the analyses were repeated using first trial data only. To see how learning during the test phase may have affected the prototype effect, the third trial data were also separately analyzed for comparison with the first trial results.

<u>First presentation only.</u>

Confidence Ratings

Analysis of first presentation data only found a trend, but no significant difference among confidence ratings of target Morphed, High, and Low image types, *F*(2,94)=2.34, *p*=.10. The trend was for High and Low image types to be rated slightly more confident than Morphed images. The distractors also did not show any differences among image types, *F*(2, 94)=0.45, *p*=.64 (See Table C1 in Appendix C for the individual condition means for both targets and distractors).

Proportion of OLD Responses

Individual condition means of the proportion of OLD responses to the first presentation of a face are listed in Table C2 in Appendix C for both targets and distractors. The repeated measures ANOVA performed on the OLD/NEW target data found no significant effect of image type, $F(2,94)=2.14$, $p=.12$. Again, there was a trend for High and Low quality images to be recognized more often than Morphed images. The ANOVA on recognition of distractor image types showed no difference among Morphed, High, and Low quality images, $F(2,94)=0.41$, $p=.66$.

d'

The repeated measures ANOVA performed on *d*' scores for the first presentation only detected a significant main effect of image type, $F(2, 94)=3.18$, $p=.046$. As shown in Figure 4.3, the pattern was that Morphed ($M=2.37$, $SD=1.85$) images were recognized less than Low ($M=3.08$, $SD=1.18$) quality images. Although present, the recognition advantage for High ($M=2.85$, $SD=1.16$) quality images over Morphed images was weaker compared to that of Low over Morphed. However, post hoc Bonferroni *t* tests did not reveal any significant differences among the three conditions.

Third presentation only.

Confidence Ratings

Individual condition means of confidence ratings to the third presentation of a face are listed in Table C1 in Appendix C for both targets and distractors. A repeated measures ANOVA of third presentation data found no significant difference among confidence ratings of target Morphed, High, and Low image types, $F(2, 94)=0.90$, $p=.41$. Analysis of distractor faces also found no significant difference among image types, $F(2,94)=0.58$, $p=.56$.

Proportion of OLD Responses

Individual condition means of the proportion of OLD responses to the third presentation of a face are listed in Table C2 in Appendix C for both targets and distractors. The repeated measures ANOVA performed on the OLD/NEW data of target images found no significant effect of image type, *F*(2,94)=0.66, *p*=.52. In comparison to the first presentation, the trend in the pattern of means was reversed with Morphed images now having a slight advantage over High and Low quality images. As in all other distractor comparisons reported from this experiment, the ANOVA yielded no differences among image types, *F*(2,94)=0.66, *p*=.52.

d'

The repeated measures ANOVA performed on *d*' scores for the third presentation did not reveal any significant differences among Morphed (*M*=3.34, *SD*=1.51), High (*M*=3.25, *SD*=1.59), and Low (*M*=3.03, *SD*=1.57) quality image types, *F*(2,94)=0.48, *p*=.62 (Figure 4.3).

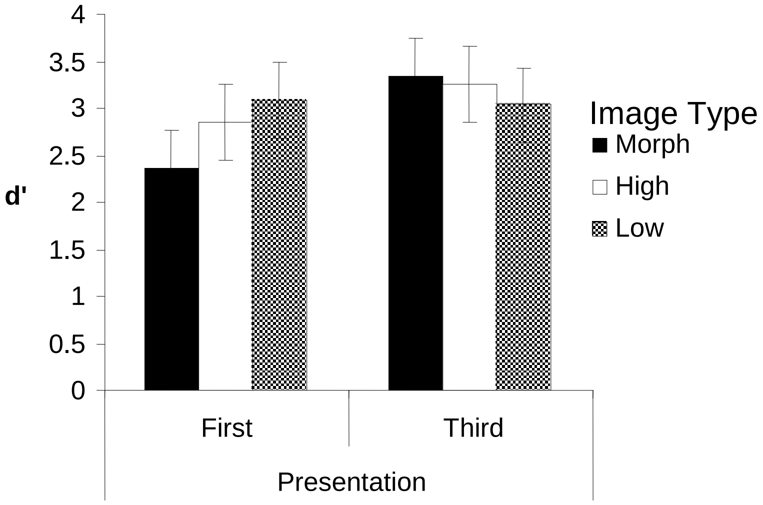


Figure 4.3. Mean discrimination scores for faces in Experiment 8 on their first or third presentation in the test phase as a function of the type of test image. Error bars represent 95% within-subject confidence intervals (Loftus & Masson, 1994).

First vs. third presentation.

Comparing the pattern of means for the confidence ratings, the proportion OLD, or the *d*' between first and third presentation of an officer's face suggested two things. The first was that recognition was higher by the third presentation compared to the first trial. There also seemed to be a tendency for participants' recognition to increase more from first to third presentation for Morphs compared to High and Low quality images. A repeated measures ANOVA was conducted to compare responses on the first and third trials by analysing both presentation and image type as independent variables. Three separate ANOVAs were carried out on the confidence ratings, the proportion of OLD responses, and the *d*' scores respectively. All three ANOVAs found a main effect of presentation. Participants rated their confidence more highly on the third presentation of an officer's face compared to the first presentation, *F*(1,47)=7.74, *MSE*=0.57, *p*=.008 Analysis of the proportion of OLD responses found the same pattern, *F*(1,47)=6.02, *MSE*=0.04, *p*=.02; *d*' scores followed the same pattern as well (see Figure 4.3), *F*(1,47)=10.55, *MSE*=1.32, *p*=.002. None of these analyses found a main effect of image type. However, the interaction between presentation and image type was significant in the analysis of confidence ratings, *F*(2,94)=3.51, *MSE*=0.66, *p*=.03. Post hoc Bonferroni paired samples *t* tests detected just one significant difference among all of the pair wise comparisons; Morphs were more confidently recognized on the third presentation of an officer's face compared to the very first, *t*(47)=3.20, *SD*=1.31, *p*=.037. The Presentation x Image Type interaction did not quite reach significance in the proportion of OLD responses and *d*' scores analyses, but the pattern of means was very similar in all three analyses.

Discussion

Contrary to what was predicted, no prototype effect was observed. The analysis of all test trials found no recognition differences among Morphed, High, and Low quality images. On the very first presentation of an officer's face, responses showed a borderline reverse prototype effect as Morphs were more poorly recognized than either the High or Low images. By the third and final presentation of an officer's face, this pattern of means had reversed with Morphs now being best recognized, but this trend towards a prototype effect did not approach significance.

Failing to observe the prototype effect in this experiment came as a great surprise. Why did participants respond so strongly to the prototype of a series of artificially constructed exemplars but not to the prototype of naturally occurring exemplars? Is the prototype effect in face recognition merely limited to laboratory experiments that use image manipulation techniques to simulate the changes in appearance that a single face undergoes from moment to moment and encounter to encounter?

It seems unlikely that natural sources of within-face variation were, for some reason, resistant to the mechanisms underlying the prototype effect in face recognition. A more likely explanation for the results of this experiment is that the differences between the two images of each person in the stimuli set were too large to give rise to a prototype effect at least after only one viewing. Since familiarity with a face can aid generalization to new exemplars of the same face (Megreya & Burton, 2007), the next experiment increased the amount of exposure participants had to the exemplars in the learning phase.

Experiment 9

In Experiment 8, the prototype effect failed to be observed using natural images of real peoples' faces. However, the particular images used did contain other sources of variation on top of what one normally considers natural within-identity face variation.

For example, variation due to the different cameras like contrast and brightness changed the appearance of the image a lot. It has already been demonstrated that participants do not always judge the two images of the same police officer as the same person even when the two images are side by side (Megreya & Burton, 2007). Megreya and Burton argued that the unfamiliarity of the participants with these faces is largely responsible for these failures in matching. In fact, they showed that after viewing a short video clip of each officer, participants were much better able to match the two views of the same officer's face.

If participants do not always judge the two images of an officer as the same person, then when seen separately, they may be unlikely to engage a prototyping mechanism at least upon the initial encounters. Familiarity with a face helps it to be recognized in unfamiliar viewing conditions and also after it has undergone novel transformations in laboratory settings. Therefore, this experiment repeated Experiment 8, but this time participants were given additional exposure to half of the faces in the study phase. By building up a stronger memory for the individual images of an officer, participants may be better able to perceive the shared identity between them, and this may change how the officer's face is represented in memory.

In the study phase, two different images of each target face were shown. As a reminder, there was a high quality and a low quality image of each face. The high and low quality images were presented once each in the lower familiarity condition and three times each in the higher familiarity condition.

The key question of Experiment 9 was if participants are given more opportunity to learn, and thus form associations, between the two images of each officer, will the prototype effect now emerge where it did not in Experiment 8?

Method

Participants

Forty-eight first and second year undergraduate Psychology students at the University of Stirling participated in this study. Data from all 48 students were collected and analysed. All other information was the same as in Experiment 8.

Materials and Apparatus

Stimuli.

The same stimuli used in Experiment 8 were used in this experiment.

Apparatus.

Same as Experiment 8

Design

This experiment was designed to investigate whether the level of preexposure to a person's face will differentially affect how strongly participants recognize it from a High, Low, or Morphed quality image. The two independent variables manipulated in this experiment were the type of image shown in the test phase and the number of repetitions of each face image during the study phase. Type of test image was manipulated just as it was in Experiment 8. Again, it had three levels: High quality, Low quality, and Morphed. In this experiment, all target faces at test had been presented in the study phase in both their High and Low image forms. Consequently, all target Morphs in the test phase will also be the average or prototype image of an officer's face. Furthermore, all target High and Low images in the test phase will have been seen before in the study phase. Remember though that targets were never presented as Morphed images in the study phase. The second IV that manipulated the number of repetitions was termed *familiarity* because varying the amount of exposure to a face should alter how familiar it is to participants. Familiarity had two levels: low (1x) and high (3x). For example, in the low familiarity condition, an officer's face was presented once in a High quality image and once in a Low quality image. Together, both IVs produced a 3 (Test Image) x 2 (Familiarity) completely within subjects factorial design.

The same two dependent variables that measured face recognition performance in Experiment 8 were also used in the current experiment. These were accuracy and confidence. Accuracy was measured as the proportion of trials participants responded OLD. Recognition confidence was measured by a three choice classification that participants made after each OLD/NEW response in the test phase. The three choices were 1 (*random guess*), 2 (*pretty sure*), or 3 (*totally certain*). Confidence was recoded to become a 6 point scale by recoding 1, 2, and 3 as 3, 2, 1 respectively for trials on which a NEW response was made and recoding as 4, 5, 6 respectively for trials on which an OLD response was made. Refer to Experiment 8 Design subsection for further details.

Counterbalancing.

Counterbalancing of stimuli across conditions of the experiment was achieved using the same stimuli rotating method as Experiment 8. Again, target and distractor faces were counterbalanced. Just like Experiment 8, the order of the test image types (High, Low, or Morphed) was counterbalanced. However, an additional counterbalancing measure was put into place to control for the effect of the order of the High and Low image of each face in the study phase. This scheme led to 48 counterbalancing conditions.

Procedure

The procedural format was very similar to that used in Experiment 8. Therefore, only aspects of the procedure that were different in the current experiment will be described.

Study phase.

Participants were shown a series of 72 images of faces and were asked to watch the sequence and to try to remember the faces. Participants were informed that some pictures would repeat, and sometimes different pictures of the same person's face would be presented. Participants were told to attend closely to each face as their recognition memory would be tested in the next phase of the experiment.

Participants did not make any responses in this phase of the experiment. Each image remained on the screen for 1200 ms followed by a blank white screen for 100 ms. Presentation of the next face in the sequence proceeded automatically after the 100 ms blank screen of the previous trial. Viewing of the entire sequence took approximately 2 min to complete.

Each participant saw 6 officers' faces in the low familiarity condition (High and Low quality image 1x each) and 6 officers' faces in the high familiarity condition (High and Low quality image 3x each). A set of 12 morphed images of filler officers' faces were presented once each in the study phase. Six of these were presented one time, and the other 6 were presented three times.

<u>Test phase.</u>

Participants entered OLD/NEW responses using the "A" and "L" keys. Confidence classifications were entered using the 1, 2, and 3 keys. After one of the three number keys was pressed, a blank white screen was presented for 600 ms. Each target was presented as a morph, a high, and a low quality image at some point during the test phase. There were an equal number of target and distractor trials of each image type. No other changes were made to the test phase from that of Experiment 8.

Results

To probe for recognition performance differences due to image type, familiarity, and order of images in the study phase, separate repeated measures ANOVAs were conducted on means of the proportion of OLD responses, *d*' scores, and confidence ratings. All statistical tests used an alpha level of .05.

Firstly, an ANOVA of the entire data set, including all three presentations of a face, was conducted. While there were no main effects of image or order, there was, as expected, a main effect of familiarity, *F*(1, 47)=88.65, *MSE*=0.06, *p*<.001. More familiar faces (*M*=.89, *SD*=.11) were called "old" more often than less familiar faces (*M*=.70, *SD*=.13).

Of the two-way interactions, only one was significant. The interaction was between the type of image shown at test and the familiarity of that face, *F*(2, 94)=5.76, *MSE*=0.03, *p*=.004. Figure 4.4 displays the pattern of means for this interaction. At lower familiarity, recognition was higher for the Morphed images compared to High and Low quality images which were recognized equally well. In contrast, after extra repetition in the study phase, High quality images were recognized better than Morphed and Low quality images which were about equal. At higher familiarity, recognition of all types of images was increased, but the increase was largest for the High quality images. Thus, the pattern of means indicated a weak trend of a prototype effect for the lower familiarity faces, but not even a hint of a prototype effect for the more familiar faces. Bonferroni post hoc paired-samples *t* tests revealed no significant differences among image types of lower familiarity faces. However, among higher familiarity faces, High quality images were recognized better than both Morphs and Low quality images (see Table C3 in Appendix C for exact statistics).

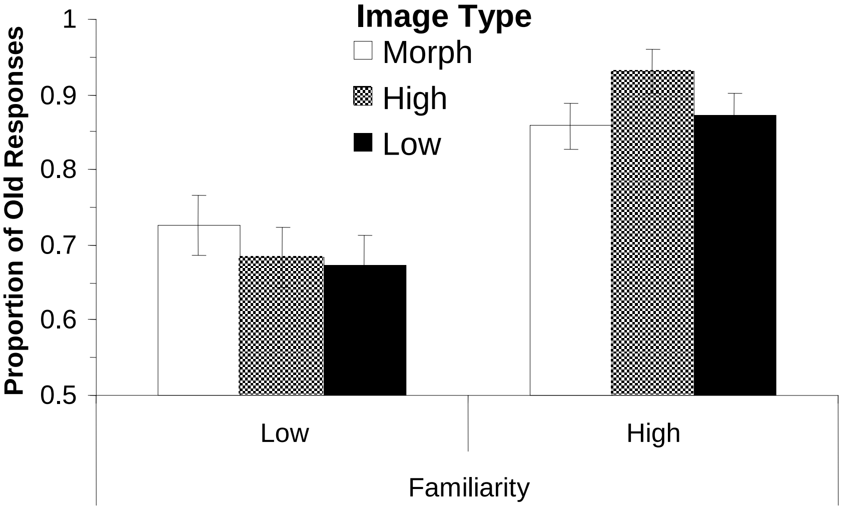


Figure 4.4. Mean proportion of old responses to target faces in Experiment 9 as a function of the type of image at test and the level of familiarity. In the low and high familiarity conditions, images were presented in the study phase once and three times respectively. Error bars represent 95% within-subject confidence intervals (Loftus & Masson, 1994).

The three-way Image x Familiarity x Order interaction just reached significance, *F*(2, 94)=3.10, *MSE*=0.04, *p*=.0496. Figure 4.5 shows the interactive pattern of condition means. For higher familiarity faces, order in the study phase did not appear to have an effect on the pattern of means of the different image types. Faces from both orders (HL and LH) were best recognized from a High quality image at test if they had been repeated three times at study. In contrast, at lower familiarity, order at study did have an impact on the effect of test image. In both orders, the Morphed image and the image that was presented first in the study phase were recognized more often than the image that was presented second at study. Recognition rates were similar for the Morphs and the first image shown in study for lower familiarity faces (although Morph was slightly higher in HL and slightly lower in the LH condition). This is an important finding because it suggests that the weak trend towards a prototype effect in the two-way interaction was actually an advantage for Morphs specifically over the second image shown in the study phase, rather than over the previously seen images in general.

Simple effects and Fisher's LSD (protected *t* tests) were calculated to tease apart the Image x Familiarity x Order three-way interaction and probe for differences among the means. When the data is broken down by familiarity, a repeated measures ANOVA confirms that order at study had no effect on recognition of higher familiarity faces, but there was an overall effect of image type, *F*(2,94)=6.39, *MSE*=0.02, *p*=.003. Indeed, Fisher's LSD revealed that High quality images were recognized significantly better than Morphed, *t*(94)=2.41, *p*=.018. High quality images were just short of being significantly better than Low quality images as well (*p*=.055). The ANOVA on the lower familiarity faces detected no recognition differences due to image type or study order on their own, but it did find a significant Image Type x Order interaction, *F*(2,94)=3.29, *MSE*=0.05, *p*=.04. At lower familiarity, the simple effect of test image was not significant when the order at study was LH, but it was significant when the order at study was HL, *F*(2,188)=3.65, *MSE*=0.05, *p*=.03. Fisher's LSD revealed that when familiarity was low and the order at study was HL, the Low quality image was recognized worse than both the Morphs, *t*(188)=2.59, *p*=.01, and the High quality image, *t*(188)=1.99, *p*=.048.

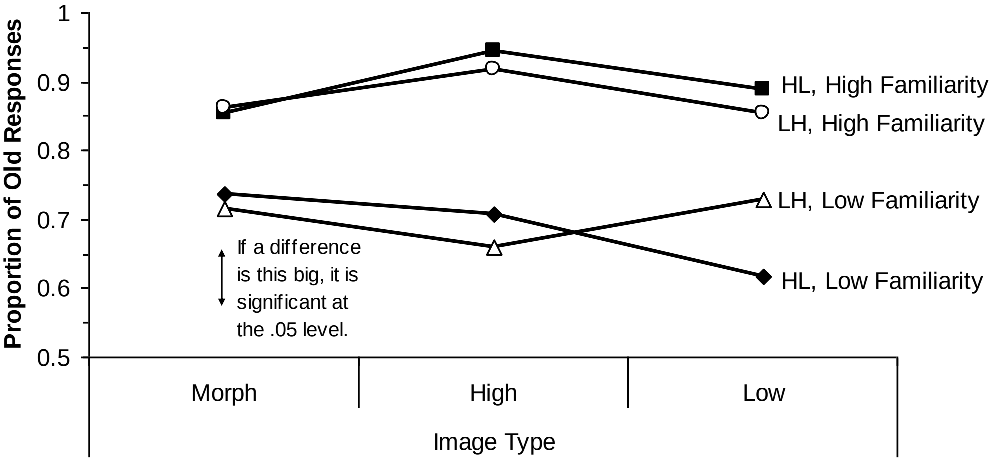


Figure 4.5. Mean proportion of old responses in Experiment 9 as a function of the type of image at test, the level of familiarity, and the order of the high and low quality images at study. HL refers to the order first the high, then the low quality image (and vice versa for LH). The single error bar represents the smallest possible significant difference between any two means and is based on the 95% within-subjects confidence interval which is appropriate for comparing the pattern of means across all conditions (Loftus & Masson, 1994).

Distractors.

A repeated measures ANOVA on the proportion of OLD responses to distractors across the three image types Morph, High, and Low detected a weak nonsignificant trend of High quality distractors being correctly rejected more than Low quality images. The mean proportion of OLD responses to Morphed distractors (*M*=.26, *SD*=.16) was just slightly higher than High (*M*=.25, *SD*=.19) and lower than Low (*M*=.30, *SD*=.17) quality distractors.

To summarize the results from the analyses of the proportion OLD data, more familiar faces were best recognized as a High quality image (Morph and Low equal) while less familiar faces were best recognized from either the Morphed or the image that was seen first in the study phase (rather than the second). Importantly, no significant prototype effects were observed. Furthermore, increasing participants' familiarity with the faces did not promote the occurrence of the prototype effect. In fact, it actually decreased the prototype effect as recognition increased more for High quality images than for Morphed images.

Subsidiary Analyses.

Additional analyses were conducted to detect prototype and familiarity effects that might have been missed by combining data across all three presentations of an officer's face in the test phase. Data from the first presentation and the third presentation were analyzed separately in two repeated measures ANOVAs that included familiarity, test image, and order in the study phase. A third ANOVA was conducted that added presentation (first or third) as a fourth independent variable. In general, the proportion of OLD responses on the first and third presentation showed a similar pattern of results to the main analyses conducted on responses to all three presentations. Participants were more likely to respond OLD to a target face on the third and final presentation that target (M=.80, SD=.14) compared to the very first time the target was presented (M=.76, SD=.13). However, this difference was not significant (p=.07). There was also a trend for the recognition of Morphs and Low quality images to improve more from the first to the third presentation compared to the High quality images which showed no improvement (see Figure 4.6).

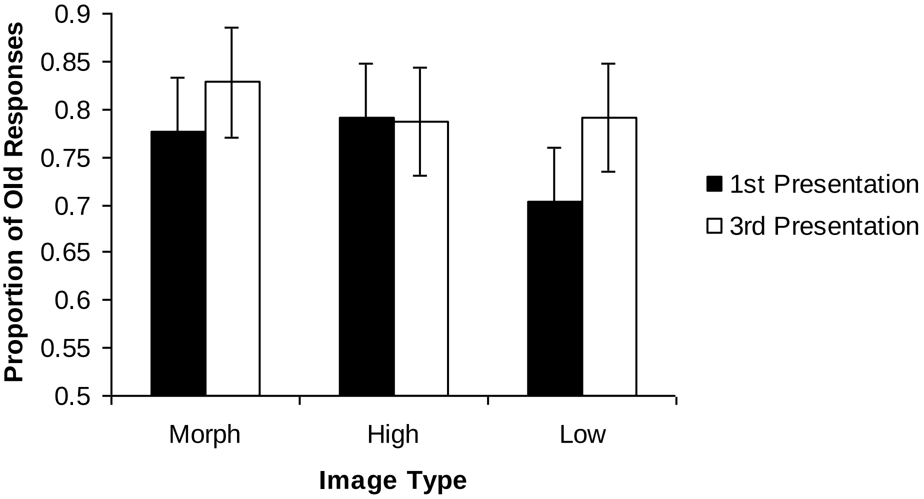


Figure 4.6. Mean proportion of old responses to target faces in Experiment 9 as a function of the type of image at test and whether it was the first or third presentation of a given person's face in the test phase. Error bars represent 95% within-subject confidence intervals (Loftus & Masson, 1994).

Distractors.

Participants were slightly more likely to correctly reject a distractor face in the third and final block of test trials (*M*=.25, *SD*=.19) than the first (*M*=.29, *SD*=.17). Although neither the main effect of trial block nor the Trial Block x Distractor Image Type interaction were significant, the trend for higher performance on distractors was driven by an improvement in correct rejection of Morph distractors compared to High or Low quality distractors.

Measures of recognition accuracy and recognition confidence.

To simply summarize the analyses of *d*' scores and confidence ratings, exact statistics will only be reported if there was a change in the significance or the pattern of means compared to the corresponding effect in the proportion OLD analysis. Likewise, a description of the pattern of means will only be given for effects that showed a different pattern in the proportion OLD analysis. The analyses of *d*'scores and confidence ratings were also broken down and analyzed in the same manner as in the proportion OLD analyses to probe for differences in recognition accuracy between the first and third presentation of an officer's face.

d' all presentations.

The repeated measures ANOVA reported for the proportion of OLD responses was repeated on the *d*' scores to measure recognition performance taking false alarm rates into account.

In addition to the main effect of familiarity, the main *d*' analysis also found a significant main effect of test image, *F*(2,94)=3.59, *MSE*=2.00, *p*=.03. According to Bonferroni paired samples *t* tests, High quality images (*M*=2.18, *SD*=0.88) were significantly better recognized than Low quality images (*M*=1.80, *SD*=0.84), *t*(47)=2.63, *SD*=1.01, *p*=.04. The mean *d*' value of Morphed images (*M*=2.04, *SD*=0.89) fell midway between, but was not statistically different from, those of High and Low quality images.

As in the proportion OLD analysis, the two-way Test Image x Familiarity interaction was significant. In addition, a trend of the two-way Test Image x Order interaction was detected, *F*(2,94)=2.49, *p*=.08. The trend was for High quality images to do a lot better, Low quality images to do a lot worse, and Morphed images to do about equally well in the study order HL compared to LH. The trend was not for a crossover interaction though. Regardless of study order, performance, from highest to lowest, was High, then Morphed, and then Low quality images. Thus, the pattern of the overall interaction trend showed that differences among test image types were much stronger in the HL study order compared to LH order. Lastly, the three-way Test Image x Familiarity x Order interaction did not quite reach significance in this experiment, *F*(2,94)=2.86, *p*=.06. Even still, the pattern of means in this interaction stayed reasonably the same as in the proportion OLD analysis.

d': first and third presentations.

The repeated measures ANOVA including presentation (first or third) as a fourth independent variable found a significant increase in recognition accuracy by the third presentation (*M*=2.43, *SD*=0.98) in the test phase compared to the first (*M*=2.06, *SD*=1.03), *F*(1,47)=4.80, *MSE*=8.29, *p*=.03. In all other respects, the results were comparable to those of the proportion OLD analyses of first and third presentations.

Confidence ratings.

Analysis of confidence ratings showed a similar pattern of results as the proportion OLD and the *d*' analyses. The main effect of familiarity was significant as was the two-way Familiarity x Test Image interaction, and the three-way Familiarity x Test Image x Order interaction. All of these effects showed the same pattern of means as the proportion OLD analysis, so they will not be discussed further. No other main effect or interactions were significant.

Confidence ratings: first and third presentations.

The repeated measures ANOVA on confidence ratings including presentation (first or third) as a fourth independent variable found the same pattern of means as the analyses of proportion OLD did on the first and third presentations did. Therefore, these results will not be reported.

Confidence ratings: distractors

A repeated measures ANOVA on confidence ratings to distractors across the three image types Morph, High, and Low revealed a nonsignificant trend of High quality distractors being more confidently rejected than Low quality images. Mean confidence ratings to Morphed distractors (*M*=2.62, *SD*=0.63) fell midway between High (*M*=2.57, *SD*=0.76) and Low (*M*=2.78, *SD*=0.69) quality distractors.

Discussion

Although the pattern of results was quite complex, some general conclusions can be made. Increasing the number of repetitions in the learning phase did indeed increase face recognition accuracy in the test phase. However, there was very limited support for the hypothesis that additional learning will lead to stronger prototype effects. In fact overall, the results suggested the opposite. While less familiar faces were best recognized from either the Morph or the image that was seen first in the study phase (rather than the second), more familiar faces were best recognized as a High quality image (Morph and Low equal).

Comparing patterns between the first and third presentation data, some small differences were found. As in Experiment 8, there was a slight morph advantage on the third but not the first presentation. It looks as though this is partially due to the Morph and Low quality images improving more across the three presentations compared to the High quality especially for the lower familiarity faces. Although all of these differences were nonsignificant, they were the only results that did support the hypothesis that additional learning would lead to increased prototype effects. The surprise twist here was that the only trend toward a strengthening of the prototype effect due to increasing the degree of learning was for learning during the test phase rather than the study phase. This will be discussed further in the discussion section at the end of this chapter.

Seeing the exemplar pairs more times in the learning phase turned out to benefit the recognition of High quality images more than the Low quality or Morphed images. It is possible that the High quality images contained more facial information relevant to face processing and discrimination than Low quality images did; and therefore, a High quality image at test may contain more retrieval cues than a Low quality image at test.

Also note that there was a trend towards a prototype effect for lower familiarity faces. However, further analysis showed that the advantage for Morphed images was limited to when they were compared with the second image presented in the initial passive view phase of the experiment but not the first image. The Morph and the first image shown in the study phase were recognized approximately equally as well. Why the ordering of image types at study had this effect on recognition remains unclear. However, the pattern does bring to mind the phrase "you never get a second chance to make a first impression." Perhaps the memory formed the first time you encounter a person's face contributes more than the second to either the stored prototype representation of that face or to the retrieved content activated upon subsequent encounters with that face. Of course, this would imply that if different face identities are stored as composite or summary representations, they may not necessarily correspond to the central tendency of the variation experienced for a given identity.

To sum up, no prototype effects were found regardless of how many repetitions of the exemplars participants saw during the study phase. The familiarity manipulation not only failed to induce a prototype effect, but it also increased the recognition of High quality images at test (after additional repetitions in the study phase). However, there was a weak trend for a prototype advantage by the third and final presentation of a face in the test phase compared to the first presentation.

Experiment 10

In Experiment 9, participants were given more exposure to the faces in the learning phase to help them associate the two studied exemplars of each face as belonging to the same person. It was hoped that additional familiarity with the exemplars might produce the prototype effect that was absent from the results of Experiment 8. However, the familiarity manipulation did not produce a prototype effect. In fact, if anything, it weakened it.

Perhaps if participants are not encouraged to form an association of identity between the two images, they will not do so. Arguably, the photograph and the video still of each officer are not obviously of the same individual. If the learning phase does not have an identity related component to it, then a connection between the two images may be unlikely to form. It is also possible that, without influences to the contrary, participants might form even more distinct representations of each of the two images after viewing additional repetitions.

Therefore, in this experiment, participants were provided with explicit cues about the identity of each face. This was achieved using the same naming task as in Experiments 5 and 6 (Chapter 3). Although manipulating whether the exemplars of each face were learned with one name or two different names had no effect on the strength of the prototype effect in Experiments 5 and 6, there were several reasons why it was worth trying the naming task again with the stimuli and recognition task used in Experiments 8 and 9. One reason is that in the experiments in Chapter 3, the prototype effect was observed, but there was a failure to alter the magnitude of the effect. In the two previous experiments in this chapter, there was a failure to observe even the basic prototype effect. Thus, the naming manipulation may affect the pattern of results differently in the present case. Also, the task at test in this experiment was to recognize the person's face, not merely the specific image of the face. Therefore, manipulating identity related information during learning in the study phase may have a different effect in a task requiring judgements specific to the identity rather than specific to a particular image. At any rate, the stimuli lent themselves well to a naming experiment since the stimuli can be learned as same or different people quite conceivably.

The key question of this experiment was will the prototype effect now emerge if participants learn to associate the two images of each face with the same name but not with two different names?

Method

Participants

Forty-eight first and second year undergraduate students enrolled in the Psychology programme at the University of Stirling participated in this study. All details regarding participation are the same as in Experiment 9.

Materials

Stimuli.

Images were the same as in Experiment 8. Male first names were chosen arbitrarily by the experimenter, but an effort was made to choose names that started with different letters of the alphabet.

Apparatus.

Same as Experiment 8.

Design

There were two independent variables in this experiment. The first was the type of image shown in the test phase, termed image type, and it had three levels: High quality, Low quality, and Morphed (see Experiment 8 for details about these terms). The second IV, termed name, was whether the two images of each officer had been learned with the same name or with two different names in the learning phases of the experiment. Thus, the name IV had two levels: same name or different name. The end result was a 3 (Test Image) x 2 (Name) completely within subjects factorial design.

Two dependent variables were used to measure face recognition in the test phase. The first DV measured the proportion of OLD responses in the Old/New forced choice task, and this data was also converted into *d*' scores to measure how well participants could discriminate between targets and distractors. The second DV was recognition confidence and was measured using a confidence classification that was made about each Old/New response. The confidence scale was from 1-3 (but was recoded as a 1-6 scale for the purpose of data analysis as described in Experiment 9). In fact, the design and procedure of the test phase of this experiment was exactly the same as the test phase in Experiment 9 except that participants were not asked to make their Old/New responses as quickly as possible, and the reaction time data were not analyzed.

Counterbalancing.

Names were counterbalanced such that each participant saw an equal number of officers with the same name as with two different names. Furthermore, across all participants, each officer was shown an equal number of times with the same name as with two different names. Names were put into pairs and assigned to a target face in each target set. The name in the pair used as the same name was reversed for half the participants.

The only other counterbalancing feature in the learning phases was the order of the High and Low quality images for each officer presented in the initial viewing phase. The learning trials were put into two blocks to achieve this. The High and Low images of each officer were put into separate blocks, and there were equal numbers of High and Low images within each block. In phase two, the guess the name task, the block order was randomized for each participant.

Targets and distractors were counterbalanced by dividing the total critical items set of 48 officers into 4 target sets of 12 officers each and rotating these sets around conditions. To be clear, the 4 conditions (A-D) were A: target, B: morphed distractor, C: high quality distractor, and D: low quality distractor. Four sets were needed because each target was presented three times at test, thus necessitating 3 times as many distractors to balance the number targets without repeating any of the distractors.

Because all three types of images of each officer were presented once each at some point throughout the test phase, the order of the three image types (Morph, High, and Low) was counterbalanced. Test trials were organized into three blocks and stimuli were rotated around the blocks. This meant that, across participants, the Morph, High, and Low images of each officer were each shown an equal number of times in the first block, the second block, and the third block.

Procedure

The experiment was divided into three parts. The first two were learning phases and the third and final part of the experiment was the recognition test phase. In all three parts, participants had unlimited time to make their responses, and images remained on the screen until a response key was pressed. Faces were always presented one at a time and positioned approximately in the centre of the display. Trial order was randomized for each participant within the constraints of any blocking done for counterbalancing purposes.

1. Face and name viewing.

In this phase, participants passively viewed images of faces with their associated name printed below it. All 12 target officers were shown in this phase, and each officer was shown in both a High and Low quality image. Half of the officers were given the same name across both images, and half of the officers were given two different names. In addition to the critical faces, there were several filler faces of morphed quality (50/50% of High and Low images). Their purpose was to provide participants with some exposure to morphed images, so that the Morphs in the test phase would not stand out as being obviously different in quality from the images seen in the learning phases. There were 9 different officers' faces used as fillers. Six were all given different names. The other 3 were presented twice each and given the same name in both occurrences.

2. Multiple choice name learning.

In this phase, participants were presented with a series of faces each with three names printed below. One name was the correct name (the name shown with that image in the previous phase) and the other two were names given to the images of other officers in the set. Participants were asked to select the correct name by pressing the key on the keyboard that corresponded to the initial letter of that name. The three names always started with different letters. Pressing a letter key on the keyboard triggered a feedback message indicating if the response was correct or incorrect. If the response was incorrect, the feedback also displayed the correct name. All of the same items from the passive viewing phase were shown in this phase.

3. Recognition test: Old/New face + confidence.

In the test phase, half the items in the series were OLD (target faces) and half were NEW (distractor faces). Participants were instructed to respond OLD if the face had been shown earlier in the experiment or NEW if the face was being shown for the very first time. Participants were asked to base their judgements according to whether they thought the same *person* had been presented earlier rather than the exact same image. Therefore, it would be correct to respond OLD to an image that had never been seen before as long as it was judged to belong to one of the people shown earlier. The "z" and "/" keys were pressed on the keyboard to indicate an OLD or NEW response respectively (response keys reversed for half of the participants). Pressing a response key triggered a prompt screen which asked for a confidence rating of 1, 2, or 3 with the following descriptions: 1. *random guess*, 2. *pretty sure*, and 3. *very certain*. Thus, participants made an Old/New decision followed by a confidence rating on that decision for each image in the test series.

The test phase began with six practice trials. Half the faces in the practice trials were new faces and half were seen before as filler faces in the learning phases. At some point during the test phase proper, all 12 target faces were shown in three different image types (Morph, High, and Low). Distractor faces were also of three different image types and in equal numbers to match the number of image types of target faces.

Results

Study Phase

Participants were more accurate at selecting the correct name in the study phase when the High and Low quality images of the same officer's face were seen initially with the same name (73% correct) compared with two different names (61% correct). Chance performance on this task was 33%.

Test Phase

Figure 4.7 shows the mean pattern of *d*' scores in each of the individual conditions. Separate repeated measures ANOVAs were conducted on participant means of the proportion of OLD responses, the *d*' scores on the OLD/NEW responses, and the confidence ratings. The ANOVAs probed for recognition differences due to whether images of the same officer were learned with the same name or two different names and the type of image shown in the test phase.

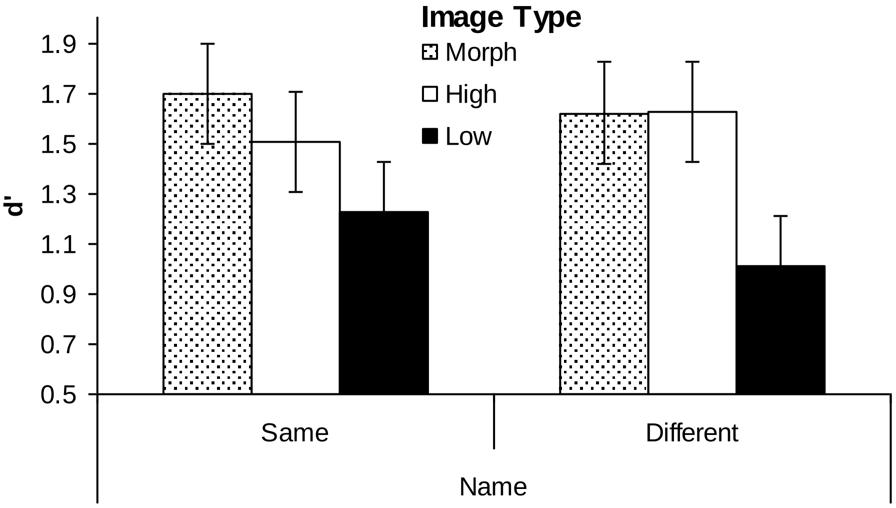


Figure 4.7. *d*' in Experiment 10 as a function of image type and whether the face was learned with one name or two different names in the study phase. Error bars represent 95% within-subject confidence intervals.

The analysis of the proportion of OLD responses did not reveal any significant differences among condition means. Both the main effect of name and test image were nonsignificant (both *p*'s > .7). However, the Test Image x Names interaction did show a weak trend, *p*=.12. The trend was for Low quality images to be the least recognized of the three image types for faces that were studied with two different names. Also, there was a trend for Low quality images to be recognized more often when they belonged to faces studied with the same name compared to two different names.

Distractors of differing image quality were not rejected equally as well however, *F*(2,94)=8.40, *MSE*=0.02, *p*<.001. Low (*M*=.36, *SD*=.19) quality distractors were falsely recognized more often than either the Highs (*M*=.27, *SD*=.15) or the Morphs (*M*=.25, *SD*=.17) which did not differ. Post hoc *t* test using Bonferroni adjustment for multiple comparisons confirmed this pattern (see Table C4 in Appendix C for exact statistics).

The *d*’ analysis did find a main effect of image type, *F*(2,94)=8.01, *MSE*=1.02, *p*=.001. Low (*M*=1.12, *SD*=0.83) quality images were recognized less accurately than Morphs (*M*=1.66, *SD*=1.02) or Highs (*M*=1.57, *SD*=1.01) which were equal. Bonferroni post hoc *t* tests confirmed this pattern (see Table C5 in Appendix C for exact statistics). The main effect of names showed no effect, and the trend in the Name x Test Image interaction weakened, *p*=.21 (relative to the same interaction in the proportion old analysis).

The pattern of results looks quite similar in the data using recognition confidence as the dependent variable. Morph and High quality images seemed to be recognized more confidently for faces previously studied with two different names compared to the same name. In contrast, Low quality images still seemed to show some advantage in the same name condition compared to the different name condition. However, the ANOVA failed to detect any significant differences.

Subsidiary analyses

First presentation.

Analyzing the first presentation only data in a separate repeated measures ANOVA of name and test image on the proportion of OLD responses also yielded no significant differences among condition means. There were nonsignificant trends in both of the main effects. The trends were as follows: 1. Faces studied with the same name were recognized more than those studied with different names. 2: High quality images were recognized more than Morphs and Low quality images which were similar.

Repeating the analysis of first presentation data using *d*' as the dependent variable did find a significant main effect of image type, *F*(2, 94)= 3.78, *MSE*=3.98, *p*=.03. Post hoc Bonferroni *t* tests revealed that High quality images were better recognized than Low quality images, but no other comparisons of High, Low, and Morphed images were significant (see Table C6 in Appendix C for exact statistics).

Repeating the analysis of first presentation data using recognition confidence as the dependent variable did find a significant main effect of image type, *F*(2, 94)= 3.44, *MSE*=1.39, *p*=.04. However, post hoc paired-samples *t* tests failed to detect any significant differences among test image types after a Bonferroni adjustment for multiple comparisons was made, but the trend was for High quality images to be recognized more confidently than Low quality images.

One thing to note is that on the first presentation of an officer's face the Name x Test Image interaction in the ANOVAs on the proportion OLD, *d*', or confidence ratings did not approach significance (all *p*'s>.4). In fact, the pattern of condition means was qualitatively different on the first presentation compared to the pattern produced by collapsing across all presentations. Learning an officer's face with the same name led to increased recognition of the Morph and Low quality images but had no effect on High quality images.

Third presentation.

Analyzing the third presentation only data in three separate repeated measures ANOVA of name and test image on the proportion of OLD, *d*', and confidence ratings respectively found only one significant effect. This was a main effect of test image and was found in the *d*' analysis, *F*(2, 94)= 4.50, *MSE*=4.37, *p*=.01. Post hoc Bonferroni *t* tests indicated that Morphs were better discriminated from distractors than Low quality images were, *t*(47)=2.95, *SD*=1.95, *p*=.015. Morphs were better discriminated than High quality images were, but this difference just missed significance, *p*=.051. The qualitative pattern of test image means was fairly consistent across all three dependent variables with Morphs being better recognized than High and Low quality images which were equal. Names had no effect on recognition on its own nor did it have an interactive effect with test image. However, a weak trend for better recognition of High quality images in the different name condition compared to the same name condition was observed across all three analyses of third presentation data. Figure 4.8 summarizes the effect of image type on recognition on the first and third presentations using the proportion of old responses as an example.

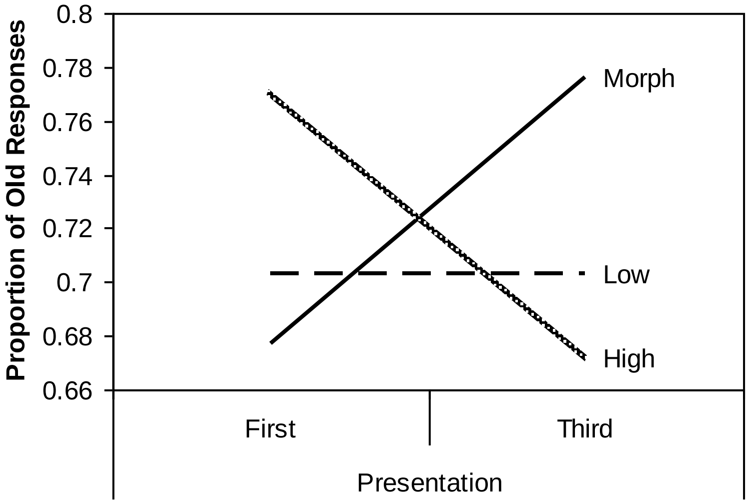


Figure 4.8. Mean proportion of old responses in Experiment 10 on the first and third presentation of a face in the test phase as a function of the type of test image.

<u>First vs. third presentation.</u>

Three more repeated measures ANOVAs were conducted to probe for recognition differences between the first and third presentations of an officer's face in the test phase. The three ANOVAs analyzed proportion old responses, *d*', and confidence ratings in turn. Each of these ANOVAs analyzed the effects of presentation, test image, and names on recognition responses. A fourth repeated measures ANOVA was conducted on the proportion of old responses to distractor images to compare false alarm rates in the first block of test trials to the third block. All post hoc comparisons were made using paired samples *t* tests with a Bonferroni adjustment for multiple comparisons.

Figure 4.9 shows the change in the proportion of old responses from the first to third presentation in the test phase across all levels of name and image type. Old responses to targets showed no overall improvement whatsoever from first to third presentation. There was also no main effect of test image. However, the influence of these variables was revealed by a significant interaction between presentation and test image, *F*(2, 94)=3.98, *MSE*=.12 , *p*=.02. Comparing third presentation responses to first presentation, Morphs were recognized more on the third, Highs less on the third, and Lows showed no change. Post hoc testing confirmed that the effect of presentation differed between Morphs and High quality images, *t*(47)= 2.85, *SD*=.48, *p*=.019. There was also a nonsignificant trend for an interaction between names and presentation, *p*=.06. Faces learned with the same name were better recognized on the first presentation at test relative to faces learned with two different names, but this trend disappeared and even slightly reversed on the third presentation at test. Results from the ANOVA on confidence ratings were entirely consistent with those of the ANOVA on the proportion of old responses just described, and thus, to avoid redundancy, they will not be reported in full.

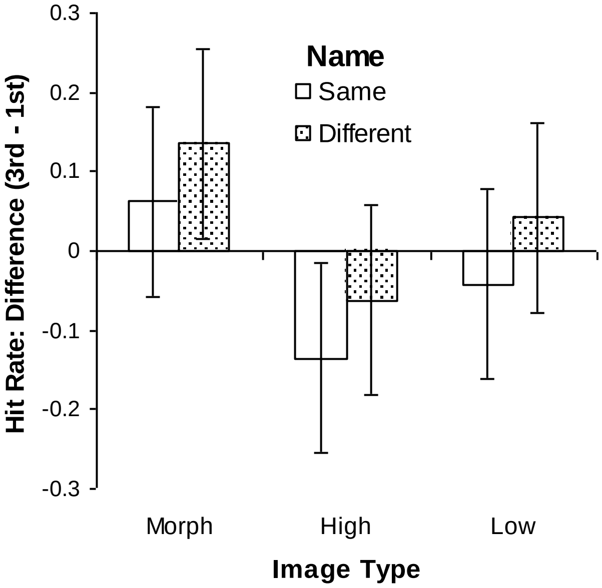


Figure 4.9. Mean difference between the proportion of old responses in the first and third presentations (3rd – 1st) of a face in the test phase of Experiment 10 as a function of image type and whether the face was learned with one name or two different names in the study phase. Error bars represent 95% within-subject confidence intervals.

The *d*' ANOVA indicated a nonsignificant trend towards better overall discrimination in the third block of trials than the first block. It also detected a significant main effect of test image, *F*(2, 94)= 5.05, *MSE*= 3.62., *p*=.01. Post hoc tests found that this difference was due to Morphs being recognized better than Low quality images, *t*(47)= 3.28, *SD*= 1.25, *p*=.006. Again there was a significant interaction between test image and presentation, *F*(2, 94)= 3.47, *MSE*= 4.73, *p*=.04. The pattern was that Morphs and Low quality images showed an improvement on the third presentation relative to the first, but High quality images showed a small decline. Although post hoc testing revealed no significant differences, the trend of Morphs improving more than High quality images bordered on significance, *p*=.057. The *d*' analysis will not differ from the proportion of old responses for any effects involving name, so the Names x Presentation trend was observed as described earlier.

Considering the *d*’ and proportion old analyses together, it seems that a reduction in the false alarm rate was at least partly responsible for improvements in discrimination observed in the data from the third presentation. An ANOVA on the proportion of old responses to distractors indicated that false alarm rates were affected by both the quality of the distractor image and whether the distractor was presented in the first or third block of test trials. The main effect of image type was significant, *F*(2, 94)= 6.73, *MSE*=.06, *p*<.01. Post hoc testing revealed that Morphed distractors were judged falsely as old less often than Low quality distractors, *t*(47)= 3.40, *SD*=.25, *p*=.004; High quality distractors were also judged as old less often than Low quality distractors, *t*(47)= 2.49, *SD*=.27, *p*=.049; but Morph and High quality distractors showed no difference. There was also a significant overall decline in the amount of old responses given to distractors in the third block of test trials compared to the first block, *F*(1, 47)= 4.95, *MSE*=.05, *p*=.03. The interaction between trial block and image type was not significant. However, inspection of the pattern of conditions means suggests that only Morph and Low quality distractors showed a notable reduction in false alarms from the first to third block of test trials (5.73%, 0.52%, and 10.43% reductions for Morph, High and Low quality distractors respectively).

In summary, recognition performance tended to be higher on trials near the end of the test phase compared to the beginning. This was demonstrated most clearly by the morphed images which received higher hit rates and confidence ratings to targets and a reduced false alarm rate to distractors. Improvements on Low quality images were mainly due to a reduction in false alarms to distractors rather than improved recognition of targets. In contrast, performance actually decreased for High quality images, and this was mainly due to decreased hits and confidence ratings to targets with false alarms showing almost no change between trials in the first and third blocks of the test phase.

Order of images in study phase.

A separate repeated measures ANOVA was conducted to check if the order of the High and Low images in the initial passive viewing phase had any impact on Old/New judgements. Order did not have any significant effect on the first trial at test or throughout the test phase (all *p*'s >.05).

Discussion

Consistent with the results of Experiments 8 and 9, Experiment 10 also did not produce a clear overall prototype effect. Even when participants learned to associate both the High and Low image of each officer's face with the same name, they still were no better at recognizing the average Morph compared to the individual exemplars they learned in the study phase.

The naming manipulation did not have any clear influence on recognition in the test phase. Overall, a face learned with the same name was about as well recognized as a face learned with two different names. There was a trend, however, for Low quality images to be better recognized after being learned with the same name in the study phase compared to all other conditions. The implication here might be that if recognition can be achieved from a close match to stored representations of the High quality image, additional activation from representations of the Low quality image would make little difference to the recognition judgement. In contrast, if the Low quality image is more difficult to recognize, additional activation from stored representations of the High quality image may sometimes make the difference between a recognition failure and a correct response. This recognition benefit for Low quality images learned with the same name could be offset by a recognition benefit for High quality images in the different name condition which might promote more attention during the learning phase. Since High quality images benefited the most from additional exposures in the study phase in Experiment 9, it is conceivable that the increased attention in the different name condition due to the greater difficulty of the task compared to the same name condition could have led to a larger recognition benefit for High quality images at test compared to Morphs and Low quality images. The end result would be no overall effect of the naming manipulation, but an interaction between naming condition and the type of image shown at test. This is essentially what was found except the interaction was only observed as a trend in the data.

Also of interest was the finding that although there was no sign of a prototype effect on the first presentation of an officer's face (in fact there was a trend in the reverse direction), on the third presentation, the prototype Morphs tended to be better recognized than the High and Low quality seen exemplars. This pattern of results was found in Experiment 8 and also as a trend in Experiment 9. Compared to performance on the first presentation, Morphs were better recognized on the third presentation, but recognition of High quality images seemed to actually get worse. Participants were not much better at recognizing Low quality targets on the third presentation, but they did get better at rejecting Low quality distractors in the final block of test trials. The effect of presenting an image first or last in the test phase will be discussed further in the chapter discussion.

Chapter Discussion

In light of the previous two chapters, the experimental findings presented in this chapter came as quite a surprise. In all three experiments, there was a failure to demonstrate the basic prototype effect.

The approach to studying natural sources of within-face variation in this chapter set up an interesting situation. It is already known that the prototype effect is quite reliable from the previous experiments and from the literature (Bruce et al., 1991; Busey & Tunnicliff, 1999; Cabeza et al., 1999), so there was good reason to expect it would be found in this situation too. On the other hand, in the line-up task used by Bruce et al. (1999), participants were rather poor at matching the video still and the photograph of the same person together (reminder: Bruce et al. used the same set of stimuli as was used in Experiments 8-10). This implies that the photo and the still are at least not obviously judged to be the same person by an unfamiliar observer. This raises the question of whether or not a prototype mechanism *should* be operating in this situation if there is no judgement of shared identity between the still and the photo and no other type of top-down information to indicate such a relationship.

In Experiment 8, I aimed simply to demonstrate the basic prototype effect using natural variants of the same person’s face. When this failed, I had to consider that to an unfamiliar viewer, exemplars of the same identity may not have been similar enough to produce the prototype effect. If part of what is important to obtaining the prototype effect is the perception of shared identity among exemplars, then trying to increase participants’ likelihood of perceiving the high and low quality images as the same person’s face was the next logical step for follow-up experiments to take. Attempts were made using two different methods. The first method was to make participants more familiar with the faces. The second method was to teach participants to associate a name with each face.

In Experiment 9, additional exposure to half of the studied faces in the learning phase was provided. The rationale was that participants would be more likely to perceive the shared identity of the high and low quality images if they were exposed to them a greater number of times. Once participants were made slightly more familiar with a face, the prototype effect might then emerge. However, Experiment 9 provided no evidence whatsoever in support of this hypothesis. Additional familiarity with the images presented in the learning phase, if anything, tended to reverse the standard pattern of the prototype effect. This result led me to consider that if the two studied exemplars were not initially perceived as the same individual, then merely presenting the images a greater number of times may do little to change this. In fact, there is evidence from the categorization literature to support this. In the absence of category feedback in the learning phase, participants cannot accurately identify the number of categories in the training set, and have difficulty learning categories except for categories that are composed of minimal distortions (Homa & Cultice, 1984). Homa et al. (2001) further showed that prior observational learning does not facilitate category learning at least when the categories are not readily discernable from the start. Therefore, unless participants receive some feedback about which images depict the same person, simply increasing the number of exposures may not help participants learn the different identities.

In Experiment 10, instead of just repeating the same images more times, I tried to explicitly teach participants that the high and low quality image were the same person’s face. I did this by using a face-name association task similar to the one used in Experiments 6 and 7. For half of the faces, the high and low quality images both were given the same name. Images of the other half of faces were learned with two different names so that I could test whether the naming manipulation did have an effect on recognition later in the test phase. To be clear, if the prototype effect was not observed in Experiments 8 and 9 because the high and low quality images of the same face were not perceived as the same identity, then learning to associate the images with the same named identity might lead the prototype morph to be better recognized than the actually seen images at test. However, the obtained pattern of results from this experiment did not support that hypothesis. Whether the images of the same person’s face were learned with one name or two, recognition of the morph, in relation to the seen exemplars, remained approximately constant. One possible explanation for the lack of naming effect is that participants learned the exemplar-name pairs in relative isolation from one another. Exemplar A might have been successfully associated with the name Bill and Exemplar B with the name Bill as well, but this did not cause Exemplars A and B to form an association with each other via the shared association with the name Bill. Similarly, one could know two different people named Bill and have no trouble keeping them straight despite their mutual association with the name Bill.

There was one more important finding from these experiments which was unanticipated. Although only a trend in Experiments 9 and 10, Experiment 8 found that the prototype morph was recognized more confidently when it was the third image of a person’s face to be presented at test rather than the first image (each face was presented as a high, low, and morphed image in some order within the test series).

While recognition of all image types showed a general improvement on the third presentation (except for recognition of high quality target images in Experiment 10 which tended to decrease), the improvement was most dramatic for morphed images. This finding was quite puzzling. Why did additional learning in the study phase not show an increase in the prototype effect (perhaps even a decrease) but learning from earlier trials in the test phase did show at least a very strong trend in that direction? I cannot offer a good explanation from these data alone, but learning in the study phase and the test phase likely had different effects on recognition. However, there was a small confound in the experimental design. To be clear, whether a morph, high, or low quality image was shown on the third trial was systematically related to which image types had been presented on the previous two test trials. For example, if a morph was presented on the third trial, then a high and a low image were presented on the first two test trials in some order. In contrast, if a high image was shown on the third test trial, then a morph and a low image were presented on the first two trials. Therefore, improvement on the morph when tested on the third trial relative to the first, could have related to having seen the high quality image earlier in the test phase. Participants might simply learn a face better from the high quality image, or the pattern of results reflects a more complex interplay between the current test image and the previously seen test images.

There is one more factor that needs to be considered in the interpretation of findings from this chapter. That is that Experiments 8-10 all used a slightly different task at test than the experiments in the previous chapters did. The task was to judge whether the test image was one of the same *people* shown earlier rather than the same exact *image*. Cabeza et al. (1999) found that these two slightly different tasks both produced the same pattern of results (i.e. the prototype effect). This finding was one of the main reasons it was anticipated that either task would produce the prototype effect. However, because the image qualities were quite salient in the stimuli that were used, a person recognition task was chosen. Based on the present results, the possibility is raised that image recognition and person recognition tasks may both be capable of producing prototype effects, but they may be unlikely to occur in person recognition when the exemplars of the same person’s face differ a lot in image quality.

Chapter 5: General Discussion

Main Objectives

The goal of the present work was to increase our understanding of how representations of faces are built from varying instances. The prototype effect was used as a tool to investigate the nature of our internal face representations. How is information derived from varying experiences with a face stored and retrieved in such a way that promotes recognition of the prototype? A closer examination of the prototype effect may reveal important clues about the representations and processes underlying human face recognition. At the very least, I aimed to highlight the importance of challenging our assumptions about the way memory works and the implications alternative theoretical perspectives may have for the field of face recognition.

The current thesis reported the results of 10 experiments. The general procedure for most of the experiments was as follows. At least two exemplars of each identity were presented in the study phase followed by a recognition task in the test phase. Among the images presented in the test series was a seen exemplar, the unseen prototype, and an unseen exemplar of each of the identities learned in the study phase. In the experiments of Chapters 2 and 3, at test, participants rated how confident they were that the *exact* image had been shown before in the study phase. Experiments in Chapter 4 used the more standard OLD/NEW face recognition task in the test phase. This task required participants to judge whether or not the test image depicted one of the people shown earlier in the experiment regardless of whether the *exact* same image had been seen earlier or not. The key comparison of interest was between recognition of the seen exemplar and the unseen prototype. The prototype effect was considered present if the recognition of unseen prototypes was higher than that of seen exemplars. Prototype effects were tested using four different types of stimuli: PROfit composites

(Chapter 2), PCA generated images (Chapter 3), morphed images (Chapter 3), and natural/unaltered images (Chapter 4). Experiments differed in terms of the variables manipulated to test potential sources of influences on the strength of the prototype effect. These manipulations tested important predictions based on different accounts of recognition memory to shed light on the origin of the prototype effect in face recognition.

Main Findings

The main finding of the majority of experiments presented in the current thesis is that the prototype effect in face recognition is a robust effect and is unaffected by a number of experimental manipulations (Chapters 2 and 3). Across several experiments, the consistent result was that after studying a set of images of the same person's face, participants were best at recognizing the average image of that face. Despite having never been presented in the study phase, the average image was more confidently recognized even compared to an image that really had been in the study phase. Unexpectedly, all of the strategic attempts to modulate the basic prototype effect were unsuccessful. So it came as a great surprise that when I tried instead to generalize my findings by using natural images of real people's faces, the prototype effect disappeared (Chapter 4).

It was ironic that all the things I thought might disrupt the prototype effect had no effect, but the one thing I thought would *not* alter the effect, the use of natural images, destroyed it near completely. Thus, we are left with two big questions. First, why were consistent prototype effects observed in the face of manipulations such as familiarity and name learning in the experiments of Chapters 2 and 3? Second, why did the prototype effect in face recognition go away when the stimuli switched from artificially constructed variations of each identity to natural images of real people?

Prototype Effects: The Role of Other Variables

Familiarity

Overall, familiarity did not change the direction or strength of the prototype effect in face recognition (Chapter 2). Although there was also some evidence that increased familiarity with the exemplars in the study phase aids their later recognition more than a morphed average, this was only true when recognizing the higher, but not the lower, quality seen exemplar (Experiment 9). Thus, the interactive effect of familiarity with the type of image shown in the test phase in Experiment 9 more likely reflected the effect of familiarity on the recognition of exemplars differing in image quality rather than how it differentially affected the recognition of seen exemplars vs. unseen prototypes. Therefore, it can be concluded that familiarity, as manipulated in the experiments of the current thesis, had roughly the same impact on the recognition of prototypes as it had on seen exemplars.

The method of familiarization is of critical importance when interpreting the role of familiarity in prototype recognition. Essentially, familiarity was manipulated by controlling how many times the seen exemplars of each face were presented in the learning part of the experiment. No matter whether a face was in the higher familiarity condition or the lower familiarity condition, participants always saw the same number of *different* exemplars of each face. The only difference was that in the higher familiarity condition, the same exemplars were repeated multiple times, but in the lower familiarity condition, the exemplars in the study phase were only shown once each. This is important because as we become familiar with a new person’s face in everyday life, we experience a wide range of variation in the appearance of their face. Being exposed to a greater number of different variants of a person's face is quite different than being exposed to the same variants a greater number of times. As Bruce (1994) points out, "experience of variations . . . within an individual may actually help rather than hinder the encoding process, just as an increase in the sample size in any analysis makes it more likely to reveal genuine differences between samples" (p. 24). Furthermore, it is well established in the categorization literature that categorization of novel patterns to newly learned categories improves with increases in the number of different exemplars studied in the category (e.g. Homa et al., 1973; Homa, Sterling, & Trepel, 1981; Homa & Vosburgh, 1976). By extension, one would expect that the greater number of variable experiences a viewer has with a person's face, the more likely it will be recognized under different viewing conditions.

If variability is important to the process of face familiarization, why then did I choose to manipulate the number of exemplar repetitions rather than the number of different exemplars? The reason was quite simple. It is difficult to conceive of any plausible account of memory that would NOT predict that encoding more exemplars of the same category (i.e. an individual's face) would increase recognition of the category's prototype more than a previously seen exemplar. In fact, increasing category size not only increases classification performance of non-studied exemplars but also decreases old/new discrimination performance in recognition memory tasks (Omohundro, 1981). Since it had already been demonstrated that the prototype effect can be observed after minimal learning (Homa et al., 2001), an account must be capable of predicting a prototype effect at least under some circumstances to be considered plausible. However, the only type of account that would not predict a stronger prototype effect after learning more exemplars from the same category would be, by necessity, incapable of predicting a prototype effect in the first place. Therefore, if familiarity had been manipulated by increasing the number of different exemplars of each face, then it would not have advanced our understanding of the prototype effect or our understanding of how we become familiar with new people's faces.

It could be argued that the method of manipulating familiarity merely increased participants' familiarity with the individual exemplars shown in the study phase rather than with the different identities of the faces per se. This is a valid criticism. However, in this particular situation, it was a more meaningful method of investigating the representations underlying the prototype effect in face recognition. By varying the number of exemplar repetitions, the participants' level of experience with a limited number of specific exemplars was increased. Because the prototype was never presented before the recognition test phase, participants had no prior experience with the prototype stimulus. Thus, the familiarity manipulation only varied the participants' experience with the seen exemplars, while leaving the novelty of the prototype unchanged. The finding that the unseen prototype was recognized more confidently than the seen exemplar even after participants were repeatedly exposed to the seen exemplars in the learning phase, was a powerful demonstration of the prototype effect in face recognition.

The following subsection will focus on how the process of familiarization relates to theories about the way we store and represent the identity of different faces in memory. Multiple trace theories of memory will not be considered in depth because they are capable of predicting almost any result given the right circumstances.

Instead, I will introduce Burton et al.'s (2005) proposal of an averaging mechanism which improves the representation of a face as it becomes familiar. Implications of their approach to the interpretation of prototype effects, particularly across levels of familiarity, will be discussed.

Theories of familiarization

How might our representations of a face change as we become more familiar with it? The most obvious account is that the more experience we have with a particular face, the more numerous and various the collection of exemplars (or episodic traces) of that face becomes. This would not only increase the chance of any given novel exemplar of a face being highly similar (i.e. above threshold) to at least one of the stored exemplars of the same identity, but would also increase the number of similar exemplars contributing to the strength of the familiarity signal produced in response to the probe. If one makes no initial assumptions about the nature of memory representations, this account of familiarity seems arguably to be the most straightforward and intuitive explanation.

However, implicit in a great deal of discussions within the face recognition literature is the assumption that each known face is represented by a unitary structure (or by a small number of different views) stored in memory. For example, any face-space model of recognition memory that represents each face as a point in multidimensional space assumes that, regardless of the number of times a person's face has been experienced, only a single representation of it will be stored (Valentine, 1991). Assuming one representation exists in memory for each known face, it follows naturally to ask how this representation might change as a face becomes more familiar. For example, it is a widely held view that the internal features become more precisely encoded or are somehow made more salient in the representations of more familiar faces. A related possibility is that participants may base their recognition judgements more on the internal vs. external features of the face to a greater extent as the face becomes more familiar (Burton et al., 2003).

Most models of face recognition are models of familiar face recognition and largely ignore the process of familiarization of newly learned faces (e.g. Bruce & Young, 1986; Burton & Bruce, 1993; Valentine, 1991). However, Burton et al. (2005) have helped to fill this theoretical gap in the literature by offering a detailed explanation of how we form stable representations of familiar faces. Burton et al. proposed that the benefits of familiarity are achieved by refining the face representations through a process of exemplar averaging. They developed an artificial system to simulate the effect of level of experience on recognizing faces across large variations in image quality. Images of 50 different celebrities' faces were collected off the internet and were morphed to a common shape. One image (shape-free) was set aside for use as a test image. A prototype representation of each person's face was made by taking the mean intensity for each pixel across images of the same identity. The number of images contributing to each average was varied between 1 (a single image) and 19. PCA was then conducted on these average prototypes, and this generated 50 eigenfaces which the test images were projected onto. It was considered a hit (i.e. correct recognition) if the nearest neighbour of the test image was the average of the same person's face. An exemplar version of the system was also developed for comparison purposes. It used the same images as the averaging system except no averages were constructed. Instead, the PCA was conducted on the set of individual images. Nearest neighbour and summed similarity matches were made between the test image and the images that the PCA was ran on.

The important findings from Burton et al.'s (2005) simulations were as follows. First, systems based on image averaging consistently outperformed systems based on the individual images. The summed similarity version of the exemplar-based system showed the worst performance, so only findings from the nearest neighbour systems will be summarized. Second, as the number of images used per face increased, so did the hit rate. Thus, it seems that an average representation will be better able to support accurate face recognition if it is built up from a greater number of images. Note also, that when the system learned only one image per face, hit rates were only 13% (chance= 2%). But as the number of images increased from 3-9, the advantage of the prototype system over the exemplar-based system remained fairly stable (~6% difference). Hit rates peaked at approximately 75% when image averages were composed from 19 different images of the same person's face. Thus, it seems that storing faces as averages or as instances are both plausible representational strategies that might underlie our poor ability at recognizing unfamiliar faces and our incredibly accurate recognition of familiar faces. Since Burton et al.'s focus was on demonstrating how an average prototype can provide a powerful, robust representation of a face, they do not specify what exactly caused their exemplar-based system to perform more accurately when there were higher numbers of images per face in the learning set. In particular, it was not mentioned how the number of images per face affected the eigenfaces generated by the PCA conducted on the sets of individual images which affects the similarity relations among images encoded in the space.

Burton et al. (2005) highlighted several characteristics of image averages which make them a good candidate for representing familiar faces. Since even unfamiliar faces can be recognized quite accurately when there is no change in image between study and test, what we gain by becoming familiar with a faces is a reduced sensitivity to properties of the image which are not related to the identity of the face. Burton et al. argued that inconsistencies among images of the same face will be removed from the representation as a natural consequence of the averaging process. Thus, the ratio of relevant to irrelevant identity information will be higher (on average) in an average image than in any given one of its constituent images. Of course, as the number of images increases, the contribution each image makes to the average representation decreases, so the average will come to reflect more what the image have in common and less what they do not have in common. If one assumes that internal features exhibit less variation across a large number of encounters than external features, then an averaging process might also explain why the internal features of a face become more important to recognition as it becomes more familiar.

To gain support for their proposal that an identity is represented by the average of all its seen images (i.e. exemplars), Burton et al. (2005) tested how well human participants could identify celebrities from their image averages. Participants performed faster in a name verification task and were also more likely to respond with the correct name (or another piece of identifying information) when presented with an average image of a celebrity (81% hits) compared to one of the individual images that the average was made from (77% hits). Thus, Burton et al. did observe a prototype-like effect in face identification using familiar faces. However, face learning was uncontrolled, so the average image could not be constructed to correspond to the exact mean of the participant's previous experiences with a given person's face. Furthermore, none of the individual or averaged images was expected to have been seen before by participants. Therefore, performance was being compared between two unseen images of a face rather than between an unseen average and a seen exemplar image, as was the case in all of the experiments in the current work (with the exception of Experiment 6).

Burton et al. (2005) also used human participants to test the hypothesis that the more images per face being averaged together, the more identifiable the resulting image will become. The reasoning here is that identification performance is a function of the degree of match between the probe image and its corresponding identity representation stored in long term memory. Participants were, in fact, faster to correctly verify the name of face from an average image when the average was built from more images. The authors suggested that experiencing a greater number of varying images improves the quality of the average representation.

Essentially the same reasoning was used to account for the prototype effects found in the current experiments within a prototype framework of representation (Experiments 1-6). Participants were more confident that they had seen the unseen prototype image of a person's face even compared to the images that they actually had seen earlier in the experiment. If participants formed an image average representation of each studied identity, then the prototype image was more confidently recognized because it approximated the prototype representation stored in memory.

It was also found that repeating the exemplars in the study phase produced no change in the strength of the prototype effect (Chapter 2). At first this may seem entirely consistent with Burton et al.'s (2005) account of face learning. Repetition of exemplars should not change the mean value for the identity which corresponds to the abstracted prototype. Thus, the degree of match between the test image and the abstracted prototype representation should not change across levels of familiarization for either the prototype or exemplar test images. But repeating the exemplars in the study phase led to dramatic increases in rated confidence for prototypes, seen, and unseen exemplars alike. It is reasonable to assume that additional repetitions may produce a prototype representation that is a slightly closer approximation of the "true mean" of a pair of images because random noise in the encoding process will tend to average out. However, I would expect this to have very minimal effects on confidence ratings and would likely only become a factor in circumstances where encoding was made difficult (e.g. brief exposure, adding random noise to images). Therefore, if the prototype effect in face recognition is due to storing an average representation for each identity, then the effect of familiarity gained through repeated exposure to the training images cannot be fully accounted for by a refinement of the underlying face representation.

Nor can the effect of additional repetitions be accounted for by an increase in the likelihood that images of the same identity will be averaged into a prototype representation. This was the reasoning used in Chapter 2 to predict a stronger prototype effect following additional repetitions, but the results of those experiments showed no such change.

To explain how repetition of exemplars increased the confidence ratings to the prototype but not the strength of the prototype effect, I will assume that whatever is causing the small advantage for prototypes over seen exemplars after minimal exposure is also responsible for the same effect observed after additional learning. Furthermore, I will also assume that the large difference in confidence ratings between faces that were and were not in the familiarization training had a different underlying cause (than the prototype effect). The prototype account of the prototype effect is that the prototype stimulus is recognized best because it has the highest degree of match with the prototype representation in memory. Thus, the relative degree of match between prototype and seen exemplar probes to their prototype representation in memory should not be altered by repeating the studied exemplars (because repetition does not alter the average value). If repetition actually improved the prototype representation, then this improvement must have affected the recognition of prototypes and seen exemplars equally. It is not obvious what sort of cognitive model would function in this manner. However, a detector-based model like Morton's (1969) logogen model might account for the pattern of results by assuming that every presentation of a person's face slightly increases the resting activation of the detector for that person's face. Thus, increasing the number of repetitions of a face in the study phase will increase the likelihood that a test face will be identified as the repeating face (or judged as old). Of course, this is not a very satisfying explanation because with repeated exposure to a face, the probability that the detector for that face will reach threshold in responses to *other* people's faces will also increase.

Recent work by Megreya and Burton (2007) demonstrated that familiarity increases participants' ability to match two different images of the same person's face and also to discriminate between images of two similar looking people's faces. This evidence suggests that as one becomes more familiar with a person's face, it will be less likely to be confused with other people's faces. McClelland and Chappell (1998) developed a model of single-item recognition to account for the effects of repetition on recognition memory. In their model, as one becomes familiar with something, they acquire more precise knowledge of what properties it does and does not have. Their approach involves estimating the likelihood of observing a set of features assuming that the probe is an instance of a particular stored item. The stored representations themselves are probability estimates that become more precise through experience. Therefore, after repeated exposures, a representation of item x will be more likely to be activated by an item x probe, but *less* likely to be activated by a probe of a different item. Importantly, as repeated exposure to item x increases, recognition of a probe of a similar item will at first increase and then gradually decrease. How many exposures are needed before the false alarm rate will peak and reverse depends on how similar the distractor item is to the target item x. This may provide a clue to why participants falsely recognize the prototype more after repeated exposure. The idea that items could be represented as a vector of probability estimates is intriguing, but their model does not generalize easily to representations of items such as individual faces which vary subtly from one encounter to the next. However, their approach to the problem may be promising if it can be adapted or extended to model learning in face recognition.

McClelland and Chappell's (1998) model does do a good job at accounting for prototype effects found using word lists. Studies of false recognition of words semantically related to words in a preceding study list have manipulated the frequency of studied words. Hall and Kozloff (1970) varied frequency of a critical target word in the study list from 1 to 7. They found that false recognition of the critical lure (a word chosen to be similar to the target word) first increased and then decreased after reaching a peak at a presentation frequency of 3. In a related study, Hall and Kozloff (1973) showed that increasing the number of *different* associate words, instead of the number of repetitions of the *same* associate word, steadily increased false recognition without any decline. Although, recognizing words in a task like this would be different than recognizing faces, the findings of Hall and Kozloff may have some implications to the current investigation. If, under certain circumstances, it is true that as the number of repetitions of an exemplar increases, recognition of a similar, but otherwise unseen, exemplar will increase to a point and then decrease, then there are at least some grounds for the suggestion that familiarity might have a nonmonotonic effect on the prototype effect. Assuming that the unseen prototype is similar to the studied exemplars, it would be falsely recognized at low levels of repetition because it activates traces of the studied exemplars with sufficient strength. At higher levels of repetitions, representations of the exemplars become more highly differentiated, and thus, the prototype stimulus is no longer a close enough match for generalization to occur.

Hintzman and Curran (1995) found similar results when the targets and critical lures differed only in terms of the plurality of a base word (e.g. bell vs. bells). Again, repetition of the target word in the study phase increased false recognition of the critical lure up to a frequency of 3, and then gradually decreased as frequency was increased beyond this point. But this gradual decline of false alarms was very slight, and even after 20 presentations of the target word in the study list, participants were still poor at rejecting the critical lure (false alarm rate ~50%). Even when participants were made aware before the study phase that they would need to remember whether the study word was singular or plural in order to distinguish it from the lures at test, participants showed very little improvement past a study list target frequency of 3.

The findings of Hintzman and Curran (1995) as well as those of Hall and Kozloff (1970) are generally consistent with what was found about the effect of repetition on the recognition of unseen exemplars in Experiments 1-3. An interesting consequence of increasing participants' familiarity with the studied exemplars was that the seen exemplars, the prototypes, and even the unseen exemplars were all rated more confidently as having been seen before. While one might anticipate increased recognition of the prototype after familiarization training, it was very surprising that unseen exemplars, falling outside the range of experienced variation, were also rated higher. In general, after familiarization training, participants went from being uncertain as to whether the unseen exemplar had been seen before to guessing with low confidence that it had been seen before. In other words, participants were actually *worse* at correctly rejecting the unseen exemplar when they were made more familiar with the studied exemplars. So just like experiments using word associates, participants made more correct rejections of the unseen exemplar (or critical lure) when the studied exemplar had been shown only once in the study phase compared to when the studied exemplar was repeated (although correct rejection was best when the unseen item was unrelated to study items). Because only two levels of familiarity were tested, one cannot be sure of what shape the function relating the number of repetitions to false recognition of the unseen exemplar would take. Hintzman and Curran's finding suggests that because the unseen exemplar shares the same basic features with the studied exemplar, performance on the unseen exemplar might show very little change one way or the other after only a few repetitions of the studied exemplar. However, Hintzman and Curran did discover a way to get participants to achieve near perfect discrimination, and this will be discussed later in the next section.

Intentional vs. incidental learning.

Comparable prototype effects were found regardless of whether participants were warned at the start of the study phase of the recognition test that would follow or not (Experiments 1 and 2). This finding does not support Bruce et al.'s (1991) finding that the prototype effect tends to be weaker when the exemplars in the study phase were learned using an intentional memorization task compared to an incidental learning task. However, the current results are not entirely inconsistent as Bruce et al. also noted that intentional learning might only weaken the prototype effect when the exemplars of the same identity were easily discriminable.

If intention does play a role in prototype recognition, it is not likely to have any effect beyond what could be achieved by other means of increasing (or decreasing) attention to the details of the images. Bruce (1994) argued that any manipulation that draws attention toward the variation among exemplars of the face has the potential to decrease the strength of the prototype effect provided that these differences were discernable in the first place. Bruce reported results from an experiment that used internal feature-shifting to create within-identity exemplars. This technique causes the length of the chin and forehead to vary among exemplars of the same face, but the length of the nose only varies among exemplars of different identities. When participants were asked to make judgements about the length of the chin in the study phase, the strength of the prototype effect was reduced compared to when participants were asked to make judgements about the length of the nose. Similarly, Neumann (1977) found that the basic instruction “try to remember each face exactly as shown” led to a prototype effect in face recognition (prototype compared to unseen exemplar), but adding “so that if I later show you an identical face you will be able to recognize it, but if I show you a very similar face you will be able to reject it” led to reverse prototype effect. Adding to the basic instruction information as to which dimensions the faces would vary on also produced a reverse prototype effect, but adding information about the dimensionality did not influence the prototype effect when the elaborated version of the basic instruction was given.

The same-face exemplars used in the current experiments were likely too similar to allow the intention of the participant to disrupt the mechanisms underlying the prototype effect. Even still, it is interesting that despite instructions to memorize each exemplar separately in the study phase and to make a judgement about the prior occurrence of the specific image in the test phase, participants were unable to reject the prototype. Since the learning task was devoid of category feedback, it seems as though the visual similarity alone resulted in participants being incapable of selectively recognizing an exemplar of a face in isolation from other similar exemplars. The implication here is that an intent to form distinct memories for two separate events (or exemplars) will not affect our ability to do so provided they are highly similar. Note that even though participants were aware of how subtle the differences between the actually studied images and their unseen counterparts would be, they nonetheless were still fairly confident that the prototype image was in fact the exact image they had seen before. Note also that it was not the case that the prototype was completely indistinguishable from its two studied exemplars. If it had been, participants would have been equally confident about the prior occurrence of both seen exemplars and prototypes, but they were even more confident about the prototypes.

Similarly, Hintzman and Curran (1995) found that only when participants were forced, rather than strongly encouraged, to judge on a trial-by-trial basis whether the study list word was singular or plural could they achieve near perfect discrimination of target and critical word pairs (e.g. target = *bells*, critical lure = *bell*). The words bell and bells are clearly distinguishable, but nonetheless highly confusable in memory. Attempting to remember the exact form of the word will not lead participants to correctly reject the lure; however, making a specific judgment about the discriminating feature will dramatically improve performance. Participants on this task likely used information about the discriminating feature in a similar way to how participants might have used information gained by judging the length of the chin in the experiment reported by Bruce (1994).

Naming

Whether exemplars of the same facial identity were all learned with the same name or with two different names during the learning phases, made no difference to the strength of the prototype effect in three separate experiments. This was true when prototype effects were observed (Experiments 5 and 6) and when they were not observed (Experiment 10). Although it is difficult to conclude based on null results that the prototype effect in face recognition cannot be influenced by naming learning, at the very least, it seemed to be the case that it is not easily influenced. The implications of this are actually very important. In particular, the naming experiments helped bridge the gap between the categorization literature and the face recognition literature.

Homa et al. (2001) investigated whether category formation is a necessary prerequisite for false recognition of the category prototype. In their study, the stimuli were faces, but the categories were three “families” of similar looking faces rather than categories of three different identities. Any two faces from the same family may or may not have resembled the same individual, but they would have at least looked quite similar. They found that participants who were trained to classify faces into one of three families were just as likely to falsely recognize the unseen category prototype as one of the training faces as participants who were simply shown the training faces and told to remember them. Participants who merely viewed the training faces were unaware of the three families of faces and did not acquire category knowledge. If it is true that unseen prototypes are falsely recognized because they correspond to an abstracted prototype of a category, then why would the unseen prototype also be falsely recognized when no categories had been learned? Thus, recognition of the unseen prototype should not be taken as evidence of an earlier abstraction process. The important implication of this to my experiments is that it may not have mattered whether or not participants associated two studied exemplars with the same name. They may have been just as likely to judge the unseen prototype as old in either case.

Based on their findings, Homa et al. (2001) concluded that classification training (with feedback) may improve a participant’s ability to correctly classify the unseen prototype as a member of a learned category, but is unlikely to improve the participant’s ability to correctly reject it as being one of the training exemplars. This would explain why no effect of naming was found in Experiments 5 and 6. Experiment 10 was a little different in that participants were asked to judge if a test face was one of the same people shown in the study phase rather than the same image. But since the name of the face was not requested, it is not entirely clear whether predictions based on a recognition or a categorization task are more appropriate. Given that name learning had no impact on how well participants recognized a target from its prototype image, results suggest participants treated the task as one of recognition memory.

Context

Using cartoon bodies to provide an associated context for learning faces did not influence the participants' confidence of having studied the unseen prototype image relative to other image types including the actually seen exemplars (Experiment 4). Applying a consistent context to the four exemplars of the same identity did not increase the strength of the prototype effect relative to applying varied contexts. Therefore, no evidence was found that shared context facilitates any sort of abstraction mechanism that may be operating on exemplars of the same identity during the study phase. A similar result was reported in an unpublished study by Bonner, Bruce, and Burton (2003) who found no effect of consistent vs. varied semantic information on learning new faces. In their experiment, over 3 days of training, they familiarized participants with a set of previously unfamiliar people's faces using short video clips alongside an auditory dialogue containing either three or nine pieces of fictitious semantic information about each face. In the consistent condition, faces were presented with the same nine pieces of semantic information which were given each time the clip was shown on each of the three days. In the varied condition, faces were presented with three pieces of semantic information on the first day, a different three on the second day, and another different three on the third day. Consistency of the semantic information during learning did not influence participants' ability to judge whether two images were of the same person or two different people in a simultaneous face matching task involving the familiarized targets and similar looking distractors. Face matching, particularly on trials where only the internal features were visible, was used as an index of familiarity because participants should get better at this task as they become more familiar with a person's face. It is possible that if context does play a role in face recognition, it has more to do with judgements of memory rather than our ability to recognize the same identity across variable viewing conditions or to learn to differentiate identical twins. For example, context might affect how well we can remember *who* a person is, but it may be unlikely to affect our ability to perceive the diagnostic elements of a face's identity.

Gruppuso, Lindsay, and Masson (2007) found that providing an associated context picture with each face image during learning affected the subjective experience of recognizing faces. In a remember/know test (see Tulving, 1985), participants were more likely to report consciously remembering a target face from the study phase if it was presented with the same studied context at test compared to a context studied with a different face or a new context. Conversely, when target faces were presented with a new context, participants were more likely to report merely "knowing" the face had been seen before than they were to report consciously remembering it. Furthermore, the false alarm rate was higher for distractor faces presented with one of the studied contexts compared to a new context. Thus, participants may sometimes attribute their memory of the context picture to their memory of the face. However, memory judgements were not simply based on memory for the face plus memory for the context because it was only when the face and context had been studied *together* that conscious recollection of target face was enhanced. This might explain why participants in Experiment 4 were more confident that the image had been seen before when it was presented with the same context at study and test despite context having no effect on the prototype effect. Reinstating the studied context at test may have increased participants recognition for the identity of the test image, but it may not have affected their ability to discriminate subtle differences among images of the same face. This again implies that participants cannot detangle their memory for a particular image of a face from the memory of that person's face in general.

Similarity of Studied Exemplars

Due in part to the lack of effect of other manipulated variables on the strength of the prototype effect, the results discussed so far have been taken as suggestive of an underlying mechanism driven primarily by exemplar similarity. The role of exemplar similarity in learning new faces was investigated in Experiments 1-3 in Chapter 2 as well as in Experiment 6 in Chapter 3. In the experiments of Chapter 2, similarity of the studied exemplars was termed "distance." Distance referred to the distance between same-face exemplars in terms of the position of the internal features along the vertical axis of the face outline. Exemplars in a near pair were more similar to one another than exemplars in a far pair. I thought that additional learning might affect the prototype effect differently depending on the discriminability of exemplars of the same identity. However, no such interaction was found. As for the effect of similarity on the strength of the prototype effect, the interpretation is a bit more complicated. Overall, similarity of the studied exemplars did not affect recognition of the average prototype relative to exemplars that were close to the average. But similarity did have an effect on recognition of seen exemplars. Seen exemplars were given higher recognition ratings when they were more similar to the other exemplar of the same face presented in the study phase. Prototype and exemplar-based models alike can easily account for this finding either by distance from the stored prototype or by the summed similarity to stored exemplars or episodic traces. Thus, prototype and exemplar-based models predict that the recognition of both prototypes and seen exemplars would increase with increasing similarity of studied exemplars for the same reasons just given. Because of this, inferring the role of similarity in prototype abstraction from changes in the strength or direction of the prototype effect in face recognition is often problematic.

The results from Experiments 1-3 conflict with those of Cabeza et al. (1999) who found that similarity of studied exemplar did affect the strength of the prototype effect using a similar feature shift manipulation. Increasing the similarity of studied exemplars increased recognition of the prototype, but had inconsistent effects on seen and unseen exemplars. However, similarity was severely confounded with the distinctiveness of studied exemplars. In their study, the range of internal feature shifts covered the whole length of the face which made the studied exemplars at medium similarity to look a bit strange, and at low similarity, they looked very bizarre. Not only would this make lower similarity exemplars more memorable, but it would also draw participants' attention to the variation in feature placement (most noticeably, size of chin/forehead). As Bruce (1994; Bruce et al., 1991) has also noted, manipulations that increase attention to the subtle details of the images will likely decrease the strength of the response to the prototype image. Bruce et al. found strong prototype effects for stimuli using a feature shift variation size somewhere in between that used in the near condition and that used in Cabeza et al.'s large variation size. Although the size of the feature shift used by Bruce et al. (in the bimodal condition) was larger than what was used here (Experiments 1-3), the internal features did not approach the borders of the face. They also noted that in unpublished experiments using variation sizes larger than those they reported, the prototype effect did not weaken. Thus, using a feature shift manipulation, similarity of studied exemplars may not have a strong effect on the prototype effect unless exemplars are made so dissimilar that they may not appear like plausible faces.

Results of Experiment 6 did not reveal any clear role of similarity on the strength of the prototype effect. Actually, similarity did not produce any significant effects at all. However, when exemplar pairs were learned with two different names, there was a trend towards a weakening of the prototype effect. In fact, in the different name condition, there was no sign of the prototype effect at medium similarity and even a slight trend for a reversal of the effect at low similarity. Even still, participants' mean rating to prototypes learned from low similarity exemplars with two different names was just over 7 on a 1-9 scale (5=0% confident; 9=100% confident that the image was one of the studied images). Pilot testing of the stimuli ensured that low similarity exemplars were perceived as two different people's faces when viewed side-by-side. This is important because if one uses the prototype effect in face recognition as a tool to probe the nature of memory representations of individual faces, then a finding that prototype effects can be observed using exemplars of different identities may severely limit the conclusions one can draw using such an approach.

Other studies have also found that unseen prototypes based on studied images of two or more different identities are, at the very least, difficult to reject as having been one of the studied images. For example, Cabeza et al. (1999) found that prototypes built from four studied images of different identities were given a confidence rating of approximately 5 (i.e. almost no confidence) by participants using a 0-9 scale similar to the one used in Experiment 6 (1-9 scale). Busey and Tunnicliff (1999) found a significant prototype effect for a 50% morph between two studied images of different people's faces, but only when the two faces being averaged belonged to similar, but not dissimilar, looking people. Importantly, however, both of these studies reported an effect of exemplar similarity. The prototype effect weakened or disappeared when the similarity of studied exemplars was decreased[4]. I did not find a significant effect of similarity on the magnitude of the prototype effect (Experiment 6), but there was a trend in that direction at least in the different name condition. Furthermore, the prototype effect found by Solso and McCarthy (1981) using a feature swapping technique is also inconsistent with the hypothesis that prototype effects should only be observed when the prototype is the central tendency of the studied exemplars from the same identity and not when they are from different identities.

Considering the results from Experiments 1-7 altogether, participants clearly showed the ability to discriminate seen exemplars from unseen exemplars of the same identity that are outside the range of variation of the studied exemplars. Because the unseen exemplars were often very similar to the studied exemplars, this implies that

---

[4] Cabeza et al. tested 6 levels of similarity between studied exemplars. There was no difference in recognition confidence ratings between prototypes and seen exemplars at levels 6 (highest similarity), 5, 3, and 2. There was a prototype effect at level 4 and a reverse prototype effect at level 1 (lowest similarity). Only level 6 similarity exemplars were judged as belonging to the same identity. Exemplars of levels 3, 2, and 1 were judged as different people's faces while exemplar in levels 4 and 5 were not confidently judged as belonging to either the same or different identities. Thus, the pattern of prototype effects was more complicated than, say, a linear function of exemplar similarity.

participants can remember the images they studied quite precisely. On the other hand, there is something about the way exemplars of the face are being stored or retrieved that also leads to false recognition of the prototype and other exemplars of the same face that lie within the range of experienced variation. In fact, recognition confidence at test was highest for the unseen exemplar that was the furthest possible distance (in an image space) away from the individual studied exemplars while still remaining within their range of variation. It is surprising that such reliable differences in confidence ratings were observed between prototype and seen exemplars even when these images were very difficult to distinguish. Particularly in Experiments 4 and 5, participants often commented, when viewing example stimuli side-by-side, that they could not see any differences between the seen exemplars and the prototype or the local averages. Furthermore, Experiment 7 demonstrated, using stimuli from Experiments 4 and 5, that participants could not distinguish the prototype image from a seen exemplar image from memory when only the one seen exemplar was presented in the study phase. And yet these two experiments (4 and 5) yielded the largest prototype effects out of all the experiments conducted. Bruce (1994) also noted that in some of Bruce et al.'s (1991) prototype experiments, differences between images were barely perceivable, and yet they had a huge impact on recognition memory.

In conclusion based on the results reported in present work (Chapters 2 and 3) and in previous studies described in this section, the prototype effect in face recognition is observable when the prototype stimulus is nearly perceptually indistinguishable from the studied exemplars and also over a much larger range of similarity values than has been assumed in the past. Although prototypes built from studied exemplars of different identities do not always yield a significant prototype effect (i.e. higher recognition of prototype images than seen exemplars), it is clear that they are difficult to correctly reject, and their false alarm rate (or rated confidence) is always more similar to the hit rate for seen exemplars than the false alarm rate to unseen exemplars even when the unseen exemplar is closer to a seen exemplar than the prototype is (provided the unseen exemplar falls outside the range of the seen exemplars).

Learning Identity From Natural Images of Faces

In previous studies, the prototype effect in face recognition has been found using Identikit and Mac-a-Mug faces (Bruce et al., 1991; Solso & McCarthy, 1981) as well as morphed photographs (Busey & Tunnicliff, 1999; Cabeza et al., 1999). Variations on identities constructed by feature swapping (Solso & McCarthy, 1981), feature displacement (Bruce et al., 1991; Cabeza et al., 1999), and morphing between two or more images of different people's faces (Busey & Tunnicliff, 1999; Cabeza et al., 1999) have all been shown to be capable of producing prototype effects under the right experimental conditions. Although the prototype effect in face recognition has been demonstrated using realistic, high quality, colour images, none of these previous studies have tested how the effect generalizes to the circumstance where the prototype is the average of a set of unaltered images of a real person's face.

Given that prototype effects can occur to an average morph between two studied images of different identities (Busey & Tunnicliff, 1999) and to the average of a set or pair of highly varying exemplars based on an artificial identity (Bruce, 1994; Bruce et al., 1991; Cabeza et al., 1999), one would expect similar effects should be found using natural variations of real identities. However, this was not what was found in Experiments 8-10. Instead, no clear recognition advantage was found for a 50%

morph of two different images of the same person’s face relative to the actually seen images. The average morph was not necessarily recognized worse than both of the seen exemplars, but it definitely did not show a clear advantage like that of prototypes in Experiments 1-6.

How did the variation in the natural images (Chapter 4) differ from the artificial variation used to construct identity face sets in Chapters 2 and 3? The natural images contained a lot of variation due to having been taken with different cameras (i.e. one was a photograph and one was a video still). This was not true of any of the artificial variants because they were all derived from images taken under highly standardized conditions. Thus, to recognize the high and low quality images of the same face, participants would have to overcome extraneous sources of image variation such as differences in contrast, brightness, graininess, and blurriness as well as small changes in expression and head angle. It was assumed that if the prototype effect was tapping a mechanism whose purpose was to form representations of individual faces from instances that vary on these dimensions, then one might observe even stronger prototype effects using natural images of identities than what had already been found using artificially generated exemplars. As it was not the goal to study the effect of “image quality” per se, it was not anticipated that the changes in image quality among the stimuli would have had the potential to destroy the recognition advantage found for prototype images in the previous experiments.

The failure to find a strong prototype effect using natural identity variants in Chapter 4 suggests one of two things. It is possible that the prototype effect in face recognition does not tap the means by which we form stable representations of individual faces.

The reasoning is based on evidence from one of Busey and Tunnicliff's (1999) experiments that used a very similar methodology but found that 50% morphs of two studied images of different (but similar looking) people's faces had a higher hit rate than the actually studied images despite never having been seen previously. The studied images in their experiment were taken under standardized conditions. Therefore, it seems reasonable to question whether the mechanisms being tapped by the prototype effect truly do operate by combining instances of the same face while keeping instances of different identities distinct in memory (again this could occur during encoding or retrieval). Alternatively, our face recognition system might have an imaged-based front end to it (Bruce, 1994). The prototype effect then could be tapping the mechanisms that build individual face representations, but if these mechanisms represent faces based on an analysis of low-level image properties (e.g. a PCA based approach), they may not be able to easily separate the superficial properties of the image from the more important identity-related information contained in the face until more variation of the face has been experienced (Burton et al., 2005).

Therefore, it is possible that when learning a previously unfamiliar face from a small number of natural images (in the present experiments only two), the image differences between the two images are too great to easily engage an abstraction mechanism that would produce an average of the two images (either during encoding or retrieval). Similarly, anything that makes a specific image more distinctive or memorable will likely facilitate its subsequent recognition. If superficial properties of the image (i.e. not identity related) often increased the memorability of an image, then the average morph will not have shared these properties. In contrast, artificially generated exemplars and their prototypes should share these properties more evenly because they are derived from the same original images. Thus, certain image properties could potentially increase the recognition of a studied exemplar, and these properties would not be shared by an average morph unless the other studied exemplar(s) (those of the same average) also shared these same properties. Since artificial variants generated from the same images are more likely to share any such properties than are natural variants taken from different images, the natural images used in Chapter 4 were more likely to be at an advantage over their prototype than were the artificial exemplars used in Chapters 2 and 3.

Thus, the high and low quality images of the same face might be less likely to be stored or retrieved in a way that promotes recognition of the average image compared to the images used in Experiments 1-6. Also some superficial properties or details of the high or low images may have made them more memorable which is a slightly different issue than the likelihood (or the extent to which) they will be stored or retrieved separately or together. Furthermore, high quality images may have differed from low quality images in terms of their memorability. Memorability of a face can be assessed by asking participants to rate how easy they think it would be to remember that face (Vokey & Read, 1992). Memorable faces often have distinctive elements or features which make the face easier to encode and retrieve. Faces high in memorability have been found to be characterized by small, local, distinctive features (O'Toole, Deffenbacher, Valentin, & Abdi, 1994). These memorability related cues are more likely to be captured by information in higher spatial frequencies (Busey, 2001) which likely contained more information in the high quality images. This may explain why, overall, high quality images tended to be recognized better than low quality images. And if a subjective sense of memorability is involved in the rejection of distractor images as Vokey and Read (1992) suggested by providing “negative evidence”, then this would also explain why low quality distractors had much higher false alarm rates than high quality distractors.

Task Demands

In the experiments of Chapters 2 and 3, an exemplar or image recognition task was used. Indeed, this is the standard task used to assess recognition prototype effects. In Chapter 2 experiments, the rating task used at test was identical to the one used by Bruce et al. (1991, Experiments 7 and 8). In this task, participants are asked to rate the familiarity of each test image on a 10 point confidence scale (e.g. 0= *definitely not a seen picture*; 9= *definitely a seen picture*), and they are told to make their judgements about the exact image rather than the face in general. Because of the concern that asking participants to rate familiarity using a scale that measures confidence would needlessly complicate the interpretation of the results, in Chapter 3 experiments, participants were simply asked to rate their recognition confidence of having seen the identical image in the study phase using the same scale as before.

It may seem counterintuitive to use an image recognition task when the goal is to learn about face recognition not picture recognition. This approach rests on the implicit assumption that whatever mechanism causes the prototype to be falsely recognized as a seen image will also cause it to be correctly recognized as a seen person’s face if the task was one of face identity recognition. Cabeza et al. (1999) tested this assumption directly by comparing image recognition confidence on the same 10 point scale described above to an old/new task where participants rated how sure they were that they had seen the same *person*’s face before in the previous study phase. Results from the two tasks were compared in an experiment using identity variants produced by feature shifting (similar to the method in Chapter 2) and in another experiment using a morphing technique that created variants across three levels of similarity. In both experiments, the two tasks produced the same pattern of results; that is, both tasks indicated higher recognition of the prototype compared to seen exemplars of the same face.

In the experiments of Chapter 4, natural images of faces and an old/new face identity recognition task were used to generalise the results from experiments in Chapters 2 and 3 using a method with greater ecological validity. However, unlike the previous experiments, no clear recognition advantage for prototypes was observed using the more natural stimuli and recognition test. The evidence from Cabeza et al.'s (1999) experiments suggest that the old/new face task and the image confidence rating task are equivalent in that they both tap the same mechanism and produce similar patterns of results. Bruce (1994) also maintained that studies of prototype effects using a test of picture recognition as a measure of face recognition probe the same representations that underlie face identification in everyday life. Although she slightly qualifies her claim by emphasizing that image recognition tasks in prototype studies tap these representations to the same extent as standard old/new face recognition tasks when target faces are studied and tested using a different image. Results from an experiment (see end of Chapter 2) that was not included in the current thesis provided a clue which suggests that the claim that the two tasks are equivalent warrants further investigation. This was the only experiment in which no prototype effect was observed apart from the experiments of Chapter 4. It too used a face identity task. Stimuli were PROfit faces similar to those in Experiments 1-3, but there were some differences not just regarding the stimuli, but also the details of the task. Based on the evidence at hand, it seems unlikely that using an old/new face identity task would have made a large difference to the observed pattern of results, but the possibility cannot be entirely ruled out.

Implications to Our Understanding of Face Recognition

One problem with the current approach of using the prototype effect to investigate how we represent newly learned faces in memory is that its generalization to face recognition in everyday life may be limited. There are several differences in the tasks and stimuli that make comparisons between the laboratory experiments reported here and everyday life quite difficult. Beyond the suspicious lack of a prototype effect using natural (vs. artificial) images (Chapter 4), the use of static images may have had some important consequences. Although this criticism holds true for the majority of face recognition research using static images as stimuli, it is of particular relevance to the current work because of the contrast between representing faces as episodes or abstractions. Like in many other studies of face recognition, static images of faces were used in the experiments because they are the easiest type of stimulus to control in terms of their properties and in their presentation to participants. However, in everyday life, we often see faces as a continuous stream of visual information, not as discrete episodes or exemplars.

Indeed, it is under naturalistic conditions, while viewing continuous motion in faces that we are most likely to experience a more or less full range of variation that would be needed to generalise to future encounters with the same person’s face. Not only does a dynamic face provide a large range of within-identity variation, but it also provides very strong top-down support to guide the abstraction of identity information. When viewing static images, the visual system cannot draw the connections among exemplars of an unfamiliar face. However, when viewing a continuous stream of visual input from a face, the identity is already known to be unchanging, so abstraction can proceed without the hindrance of trying to judge the likeness of one instance to the next.

Curiously, in the literature regarding episodic theory of memory, the issue of how we break experiences into discrete episodes is largely ignored. Ironically, the theory that attempts to explain all acts of memory in terms of the storage and retrieval of events has very little to say about what actually constitutes an event. This problem has been partially sidestepped in the vast majority of experiments by using word stimuli which due to their static and discrete nature, map much more easily onto the concept of an episode than do live faces.

Homa et al.’s (2001) finding that participants were just as likely to judge the unseen category prototype as old as a studied exemplar regardless of whether participants were even aware of the experimental categories suggests another limitation of the current approach. Whatever mechanisms underlie the prototype effect in recognition may not be the same mechanisms that underlie the prototype effect in categorization or concept formation. This would be more problematic for a prototype account than an exemplar-based account because the prototype account is that the unseen prototype is falsely recognized because it corresponds to a representation of a category’s central tendency abstracted from experiences with category members. Therefore, if the unseen prototype is falsely recognized as old when no category prototypes had been abstracted, then a pure prototype view is clearly inadequate. An exemplar view can better accommodate this finding because it does not posit that false recognition of category prototypes is due to representing categories as abstracted prototypes. It is the similarity of the prototype to the exemplars encoded in memory that matters most, and many exemplar models will compute recognition based on the summed similarity of a probe (e.g. category prototype) to all exemplars stored in memory regardless of category membership. Therefore, if we want to understand how we represent faces, then measuring how confident participants are that an image was shown earlier in the experiment is very different from measuring their recognition of an image as a specific familiar person. The prototype view has difficulty explaining the prototype effect under circumstances where abstraction is unlikely, and while the exemplar view can explain this, the explanation it provides suggests that the prototype effect in recognition does not tap into the same processes that would be used to identify a face (recognize the identity).

Suggestions for Further Research

Future directions for investigating the prototype effect in face recognition

Beyond the current work, there has been very little research focussing on the degree of learning in prototype face recognition. Since Homa et al. (2001) found an effect of repetition on old/new judgements of category prototypes, it would be informative to follow up on his findings using categories that are more like different identities rather than different "families" as in their study. Homa et al. found that the prototype effect strengthened as the number of repetitions of studied exemplars increased, but only for categories constructed using a feature swapping technique. In the feature swapping method, the prototype was actually a composite of the most common features among its exemplars. In contrast, they observed weaker prototype effects following additional repetitions for face families made by slightly distorting the features of the face so that the prototype did not share any totally identical features with its category members. However, in the later *ill-defined* type of category, the category members themselves also did not share identical features with each other, so this may also have caused the pattern of results they observed. Therefore, future research should test for an effect of repetition using categories based on exemplars that are re-combinations of a fixed set of similar features. The prototype at test would be either a composite of the most frequent feature values or a morphed average of the studied exemplars. Although comparing recognition between two different types of images is problematic, comparing differences in recognition across levels of study repetitions could lead to a better understanding of why increased learning of certain types of perceptual categories lead to recognition prototype effects and some do not.

To understand how the prototype effect might reveal important information about how we learn faces in more naturalistic situations, future research should move away from presenting faces as discrete instances. Since it has been argued that viewing a dynamic face may provide ideal conditions for abstraction of an average identity prototype (see Bruce, 1994; Burton et al., 2005), presenting faces in video clips would provide a more meaningful approach to the problem of how we form stable representations of familiar faces. Prototypes could be made by image averaging the individual video frames. Likewise, using 3D face stimuli made from 3D scans of peoples' faces would help address the issue of whether we can abstract anything more from 3D images relative to 2D images as well as how averaging and generalisation may operate between different viewpoints. Since 3D scans can be morphed just like 2D images, extending the current research in this way should not be problematic.

Beyond Probing With Prototypes: the Potential of Face Adaptation

What other approaches could be used to probe the underlying nature of face identity representations? Studies of adaptation effects may have more potential to reveal important information that could deepen or change our understanding of how faces are represented.

Exposure to an obviously distorted exemplar of a face can influence subsequent judgements about that face. Most commonly, participants are asked to make judgements regarding the veridicality or normality of a face image after being exposed to a distorted image of the same person’s face earlier in the experiment. The basic finding is that exposure to an extreme distortion of a person’s face makes participants more likely to judge an exemplar that is slightly distorted (distorted in the same direction at a lower magnitude) as being a veridical image of the person. This adaptation effect has been demonstrated using famous faces on immediate tests of recognition and also following 5 min (Carbon & Leder, 2005) or 24 h delays (Carbon et al., 2007) between exposure and test. One interpretation of this effect is that it reflects the updating of existing face representations to include new information from our most recent encounters with a face.

However, many studies have demonstrated similar effects using unfamiliar faces (e.g. Rhodes, Jeffery, Clifford, & Leopold, 2007; Robbins, McKone, & Edwards, 2007; Webster & MacLin, 1999). These adaptation effects have not been interpreted as due to updating a stored memory representation for three good reasons. First, unfamiliar faces are assumed to have no “face” representation (e.g. a FRU in the Bruce and Young 1986 model). Webster and MacLin asked participants to judge whether they thought images of faces had been distorted or not. Despite having never seen the original images, participants’ judgements were strongly influenced by exposing them to a highly distorted face before they made their judgements. Second, the adaptor image that participants are exposed to need not be the same identity as the test face. The magnitude of the adaptation effect is similar when the exposure and test faces are the same identity (based on the same original image) compared to when they are different identities (although transfer is sometimes found to be slightly higher when adapting to the same identity or a similar face, this has not been consistently observed). Thus, even if participants were given an opportunity to form some sort of face representation based on a verdical image of an unfamiliar target, they would not be expected to update that representation to include an extremely distorted image of a different person’s face. Third, unfamiliar face adaptation effects are short lived, lasting only a few seconds (Rhodes et al., 2007). Adaptation effects from exposure to highly distorted unfamiliar faces are thought to be due to a more general recalibration of the perceptual system rather than adjustments to representations of individual faces.

It is not clear how adaptation effects found using familiar and unfamiliar faces differ. To what extent does adaptation to a distorted familiar face transfer to the perception of unfamiliar faces and vice versa? Studying face adaptation effects could reveal a lot about how we form representations of familiar faces and also how we maintain and update existing face representations. A closer examination of decay rates could also be illuminating. If strong evidence that identity specific adaptation can be observed after delays as long as one day or more, and if the strength of the effect decreases over time, then this would pose a serious challenge to a pure prototype account of face representation. If adaptation effects found using familiar faces reflect a permanent change to the underlying representation, then how could the change weaken without storing any information about this change or the exemplar that caused it? For example, one might assume that the abstraction mechanism makes an adjustment to a prototype representation every time a new exemplar is encountered, and the amount of adjustment is proportional to the number of exemplars the existing prototype has been based on. Even this assumption is beyond the scope of current prototype based models which maintain that the mean alone is sufficient to represent a person's face. Similarly, one could further assume that the contribution of each exemplar to its prototype is a function of its recency. Thus, some record of the exemplars (or the changes they produced) would need to be retained so that their contribution could be reduced from the prototype over time. Of course, a simpler explanation is that all exemplars are stored and prototypes are computed ad hoc at the time of retrieval with the most recent exemplars weighted more heavily. A hybrid prototype + exemplar model could also provide a reasonable account. Exposure to a distorted image of a known face could cause a permanent change to the prototype, but a trace of the distorted image would also be laid down and would decay over time. Over short delays, adaptation effects might reflect the joint activation of the prototype and the trace of the distorted image making the test image of the same identity appear slightly distorted in the opposite direction.

If it is assumed that this type of adaptation effect is due to changes in the representations of individual faces, then this methodology provides a framework for testing hypotheses about the properties of such representations and the process by which they are updated. To understand what adaptation effects reveal about the perception and recognition of identity in faces, future adaptation studies must examine the relationships between the degree of learning (i.e. familiarity), similarity between exposure and test images, and variables that modulate retinotopic adaptation (e.g. size, spatial location etc.). In addition to investigating how these variables affect the strength of adaptation, it is also important to map the rate of decay as a function of the duration of delay between the exposure and test phases.

Concluding Remarks

What does it mean to form a representation of an individual face in memory? To assume such a representation exists is common place among researchers of face recognition. Although to some this may not seem like a contentious assumption to make, I argue otherwise. It is often the most basic assumptions we make about the nature of memory that decide not only the methodologies we use to investigate it, but even the types of questions we, as cognitive psychologists, think are meaningful to ask. And as our assumptions drive the questions to which we seek answers, so too will our assumptions drive the answers that we find. The more we narrow our views, the more we will find the same answer again and again. Even when all evidence points in the same direction, do not forget to try looking down another; it just may take you even farther.

The interpretation of prototype effects as being indicative of the abstraction of category prototypes was popularized by Posner and Keele (1968, 1970), and is still very influential in current thinking about face recognition. Various forms of exemplar models have been proposed in the more general literature to account for this effect, but they have not received much attention in the area of face recognition specifically. To be fair, the prototype vs. exemplar distinction is not of obvious relevance to all of the questions face recognition researchers seek answers to. However, it is extremely important to any question concerning the nature of identity representation in faces. The case of the prototype in face recognition exemplifies this. What causes this effect is an interesting question in itself. I am not the first to point out that we really do not know (cf. Bruce, 1994; Bruce et al., 1991; Cabeza et al., 1999). Nonetheless, I will attempt to draw some conclusions from the present work based on assumptions I think are reasonable.

Information about the faces of many different individuals is stored in memory, and this information can be accessed when needed. The experiments showed that faces are stored or retrieved in a way that facilitates recognition of the central value of a set of varying exemplars of similar appearance. Moreover, it is the range of experienced variation that seems to be particularly important because changes in recognition responses due to the type of exemplar shown at test were most obviously characterized by a reduction in the perceived "oldness" of an unseen image that was outside the range of variation of the studied exemplars. This pattern was observed even when the unseen exemplars were highly similar to the seen exemplars. Because the range, not simply the average value, of variation seems to be most important, a purely prototype account of the prototype effect in face recognition is not plausible. At least over the time frame tested here, some information reflecting the range of experienced variation must be retained in memory. If prototypes were abstracted, they were not the only representations used to perform the task. On the other hand, evidence from the present work clearly demonstrated that information from multiple exemplars *was* being combined in memory, if not during encoding, then at the time of retrieval. It is in this way, that we can say a representation of a face is formed or built up out of varying instances. Assuming from the start some sort of pre-existing representation of an individual face, will lead to unfounded claims about how we represent different identities, how we become familiar with a face, and of course, how we make recognition judgements about pictures of faces presented in experiments.

It was surprising that no prototype effects were found using natural variations of identities. This result highlighted the possibility that the prototype effect in face recognition is not due to an averaging mechanism that operates specifically on exemplars of the same identity. Although such a mechanism may operate in this way under the right conditions, it is an oversimplification to describe it in this way. Exemplars of the same identity often are likely to be stored or retrieved in combination. But it is not *because* they belong to the same identity. Sometimes studying exemplars of two different identities may produce a prototype effect (Busey & Tunnicliff, 1999), and sometimes studying two exemplars of the same identity will not produce a prototype effect. The differences between the high and low quality exemplars that were used here leads one to surmise that superficial image properties, at least for unfamiliar faces, place heavy constraints on which images of shared identity will give rise to a prototype effect. Future research is needed to understand the role of identity in prototype recognition that goes above and beyond lower level image similarity. Furthermore, a deeper understanding of how memories formed from viewing the variation of faces in continuous motion are different than those formed from viewing static images may be the key to understanding why humans are so incredibly good at recognizing familiar faces.

Reference List

Angeli, A., Davidoff, J., & Valentine, T. (2001). Distinctiveness induces categorical perception of unfamiliar faces. Perception, 30, Supplement, 58.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(4), 629-654.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization (pp. 101-140). Cambridge: Cambridge University Press.

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. Memory & Cognition, 12(3), 219-228.

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

Barton, J. J. S., Keenan, J. P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. British Journal of Psychology, 92(3), 527-549.

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

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

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

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

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

Bradshaw, J. L., & Wallace, G. (1971). Models for the processing and identification of faces. Perception & Psychophysics, 9(5), 443-448.

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

Bruce, V. (1979). Searching for politicians: An information-processing approach to face recognition. The Quarterly Journal of Experimental Psychology, 31(3), 373-396.

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

Bruce, V. (1994). Stability from variation: The case of face recognition. The Quarterly Journal of Experimental Psychology a: Human Experimental Psychology, 47(1), 5-28.

Bruce, V., & Burton, M. (2002). Learning new faces. In M. Fahle & T. Poggio (Eds.), Perceptual Learning (pp. 317-334). Cambridge: MIT Press.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. Cognition, 38(2), 109-144.

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

Bruce, V., 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(4), 339-360.

Bruce, V., & Valentine, T. (1985). Identity priming in the recognition of familiar faces. British Journal of Psychology, 76 (3), 373-383.

Bruck, M., Cavanagh, P., & Ceci, S. J. (1991). Fortysomething: Recognizing faces at one's 25th reunion. Memory & Cognition, 19(3), 221-228.

Bruyer, R., Lafalize, A., & Distefano, M. (1991). Age decisions on familiar and unfamiliar faces. Behavioural Processes, 24, 21-35.

Bruyer, R., Leclere, S., & Quinet, P. (2004). Ethnic categorisation of faces is not independent of face identity. Perception , 33, 169-179.

Bruyer, R., Mejias, S., & Doublet, S. (2007). Effect of face familiarity on age decision. Acta Psychologica, 124(2), 159-176.

Burton, A. M., & Bruce, V. (1993). Naming faces and naming names: Exploring an interactive activation model of person recognition. Memory,1, 457-480.

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

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. Cognitive Psychology, 51(3), 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(3), 243-248.

Busey, T. A. (2001). Formal models of familiarity and memorability in face recognition. In M. J. Wenger & J. T. Townsend (Eds.), Computational, geometric, and process perspectives on facial cognition: Contexts and challenges (pp. 147-191). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5), 1210-1235.

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. Perception & Psychophysics, 65(8), 1296-1306.

Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: Extension and limits. Memory & Cognition, 27(1), 139-151.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. Psychological Science, 11(5), 429-433.

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. Visual Cognition, 8(2), 237-262.

Campanella, S., Hanoteau, C., Seron, X., Joassin, F., & Bruyer, R. (2003). Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique. Visual Cognition, 10(2), 129-156.

Carbon, C., & Leder, H. (2005). Face adaptation: Changing stable representations of familiar faces within minutes? Advances in Cognitive Psychology, 1(1), 1-7.

Carbon, C., Strobach, T., Langton, S. R. H., Harsanyi, G., Leder, H., & Kovacs, G. (2007). Adaptation effects of highly familiar faces: Immediate and long lasting. Memory & Cognition, 35(8), 1966-1976.

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

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. Journal of Experimental Psychology, 58, 17-22.

Deffenbacher, K. A., Vetter, T., Johanson, J., & O'Toole, A. J. (1998). Facial aging, attractiveness, and distinctiveness. Perception, 27(10), 1233-1243.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. Journal of Experimental Psychology: Human Learning & Memory, 7(6), 397-417.

Ellis, H., Shepherd, J., & Davies, G. (1975). An investigation of the use of the photo-fit technique for recalling faces. British Journal of Psychology, 66(1), 29-37.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. Perception, 8(4), 431-439.

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

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. Journal of Experimental Psychology, 90(1), 65-74.

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

Ganel, T., & Goshen-Gottstein, Y. (2002). The perceptual integrality of sex and identity of faces: Further evidence for the single-route hypothesis. Journal of Experimental Psychology: Human Perception and Performance, 28, 854-867.

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

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91(1), 1-67.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. Perception, 34(1), 77-86.

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

Goldman, D., & Homa, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. Journal of Experimental Psychology: Human Learning and Memory, 3(4), 375-385.

Gruppuso, V., Lindsay, D. S., & Masson, M. E. J. (2007). I'd know that face anywhere. Psychonomic Bulletin & Review, 14(6), 1085-1089.

Haig, N. D. (1986). Exploring recognition with interchanged facial features. Perception, 15(3), 235-247.

Hall, J. W., & Kozloff, E. E. (1970). False recognitions as a function of number of presentations. American Journal of Psychology, 83(2), 272-279.

Hall, J. W., & Kozloff, E. E. (1973). False recognitions of associates of converging versus repeated words. American Journal of Psychology, 86(1), 133-139.

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

Hannigan, S. L., & Reinitz, M. T. (2000). Influences of temporal factors on memory conjunction errors. Applied Cognitive Psychology, 14(4), 309-321.

Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. Science, 180(4091), 1194-1196.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning & Verbal Behavior, 16(3), 321-338.

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

Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. Journal of Experimental Psychology: Human Perception and Performance, 22(4), 986-1004.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. Psychological Review, 93(4), 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review, 95(4), 528-551.

Hintzman, D. L., & Curran, T. (1995). When encoding fails: Instructions, feedback, and registration without learning. Memory & Cognition, 23(2), 213-226.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. Perception, 23(1), 65-74.

Homa, D. (1978). Abstraction of ill-defined form. Journal of Experimental Psychology: Human Learning & Memory, 4(5), 407-416.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. Journal of Experimental Psychology, 101(1), 116-122.

Homa, D., & Cultice, J. C. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(1), 83-94.

Homa, D., Goldhardt, B., Burruel-Homa, L., & Smith, J. C. (1993). Influence of manipulated category knowledge on prototype classification and recognition. Memory & Cognition, 21(4), 529-538.

Homa, D., Smith, C., Macak, C., Johovich, J., & Osorio, D. (2001). Recognition of facial prototypes: The importance of categorical structure and degree of learning. Journal of Memory and Language, 44(3), 443-474.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. Journal of Experimental Psychology: Human Learning & Memory, 7(6), 418-439.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. Journal of Experimental Psychology: Human Learning & Memory, 2(3), 322-330.

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

Inn, D., Walden, K. J., & Solso, R. L. (1993). Facial prototype formation in children. Bulletin of the Psychonomic Society, 31(3), 197-200.

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). The effects of visual object priming on brain activation before and after recognition. Current Biology, 10(17), 1017-1024.

Jones, T. C., Bartlett, J. C., & Wade, K. A. (2006). Nonverbal conjunction errors in recognition memory: Support for familiarity but not for feature bundling. Journal of Memory and Language, 55(1), 138-155.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. Psychological Review, 93(2), 136-153.

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

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

Lander, K., & Metcalfe, S. (2007). The influence of positive and negative facial expressions on face familiarity. Memory, 15(1), 63-69.

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

Leder, H., & Carbon, C. (2006). Face-specific configural processing of relational information. British Journal of Psychology, 97, 19-29.

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

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53(2), 513-536.

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

Lee, K. J., & Perrett, D. I. (2000). Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments. Perception, 29(11), 1291-1312.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. Perception and Psychophysics, 62, 386-401.

Lewis, M. B., & Johnston, R. A. (1999). A unified account of the effects of caricaturing faces. Visual Cognition, 6(1), 1-41.

Liu, C. H., Seetzen, H., Buton, A. M., & Chaudhuri, A. (2003). Face recognition is robust with incongruent image resolution: Relationship to security video images. Journal of Experimental Psychology: Applied, 9(1), 33-41.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. Psychonomic Bulletin and Review, 1(4), 476-490.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. Canadian Journal of Experimental Psychology, 57(3), 203-220.

Matthews, M. (1978). Discrimination of identikit constructions of faces: Evidence for a dual processing strategy. Perception & Psychophysics, 23, 153-161.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. Psychological Review, 105(4), 724-760.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? Trends in Cognitive Sciences, 11(1), 8-15.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. Journal of Experimental Psychology: Human Perception and Performance, 27(3), 573-599.

Mckone, E., & Peh, Y. (2006). Memory conjunction errors for realistic faces are consistent with configural processing. Psychonomic Bulletin & Review, 13(1), 106-111.

Medin, D. L. (1989). Concepts and conceptual structure. American Psychologist, 44(12), 1469-1481.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. Psychological Review, 85(3), 207-238.

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

Megreya, A. M., & Burton, A. M. (2006b). Recognising faces seen alone or with others: When two heads are worse than one. Applied Cognitive Psychology, 20(7), 957-972.

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

Metcalfe, J. (1990). Composite holographic associative recall model (charm) and blended memories in eyewitness testimony. Journal of Experimental Psychology: General, 119(2), 145-160.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. Perception, 31(5), 553-566.

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76(2), 165-178.

Moscovitch, M., & Moscovitch, D. A. (2000). Super face-inversion effects for isolated internal or external features, and for fractured faces. Cognitive Neuropsychology, 17(1), 201-219.

Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. Psychological Science, 11(6), 492-496.

Nachson, I., Moscovitch, M., & Umilta, C. (1995). The contribution of external and internal features to the matching of unfamiliar faces. Psychological Research, 58, 31-37.

Neumann, P. G. (1974). An attribute frequency model for the abstraction of prototypes. Memory & Cognition, 2(2), 241-248.

Neumann, P. G. (1977). Visual prototype formation with discontinuous representation of dimensions of variability. Memory & Cognition, 5(2), 187-197.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(1), 104-114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115(1), 39-57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14 (4), 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(1), 3-27.

O'Donnell, C. (2003). Investigating face familiarisation. Unpublished doctoral dissertation, University of Stirling, United Kingdom.

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

Omohundro, J. (1981). Recognition vs. classification of ill-defined category exemplars. Memory & Cognition, 9(3), 324-331.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. Nature Reviews Neuroscience, 5(4), 291-303.

Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. The

Quarterly Journal of Experimental Psychology a: Human Experimental Psychology, 54(1), 197-235.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. Journal of Memory and Language, 35(2), 261-285.

Phillips, R. J., & Rawles, R. E. (1979). Recognition of upright and inverted faces: A correlational study. Perception, 8(5), 577-583.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77(3), 353-363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. Journal of Experimental Psychology, 83 (2), 304-308.

Reed, S. K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3(3), 382-407.

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. Memory & Cognition, 20(1), 1-11.

Reinitz, M. T., & Hannigan, S. L. (2001). Effects of simultaneous stimulus presentation and attention switching on memory conjunction errors. Journal of Memory and Language, 44(2), 206-219.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(1), 161-168.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? Cognition, 47(1), 25-57.

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

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

Rhodes, G., Jeffery, L., Clifford, C. W. G., & Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. Vision Research, 47(17), 2291-2296.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. Psychological Science, 14, 558-566.

Rhodes, G., & Moody, J. (1990). Memory representations of unfamiliar faces: Coding of distinctive information. New Zealand Journal of Psychology, 19(2), 70-78.

Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adapter position effects and neural models. Journal of Experimental Psychology: Human Perception and Performance, 33(3), 570-592.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 803-814.

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), Studies in cross-cultural psychology (Vol. 1, pp. 1-49). London: Academic Press.

Rossion, B. (2002). Is sex categorization from faces really parallel to face recognition? Visual Cognition, 9, 1003-1020.

Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., & Crommelinck, M. (2001). How does the brain discriminate familiar and unfamiliar faces? A pet study of face categorical perception. Journal of Cognitive Neuroscience, 13(7), 1019-1034.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. Cognitive Psychology, 15(3), 346-378.

Ryu, J., & Chaudhuri, A. (2006). Representations of familiar and unfamiliar faces as revealed by viewpoint-aftereffects. Vision Research, 46(23), 4059-4063.

Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. Neuropsychology, 11(3), 331-342.

Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. Journal of Memory and Language, 35(2), 319-334.

Schooler, J. W., & Tanaka, J. W. (1991). Composites, compromises, and charm: What is the evidence for blend memory representations? Journal of Experimental Psychology: General, 120 (1), 96-100.

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

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. Journal of Experimental Psychology: Human Perception and Performance, 24(6), 1748-1765.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. Journal of Experimental Psychology: Human Perception and Performance, 22(4), 904-915.

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

Sergent, J. (1986). Microgenesis in face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.) Aspects of face processing (pp. 17-73). Boston, US: Martinus Injhoff Publishers.

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. Journal of Experimental Psychology, 55(6), 509-523.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. Science, 237(4820), 1317-1323.

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(2), 267-287.

Smith, E. E., & Nielsen, G. D. (1970). Representations and retrieval processes in short-term memory: Recognition and recall of faces. Journal of Experimental Psychology, 85, 397-405.

Solso, R. L., Heck, M., & Mearns, C. (1993). Prototype formation in very short-term memory. Bulletin of the Psychonomic Society, 31(3), 185-188.

Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. British Journal of Psychology, 72(4), 499-503.

Solso, R. L., & Raynis, S. A. (1979). Prototype formation from imaged, kinesthetically, and visually presented geometric figures. Journal of Experimental Psychology: Human Perception and Performance, 5(4), 701-712.

Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. British Journal of Psychology, 89, 39-57.

Stevenage, S. V., Lee, E. A., & Donnelly, N. (2005). The role of familiarity in a face classification task using thatcherized faces. The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 58(6), 1103-1118.

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(5), 760-766.

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

Thompson, P. (1980). Margaret Thatcher: A new illusion. Perception, 9(4), 483-484.

Tong, F., & . Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. Journal of Experimental Psychology: Human Perception & Performance, 25, 1016-1035.

Tulving, E. (1985). How many memory systems are there? American Psychologist, 40(4), 385-398.

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

Valentine, T. (2001). Face-space models of face recognition. In M. J. Wenger & J. T. Townsend (Eds.), Computational, geometric, and process perspectives on facial cognition: Contexts and challenges (pp. 83-113). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. Visual Cognition, 15(4), 420-467.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. Memory & Cognition, 20(3), 291-302.

Warrington, E. K., & James, M. (1967). An experimental investigation of facial recognition in patients with unilateral cerebral lesions. Cortex, 3(3), 317-326.

Webster, M. A., & Maclin, O. H. (1999). Figural aftereffects in the perception of faces. Psychonomic Bulletin & Review, 6(4), 647-653.

Whittlesea, B. W. (1987). Preservation of specific experiences in the representation of general knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(1), 3-17.

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. Psychological Science, 17(9), 799-806.

Wolberg, G. (1990). Digital image warping. Los Alamitos, CA: IEEE Computer Science Society Press.

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

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. Perception, 14, 737-746.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. Perception, 16(6), 747-759.

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old-new recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(4), 1022-1041.

Appendix A

Table A1. Post hoc Bonferroni paired-samples *t* tests comparing mean confidence ratings among prototype, seen, and unseen exemplar types in Experiment 1.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Prototype - Seen	.57*	.87	3.21	23	.012
Pair 2	Prototype - Unseen	.97*	1.02	4.68	23	.000
Pair 3	Seen - Unseen	.40*	.70	2.80	23	.031

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table A2. Post hoc Bonferroni paired-samples *t* tests comparing mean confidence ratings among prototype, seen, and unseen exemplar types in Experiment 2.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Prototype - Seen	.51*	.64	3.92	23	.002
Pair 2	Prototype - Unseen	1.01*	.88	5.62	23	.000
Pair 3	Seen - Unseen	.49*	.71	3.42	23	.007

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table A3. Post hoc Bonferroni paired-samples *t* tests comparing mean confidence ratings among conditions in the Distance x Exemplar interaction in Experiment 2.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Far Prototype - Far Seen	.67	1.05	3.10	23	.075
Pair 2	Far Prototype - Far Unseen	.20	.97	1.00	23	1.000
Pair 3	Far Seen - Far Unseen	-.47	1.07	-2.15	23	.629
Pair 4	Near Prototype - Near Seen	.36	.74	2.34	23	.424
Pair 5	Near Prototype - Near Unseen	1.81*	1.39	6.38	23	.000
Pair 6	Near Seen - Near Unseen	1.46*	1.09	6.57	23	.000

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni (this adjustment was made based on all 16 possible pairwise comparisons of Distance x Exemplar Type although only 6 are included in the table).

Appendix B

Table B1. Post hoc Bonferroni paired-samples *t* tests comparing mean confidence ratings among prototype, local average, seen, and unseen exemplar types in Experiment 4.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Prototype - Local Average	.43*	.68	4.40	47	.000
Pair 2	Prototype - Seen Exemplar	.98*	.97	6.99	47	.000
Pair 3	Prototype - Unseen	3.28*	1.21	18.85	47	.000
Pair 4	Seen Exemplar - Local Average	-.55*	1.03	-3.66	47	.004
Pair 5	Local Average - Unseen	2.85*	1.23	16.07	47	.000
Pair 6	Seen Exemplar - Unseen	2.30*	1.23	12.98	47	.000

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table B2. Post hoc Bonferroni paired-samples *t* tests comparing mean confidence ratings among faces presented with the same, different, or no context in Experiment 4.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Same - Different	.34*	.82	2.88	47	.020
Pair 2	Same - None	.44*	.72	4.20	47	.000
Pair 3	Different - None	.10	.52	1.35	47	.565

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table B3. Post hoc Bonferroni paired-samples *t* tests comparing mean confidence ratings among prototype, local average, and unseen exemplar types in Experiment 6.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Prototype - Local Average	.20	.68	2.06	47	.136
Pair 2	Prototype - Unseen Exemplar	2.69*	1.10	16.86	47	.000
Pair 3	Local Average - Unseen Exemplar	2.49*	1.09	15.80	47	.000

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table B4. Post hoc Bonferroni paired-samples *t* tests comparing mean confidence ratings among prototype, local average, seen, and unseen exemplar types in Experiment 7.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Prototype - Local Average	.70	1.62	2.46	31	.118
Pair 2	Prototype - Seen Exemplar	.02	1.14	.10	31	1.000
Pair 3	Prototype - Unseen Exemplar	1.80*	1.53	6.64	31	.000
Pair 4	Local Average - Seen Exemplar	-.68*	1.31	-2.96	31	.035
Pair 5	Local Average - Unseen Exemplar	1.10*	1.55	4.01	31	.002
Pair 6	Seen Exemplar - Unseen Exemplar	1.78*	1.39	7.23	31	.000

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Appendix C

Table C1. Mean confidence ratings to targets and distractors on the first and third (final) presentation of an officer's face varying across image type in Experiment 8.

		Presentation			
		First		Third	
	Image Type	Mean	Std. Deviation	Mean	Std. Deviation
Target	Morph	4.95	1.10	5.55	.72
	High	5.28	1.00	5.36	.90
	Low	5.30	.76	5.35	.86
Distractor*	Morph	2.56	.96	2.47	.97
	High	2.41	.78	2.34	.96
	Low	2.44	.92	2.55	1.00

*. Distractor faces were only presented once each; thus, distractor summary statistics are based on the first and third block of trials.

Table C2. Mean proportion of OLD responses to targets and distractors on the first and third (final) presentation of an officer's face varying across image type in Experiment 8.

		Presentation			
		First		Third	
	Image Type	Mean	Std. Deviation	Mean	Std. Deviation
Target	Morph	.79	.31	.94	.17
	High	.88	.22	.90	.23
	Low	.89	.21	.90	.21
Distractor*	Morph	.24	.24	.20	.25
	High	.22	.20	.17	.23
	Low	.20	.25	.22	.26

*. Distractor faces were only presented once each; thus, distractor summary statistics are based on the first and third block of trials.

Table C3. Post hoc Bonferroni paired-samples *t* tests of the Familiarity x Test Image interaction on the mean proportion of OLD responses in Experiment 9.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Morph (Low Familarity) - High (Low Familiarity)	.04	.21	1.38	47	1.000
Pair 2	Morph (Low Familiarity) - Low (Low Familiarity)	.05	.21	1.73	47	1.000
Pair 3	High (Low Familiarity) - Low (Low Familiarity)	.01	.23	.32	47	1.000
Pair 4	Morph (High Familiarity) - High (High Familiarity)	-.07*	.14	-3.69	47	.009
Pair 5	Morph (High Familiarity) - Low (High Familiarity)	-.01	.18	-.53	47	1.000
Pair 6	High (High Familiarity) - Low (High Familiarity)	.06*	.13	3.13	47	.000

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni (this adjustment was made based on all 16 possible pairwise comparisons of Familiarity x Test Image although only 6 are included in the table).

Table C4. Post hoc Bonferroni paired-samples *t* tests on the effect of the type of Image (Morph, High, and Low quality) on the proportion of OLD responses to distractor faces in Experiment 10.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Morph - High	-.02	.18	-.69	47	1.000
Pair 2	Morph - Low	-.11*	.21	-3.69	47	.002
Pair 3	High - Low	-.09*	.21	-3.03	47	.012

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table C5. Post hoc Bonferroni paired-samples *t* tests on the main effect of test image (Morph, High, and Low quality) on *d*’ scores in Experiment 10.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Morph - High	.09	1.01	.63	47	1.000
Pair 2	Morph - Low	.54*	.96	3.94	47	.001
Pair 3	High - Low	.45*	1.05	2.97	47	.014

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table C6. Post hoc Bonferroni paired-samples *t* tests on the main effect of test image (Morph, High, and Low quality) on *d*’ scores for the first presentation of an officer’s face in Experiment 10.

		Paired Differences				
		Mean	Std. Deviation	t	df	Sig. (2-tailed)[a]
Pair 1	Morph - High	-.44	2.18	-1.38	47	.519
Pair 2	Morph - Low	.35	2.00	1.23	47	.677
Pair 3	High - Low	.79*	1.78	3.07	47	.011

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.