The effect of comparison on the perceived similarity of complex visual stimuli

Paula Christiane Engelbrecht

Thesis submitted to the University of London for the degree of Doctor of Philosophy

2018

Department of Psychology Goldsmiths, University of London

Declaration

I, Paula Engelbrecht, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed: ____________________________________ Date: 03/09/2018

Acknowledgements

This thesis is dedicated to my two supervisors Prof Tim Valentine and Dr Karina Linnell. I am grateful to both for allowing me to transfer my research to their supervision, and for providing me with invaluable support, advice and feedback.

Abstract

Comparison has been shown to decrease the perceived similarity of artificial face stimuli that are difficult to discriminate (Mundy, Honey, & Dwyer, 2007). This thesis presents seven experiments that examined the effect of comparison on the perceived similarity of a wider range of unfamiliar face stimuli. Participants were asked to compare two faces before either rating their perceived similarity, or deciding whether they are the same person. In the first five experiments participants were shown face pairs that ranged in phenotypic similarity—the degree to which the two faces look alike. With the exception of highly similar face morphs comparison was found to increase the perceived similarity of both phenotypically similar and dissimilar face pairs, relative to a no-comparison control. This finding suggests that for most naturally occurring face stimuli, comparison results in an increase in perceived similarity. In the last two experiments the quality of one of the stimuli in each pair was degraded to simulate the effects of poor quality video footage. A comparison-related decrease in perceived similarity was found in both experiments. This finding suggests that pictorial differences between face stimuli—including differences in image quality, camera distance and lighting, variations in pose and facial expression, and the presence of disguises such as hats and sunglasses—play an important role in mediating the effect of comparison on perceived similarity.

Table of Contents

Figures

Tables

Chapter 1: Research rationale, background, and approach

1.1 Introduction and overview

1.1.1 Research rationale

Mistaken eyewitness identifications contributed towards more than 70% of the wrongful convictions that the US Innocence Project helped overturn through DNA evidence (The Innocence Project, n.d.). One factor that is likely to have contributed to some of these mistaken identifications is the presence of a weapon, which can interfere with witnesses' ability to process the physical appearance of the assailant (Kramer, Buckhout, & Eugenio, 1990; Loftus, Loftus, & Messo, 1987; Pickel, Ross, & Truelove, 2006). Events that occur after a crime has taken place can also influence the accuracy of eyewitness testimonies. For example, it has been shown that witnesses' memories of a crime can be altered by information they are exposed to during criminal investigations, including the wording that is used during questioning (Johnson & Raye, 1998; Loftus, 2005; Loftus & Hoffman, 1989). Research on the limitations of human cognition and their impact on the accuracy of eyewitness testimonies has informed best practices in criminal procedures. For example, the use of standardised witness instructions that state that the perpetrator may or may not be present in a police identity parade is recommended to avoid biasing witnesses towards making a positive identification (National Academy of Sciences, 2014).

Inaccurate eyewitness testimonies are not the only factor that can result in miscarriages of justice. The case of Christopher Seddon, who was wrongfully arrested on suspicion of shoplifting highlights that mistaken identifications can occur even when visual representations of both the suspect and perpetrator are available for comparison. Upon exiting a supermarket, Christopher Seddon was seized and handcuffed by police. He had been wrongly identified by a security guard based on a still from closed-circuit television (CCTV) footage stored on the guard's mobile phone. It cost Christopher Seddon £9,600 in legal fees and expert reports to clear his name (Brignall, 2016). Empirical studies confirm that the process of deciding whether a person present matches a person shown in a photo or video is error prone (Davis & Valentine, 2009; Kemp, Towell, & Pike, 1997; Megreya & Burton, 2008; White, Kemp, Jenkins, Matheson, & Burton, 2014). The photo of the shoplifter stored on the security guard's phone was distorted through compression. It can therefore be argued that mistaken identifications only occur when poor quality images are used. However, empirical evidence demonstrates that the process of matching two unfamiliar faces from high quality photos or video is susceptible to error (Bruce et al., 1999; Henderson, Bruce, & Burton, 2001). In addition to wrongful arrests, misidentifications have also lead to wrongful deportations and detentions. For example, Atterbell Maplanka was barred from re-entering the UK after attending his mother's funeral in Zimbabwe (Corcoran, 2015). He was detained because his face shape had changed due to weight gain, and he no longer closely resembled his passport photo. Empirical evidence confirms that images of the same unfamiliar face captured under different viewing conditions are easily mistaken for different individuals (Jenkins, White, Van Montfort, & Burton, 2011; see also Andrews, Jenkins, Cursiter, & Burton, 2015).

The limitations of unfamiliar face matching performance outlined above "are not widely appreciated even within cognitive psychology, and seldom penetrate cognate fields in engineering and law" (Jenkins & Burton, 2011, p.1673). It is therefore perhaps unsurprising that the underlying cognitive factors that drive face matching performance are not well understood (Bobak, Hancock, & Bate, 2016). One factor that may affect performance on face matching tasks is prior comparison of the to-be-matched stimuli. Comparison—the process of detecting and assessing similarities and differences—is central to cognition, and has been widely studied. It has been invoked in theories of memory retrieval (Hintzman, 1986 experiment 1b; Raaijmakers & Shiffrin, 1981), language processing (Bailey & Hahn, 2001; Hahn & Bailey, 2005), object recognition (Tarr & Gauthier, 1998) categorisation (Hampton, 1995; Lamberts, 2000; Medin & Schaffer, 1978; Nosofsky, 1986), decision making (Medin, Goldstone, & Markman, 1995), problem solving (Bassok, 1990; Holyoak & Koh, 1987; Novick, 1988, 1990, Ross, 1987, 1989), induction (Osherson, Smith, Wilkie, López, & Shafir, 1990; Rips, 1975), and social cognition (Smith & Zárate, 1992). Yet, even though comparison plays a crucial part in a variety of cognitive processes, the role of comparison in interpreting naturalistic visual stimuli such as faces remains relatively unexplored. To close this gap this thesis explores the effects of similarity-focused and difference-focused comparisons on the perceived similarity of unfamiliar face pairs that vary in their degree of resemblance.

1.1.2 Research aims

The overarching goal of this thesis is to understand the effect of comparison on the perceived similarity of unfamiliar faces. Achieving this aim is of both applied and theoretical interest. Various contextual factors have been shown to affect face matching accuracy, including poor sleep (Beattie, Walsh, McLaren, Biello, & White, 2016), anxiety levels (Attwood, Penton-Voak, Burton, & Munafò, 2013), and task duration (Alenezi & Bindemann, 2013; Alenezi, Bindemann, Fysh, & Johnston, 2015). A study in which participants were asked to match 1000 face pairs found that, whereas trial repetition results in a small improvement in accuracy on same trials, accuracy on mismatch trials decreases (Alenezi & Bindemann, 2013). The authors interpret this finding as suggesting "that observers find it increasingly difficult to tell the faces of different people apart or, in other words, that different facial identities are increasingly looking the same with continuous exposure to this task" (Alenezi & Bindemann, 2013, p. 750). There is some evidence to suggest that acute anxiety has the opposite effect on face matching performance (Attwood et al., 2013). Manipulating acute anxiety through the administration of CO_2 has been shown to reduce the number of correct 'same' responses (Attwood et al., 2013). The anxiety manipulation had no effect on the participants' ability to correctly identify mismatches. These findings suggest that anxiety might reduce the perceived similarity of face pairs. Like task duration and anxiety, the history of prior comparisons to which a given face pair has been subjected may also affect perceived similarity. To improve face matching procedures in applied settings it is important to know whether, and under what circumstances, comparison affects performance.

Structural alignment theory is the dominant theoretical account of comparison. The theory has been developed and tested using a range of stimuli including word pairs (Markman & Gentner, 1993a, 1993b), category descriptions (Smith & Gentner, 2014) line drawings of causal and relational scenes (Gentner & Markman, 1997; Markman, 1996; Markman & Gentner, 1993c), and drawings of objects that belong to different categories (Boroditsky, 2007; Gentner & Namy, 1999). The theory predicts that comparison should increase the perceived similarity of objects that share a common structure because the commonalities highlighted by the alignment process are weighted more highly than the differences (Markman & Gentner, 1993b, 1996). Yet, the finding that the opportunity to compare stimuli that are difficult to discriminate improves performance on discrimination tasks (Mundy, Honey, & Dwyer, 2007, 2009) suggests that this prediction might be limited to objects that are easy to tell apart.

1.1.2.1 The role of stimulus similarity

The first aim of this thesis is to investigate the role of stimulus similarity in mediating the effect of comparison on the perceived similarity of unfamiliar faces. A study investigating the effects of comparison on the perceived similarity of both familiar and novel objects has found that comparison increases the perceived similarity of similar objects (Boroditsky, 2007). However, evidence from the discrimination learning literature shows that the opposite holds true for complex visual stimuli that are difficult to tell apart, including faces (Mundy et al., 2007, 2014; Mundy, Honey, & Dwyer, 2009). The face stimuli used in discrimination learning studies have been designed to be very difficult to discriminate (Mundy et al., 2007, 2014). They were created by morphing two similar faces and then selecting face pairs from the set of intermediate faces. Whilst these highly artificial stimuli are well-suited to their use in perceptual learning studies, they are not representative of the types of face stimuli that are compared in applied settings, such as photo IDs and CCTV footage (Bindemann, Attard, Leach, & Johnston, 2013; Bindemann & Sandford, 2011; Davis & Valentine, 2009; Megreya, Sandford, & Burton, 2013).

It therefore remains to be established whether the findings from the discrimination learning literature extend to the kinds of nonhomogeneous face stimuli that can result in mistaken identifications in applied settings. It has been argued that unfamiliar face matching performance is heavily reliant on image-comparison techniques rather than specialist face-processing (Burton, 2013; Jenkins & Burton, 2011; Jenkins et al., 2011). It is therefore unsurprising that the likelihood of matching two unfamiliar faces successfully is dependent on surface level similarities and differences including lighting, camera angle, facial expression and orientation (Hancock, Bruce, & Mike Burton, 2000; Jenkins et al., 2011; Megreya & Burton, 2006). It follows that the use of artificially homogenised face stimuli, such as those deployed in the discrimination learning literature, might reduce the generalisability of experimental findings on face matching performance (Burton, 2013). Understanding whether comparison increases or decreases the perceived similarity of faces captured in naturalistic, and often suboptimal settings (e.g. due to the angle from which CCTV footage is recorded) can inform how these comparisons should be conducted in forensic settings. The role of pictorial and phenotypic similarities and differences in mediating the effect of comparison on the perceived similarity of faces will be further discussed in section 1.2.1.

1.1.2.2 The role of the comparison task

The second aim of this thesis is to understand how the nature of the comparison task influences perceived similarity. There are many task specific and contextual factors that have the potential to influence the outcome of a comparison, including the time available to complete the comparison, and the availability of contextual information that might introduce bias (Dror, Péron, Hind, & Charlton, 2005; Kukucka & Kassin, 2014; Osborne, Woods, Kieser, & Zajac, 2014; Searston & Tangen, 2015; but see Kerstholt et al., 2010). An understanding of the effects of the comparison task on the perceived similarity of unfamiliar faces has important practical implications. It can inform best practices in applied contexts that require face matching, including criminal investigations and border control. Before considering mediating factors such as the influence of prior knowledge, it is important to first understand whether variations in the comparison task itself affect perceived similarity.

This thesis focuses on three aspects of the comparison task. The first aspect is whether the comparison involves finding differences or commonalities. Morphological comparison analysis is a forensic image comparison technique that is commonly used by forensic experts in the UK (Wilkinson & Evans, 2009). "With morphological comparison analysis facial features are classified into discrete categories, providing an indication of whether these are similar across images" (Davis, Valentine, & Wilkinson, 2012, p. 136). The outcomes of a morphological comparisons, and of face comparisons more generally, might be affected by whether the person who performs the comparison focusses on finding commonalities or differences.

Studies on the acquisition of perceptual expertise suggest that similarity- and difference-focused comparisons may have different effects on perceived similarity (Tanaka, Curran, & Sheinberg, 2005; Tanaka & Pierce, 2009). In one of these studies participants were required to categorise images of different bird species over six days of training. Half of the participants performed categorisations at the basic (family) level (e.g. "owl"), the other half were trained to categorise exemplars at the subordinate (species) level (e.g. "barn owl"). The study found that subordinate-level training improved discrimination accuracy for new exemplars of bird species that were shown during training, and for novel bird species. These findings have been interpreted as showing that subordinate-level training tunes participants' attention towards fine-grained visual cues that distinguish different species of the same family (Tanaka et al., 2005). A similar pattern of results was obtained in a study in which participants were trained to discriminate faces that belong to a different race from their own (Tanaka & Pierce, 2009). The study found that individuation training (operationalised as learning face names) resulted in better performance on an old-new recognition test than classification training (operationalised as labelling the racial category of a face). Based on the above findings it can be argued that other manipulations that direct attention towards individuating features, such as difference-focused comparisons, may reduce the perceived similarity of the compared stimuli. Similarly, manipulations that direct attention towards commonalities, such as similarity-focused comparisons, may increase the perceived similarity of the compared stimuli. This thesis will explore whether similarity-focused and difference-focused comparisons have different effects on the perceived similarity of unfamiliar faces.

A second aspect of the comparison task which may influence perceived similarity is the presence or absence of time-pressure. Whereas stimulus presentations were timed in studies that have found comparison-related decreases in perceived similarity (Mundy et al., 2014, 2007; Mundy, Honey, & Dwyer, 2009), no strict time constraints were imposed in studies that have found increases in perceived similarity (Boroditsky, 2007; Hassin, 2001). These differences might play an important role in determining the direction of the effect of comparison on perceived similarity. A third difference between the two paradigms is that, whereas participants in studies that have found comparison related increases in similarity had to externalise the comparison (either by writing down shared and distinctive features, or by rating similarity), participants in discrimination learning studies did not. The role of time constraints, and task explicitness will be further discussed in sections 1.2.2 and 1.2.3.

1.1.3 Summary and overview of experiments

There is evidence to suggest that performance on unfamiliar face matching tasks can be error prone, even when high-quality video stills and photographs are used (Bruce et al., 1999; Henderson et al., 2001). It is important to understand how common processes—such as comparison—influence the likelihood that correct identifications are made in these contexts. According to structural alignment theory, comparison will increase the perceived similarity of the compared faces (Markman & Gentner, 1993b). According to theoretical accounts that have been developed to explain discrimination learning, such as habituation, comparison will result in a decrease in perceived similarity (Honey & Bateson, 1996; Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009). Studies which have found comparison related increases, and studies which have found comparison related decreases in perceived similarity, have deployed different stimuli and comparison tasks. This thesis examines the role of stimulus similarity and comparison task in mediating the effect of comparison on the perceived similarity of unfamiliar faces. Understanding the relationship between these variables can inform both our understanding of face matching performance in applied settings, and contribute towards the development of our theoretical understanding of comparison.

This thesis presents seven experiments on the effect of comparison on the perceived similarity of unfamiliar faces. In the first four experiments participants were asked to compare pairs of faces and rate the perceived similarity between them. There were two comparison tasks: a similarity-focussed comparison task (in which participants were asked to write down three similarities between two faces), and a difference-focussed comparison task (in which participants were asked to write down three differences). After completion of a comparison task participants were asked to rate the similarity of the compared faces on a rating scale. The effect of comparison on perceived similarity was measured by comparing similarity ratings obtained after completion of a comparison task to a similarity rating for a face pair that was not subjected to a prior comparison. In the fifth experiment participants were shown face pairs for a fixed duration of time, before rating the similarity of the pair. They were asked to compare the faces for the duration of the presentation but were not required to write down any information about them. In Experiments 6 and 7 a different measure of perceived similarity was used. Instead of rating the perceived similarity between faces, participants were asked to decide whether two faces were the same person or not. Perceived similarity was measured by comparing the number of 'same' and 'different' responses obtained after completing a comparison task, to the number of 'same' and 'different' responses for a no-comparison control pair. The theoretical motivation for these experiments is further developed in section 1.2.

1.2 Background

The overarching goal of this thesis is to understand the effect of comparison on the perceived similarity of unfamiliar faces. A natural starting point is to review the existing research on the relationship between comparison and perceived similarity, and assess how it extrapolates to faces. This body of research has not generated consistent results. Some experiments have shown that comparison increases the perceived similarity of visual stimuli (Boroditsky, 2007), others have found the opposite (Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009). One possible explanation for this discrepancy is the relative similarity of the compared stimuli.

Comparison has been shown to both decrease the perceived similarity of stimuli that are difficult to tell apart (Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009), and to increase the perceived similarity of stimuli that are similar, but easy to discriminate (Boroditsky, 2007). It is not clear which of these sets of findings is the most applicable to unfamiliar faces. Studies that have found comparison-related decreases in perceived similarity have used face stimuli that are homogenous and difficult to discriminate (see Figure 3 for examples). This begs the question—can the finding that comparison decreases the perceived similarity of highly similar faces be extrapolated to naturalistic face stimuli? The stimuli for which comparison-related increases have been observed are generally more heterogeneous than faces, and too dissimilar to be mistaken for each other (see Figures 4 and 5 for examples). Again, this begs the question whether the finding that comparison increases the perceived similarity of similar visual stimuli can be extrapolated to faces. Section 1.2.1 reviews the existing literature on the effects of comparison on the perceived similarity of visual stimuli, and evaluates its applicability to unfamiliar faces.

To understand the effect of comparison on the perceived similarity of unfamiliar faces it is crucial to know how variations in the task context affect the relationship between comparison and perceived similarity. One contextual factor known to vary in applied settings is the time available to make a similarity judgement. Face comparisons are routinely undertaken in a range of professional settings. Some of these, such as border control, will require relatively speedy decisions. In other contexts, such as when deciding whether a suspect matches video footage of the perpetrator, the investigating officer may be able to deliberate for longer. These differences are reflected in the experimental paradigms that have been deployed in studies that assessed the role of comparison in similarity. Whereas studies that deploy direct measures of similarity (such as asking participants to rate the similarity of two stimuli on a rating scale) do not impose time limits on the comparison task, or response deadlines, studies that deploy indirect measures of similarity (such as response times and error rates) limit both the time available to compare the stimuli, and the time available to make a response. The potential impact of responding under time constraints on perceived similarity is discussed in section 1.2.3. The role of imposing time limits on the comparison task is explored in section 1.2.2.

Another task variable that might influence the effect of comparison on perceived similarity is the focus of the comparison process. For example, a person who is focused on finding differences between two stimuli might rate them as less similar than a person who is focused on finding similarities. Section 1.2.2 discusses the different ways in which comparison has been operationalized and the potential role that these differences might play in mediating the effect of comparison on perceived similarity.

Several theories have been developed to explain comparison. Structural alignment theory conceptualises comparison as an iterative mapping process that results in an alignment between two representations (Markman & Gentner, 1993b, 1993a, 1993c). According to representational distortion theory comparison is a process of distorting one stimulus representation into another (Hahn, Chater, & Richardson, 2003; Hodgetts, Hahn, & Chater, 2009). Finally, according to the habituation account, perceived similarity is the product of an attentional bias towards unique stimulus attributes, shaped by habituation to common stimulus elements during comparison (Mundy et al., 2007). The different theories of comparison, and their relative merits as accounts for the effects of comparison on the perceived similarity of visual stimuli, are reviewed in section 1.2.4.

1.2.1 The effect of comparison on the perceived similarity of visual stimuli

One of the aims of this thesis is to understand whether stimulus similarity plays a role in mediating the effects of comparison on the perceived similarity of unfamiliar faces. To further this aim, this section reviews the existing body of research on the relationship between stimulus similarity and perceptual similarity, and evaluates the extent to which the findings can be extrapolated to faces. The visual stimuli that have been deployed to study the effects of comparison on perceived similarity vary along a range of dimensions. To evaluate the relevance of existing research it is important to understand which of these dimensions are most likely to impact on the perceived similarity of unfamiliar faces. The next section therefore outlines the stimulus attributes that affect the perceived similarity of unfamiliar face pairs.

1.2.1.1 Stimulus determinants of face similarity

Human faces vary widely in the degree to which they resemble each other. It has been argued that this phenotypic diversity of faces has evolved to support facial recognition (Sheehan & Nachman, 2014). The faces of two people of the same gender and race, and of similar age, picked at random, might have little in common in terms of their overall face shape, the shape and configuration of their facial features, and the colour and texture of their skin. On the other hand, two monozygotic twins might resemble each other so closely, that even they have difficulties telling themselves apart from their twin (Martini, Bufalari, Stazi, & Aglioti, 2015). This variation is illustrated in Figure 1. The top row depicts four Caucasian males of a similar age, who vary in the degree to which they resemble each other. The middle row depicts three brothers, whose faces share several commonalities, for example, a similarly shaped nose, but look dissimilar in other respects. The bottom row depicts monozygotic twins. The two brothers bear a very close resemblance, and are only differentiable by their hairstyles and facial hair.

In addition to facial resemblance, pictorial differences also affect the perceived similarity of unfamiliar faces. Some of the image attributes that have been shown to affect face recognition and matching performance are lighting (Hill & Bruce, 1996; Hong Liu, Collin, Burton, & Chaudhuri, 1999; Johnston, Hill, & Carman, 1992), differences in camera angle (Jenkins & Burton, 2011), differences in pose (Bruce, 1982; Bruce, Valentine, & Baddeley, 1987; Hill & Bruce, 1996; Krouse, 1981), changes in facial expression (Bruce, 1982), hairstyle (Righi, Peissig, & Tarr, 2012), and the presence of sunglasses (Hockley, Hemsworth, & Consoli, 1999; Mansour et al., 2012; Vokey & Hockley, 2012), prescription glasses (Righi et al., 2012), and headwear (Lee, Wilkinson, Memon, & Houston, 2009; Mansour et al., 2012; Wilkinson & Evans, 2009). The range of pictorial differences that can impact on the perceived similarity of face stimuli is illustrated in Figure 2.

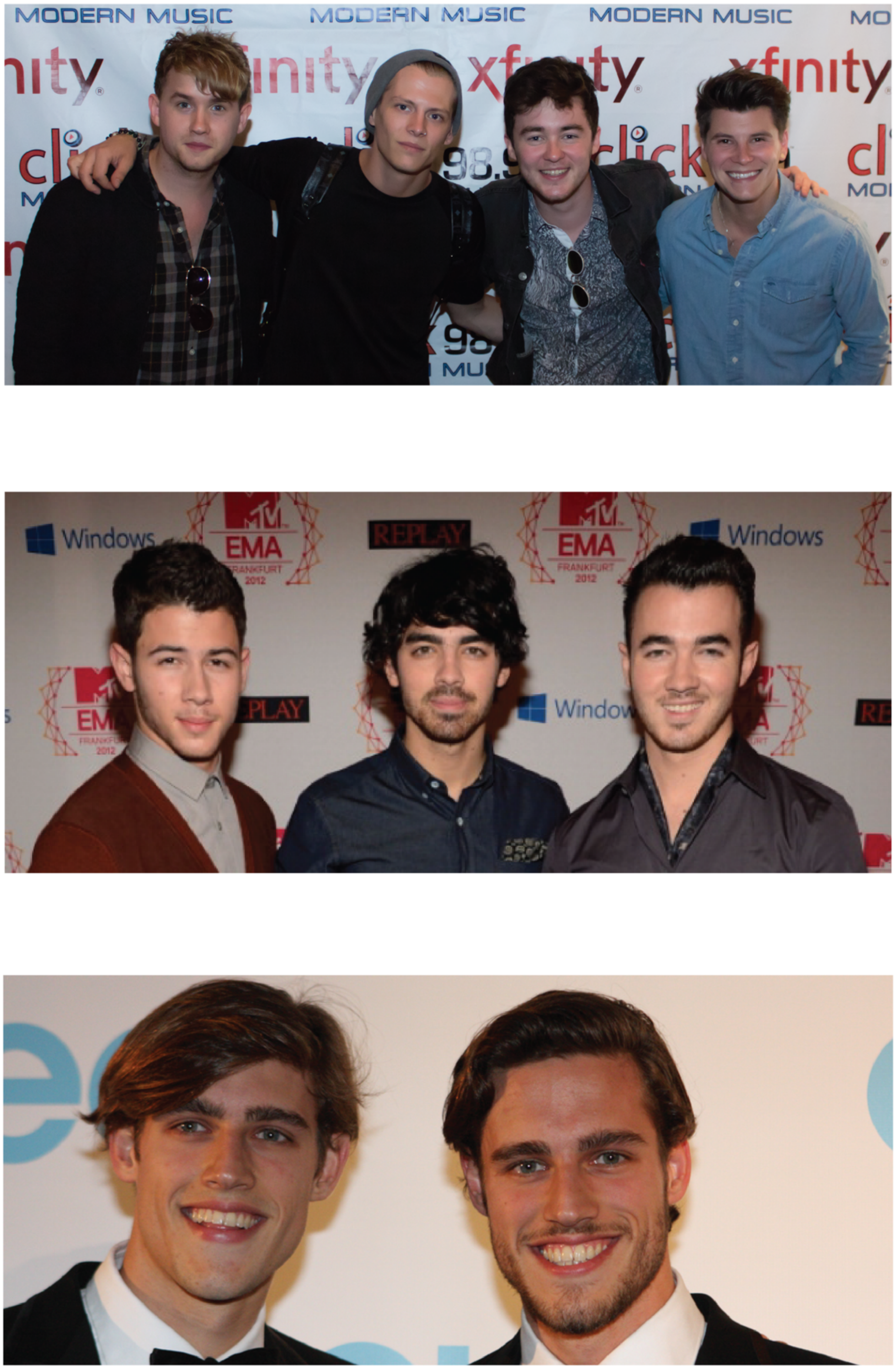


Figure 1: *Male faces that vary in the degree to which they resemble each other. The top row depicts four unrelated males of a similar age. The middle row depicts three brothers who were born between 1987 and 1992. The bottom row shows monozygotic twin brothers. The images were derived from a Google image search and are free for reuse and modification under a Creative Commons license.*

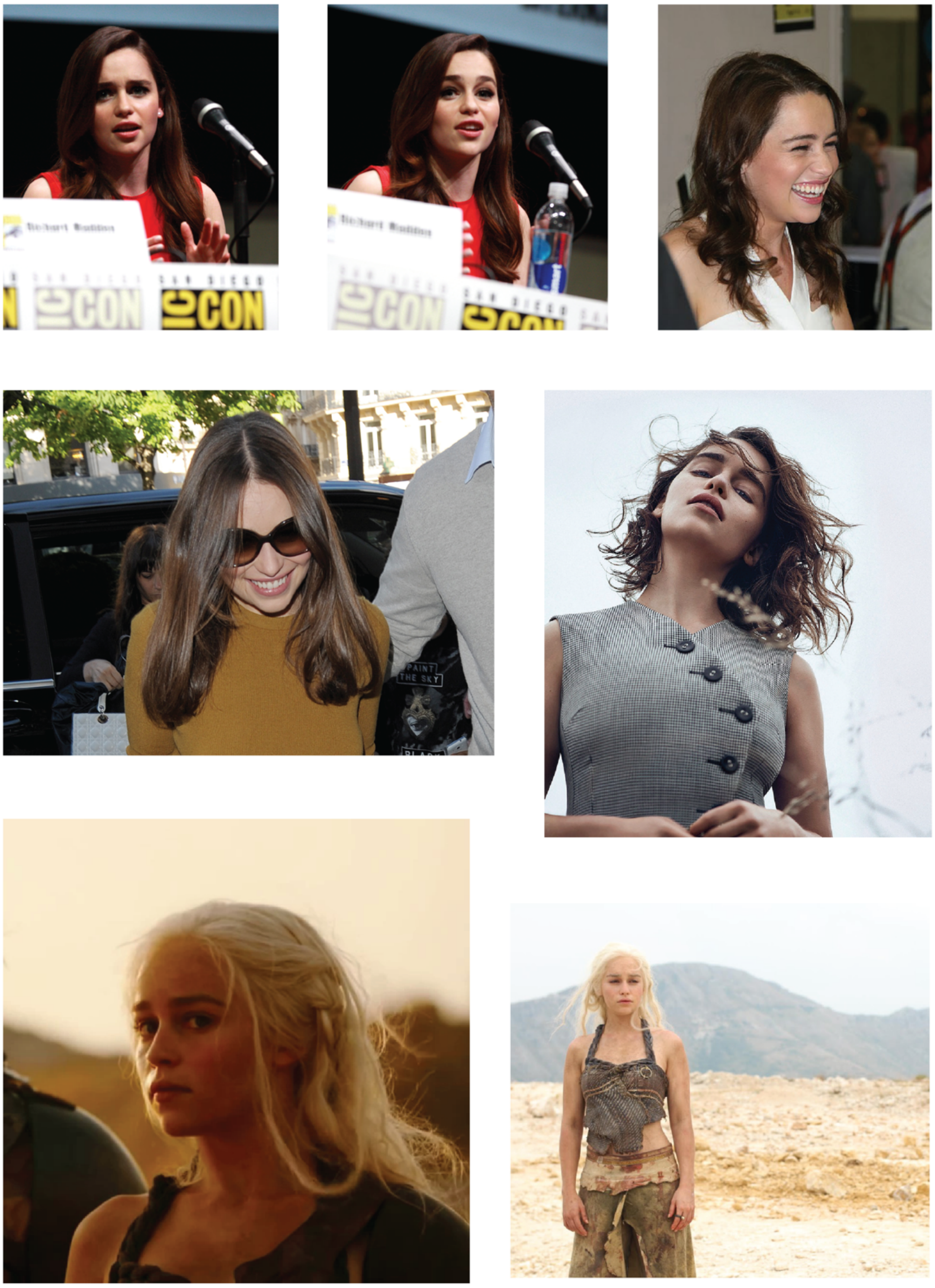


Figure 2: *Photos of the actress Emilia Clarke to illustrate how changes in pose, facial expression, lighting, camera angle, and wearing a wig, or sunglasses can affect the similarity between two images of the same person. The images were derived from a Google image search and are free for reuse and modification under a Creative Commons license.*

The impact of pictorial differences on peoples' ability to recognise and match unfamiliar faces has been used to argue that unfamiliar faces are processed like any other complex pictorial stimulus (Hancock et al., 2000). However, it has been shown that short pre-exposures to similar faces improves discrimination of changes to their internal features, and that this improvement persist when the faces are presented at different orientations during training and test (Dwyer, Mundy, Vladeanu, & Honey, 2009; see also Clutterbuck & Johnston, 2005). The authors argue that the fact that these "well-established results from the face processing literature have been replicated with the current stimuli and testing procedures confirms that the current stimuli are indeed being processed as faces" (Dwyer et al., 2009, p. 352). Two factors temper the strength of this conclusion. First, the change in orientation between pre-exposure and test was quite small, and there were no other pictorial differences between the images. This limits the extent to which these findings can be extrapolated to other face stimuli. Thus, whilst it might be true that face stimuli that are similar at the pictorial level are processed as faces, this might not be the case for other face stimuli. Second, studies on face recognition learning have shown that pictorial factors affect performance, even over multiple training sessions (Bruce, 1982; Hill, Schyns, & Akamatsu, 1997; Krouse, 1981; Longmore, Liu, & Young, 2008). These findings suggest that, even if faces are processed differently than other complex pictorial stimuli, pictorial differences are nevertheless likely to have an impact on the perceived similarity of faces.

In sum, both pictorial differences and differences inherent to the compared faces have an impact on the perceived similarity of unfamiliar faces. Both factors will therefore be considered when evaluating whether, and to what degree, existing findings on the effect of comparison on the perceived similarity of visual stimuli, can be extrapolated to faces.

1.2.1.2 The effect of comparison on the perceived similarity of similar visual stimuli that are difficult to discriminate

The findings from studies on discrimination learning suggest that the opportunity to compare similar visual stimuli that are difficult to tell apart decreases their perceived similarity. A comparison-related decrease in perceived similarity has been found for a range of similar stimuli, including unfamiliar faces (Mundy et al., 2007, 2014), checkerboard patterns (Lavis & Mitchell, 2006; Mundy, Honey, & Dwyer, 2009), dot patterns (Mundy et al., 2014), scenes (Mundy et al., 2014), and geometric shapes (Nelson & Sanjuan, 2009). Examples of these stimuli are shown in Figure 3.

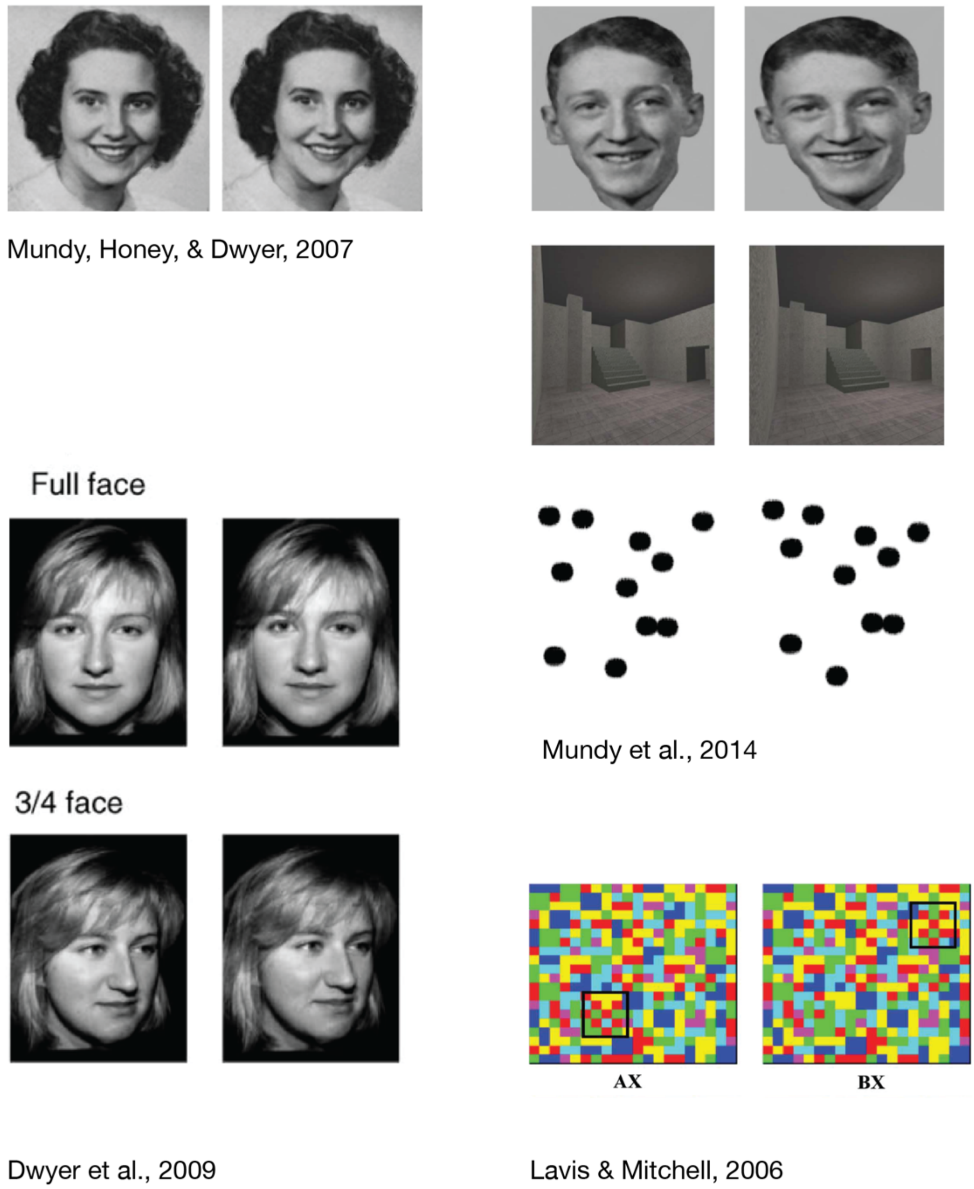


Figure 3: *Examples of the stimuli used in studies of discrimination learning.*

Intermixed exposure schedules facilitate comparison and have been shown to improve discrimination performance relative to blocked exposure for faces (Mundy et al., 2007, 2014; Mundy, Honey, Downing, et al., 2009), checkerboard patterns (Lavis, Kadib, Mitchell, & Hall, 2011; Lavis & Mitchell, 2006) dot patterns (Mundy et al., 2014), and scenes (Mundy et al., 2014). For example, in a study conducted by Lavis and Mitchell (2006) participants were asked to find differences between similar checkerboard patterns (see Figure 3). Comparison was manipulated by modifying the exposure schedule. One stimulus pair was allocated to the blocked exposure condition, the other to the intermixed exposure condition. In the intermixed exposure condition the two stimuli were presented in alternation, 60 times each. In the blocked exposure condition, one stimulus in a pair was presented 60 times in succession, before presenting the second stimulus in the same manner. In the test phase of the first experiment participants were asked to categorize the stimuli as either 'left' or 'right'. In the second experiment, they were asked to discriminate between them on a same–different task. Performance on both tasks was more accurate for intermixed exposure than for blocked exposure. These differences cannot be explained by invoking stimulus familiarity because all the checkerboard patterns were seen an equal number of times. Instead, the superior discrimination performance in the intermixed-exposure condition suggests that the opportunity to compare similar checkerboard patterns reduces their perceived similarity. The intermixed exposure advantage is not limited to exposure schedules involving dozens of repetitions. Research on face discrimination learning has shown that five exposures of 2 s duration to each face in a pair are sufficient to generate this advantage (Mundy et al., 2014; Mundy, Honey, Downing, et al., 2009). This short exposure schedule has also been shown to generate better discrimination performance for faces presented at different angles between exposure and test, and for inverted faces (Dwyer et al., 2009).

Simultaneous exposure to a stimulus pair affords easy comparison and should therefore be particularly effective in facilitating discrimination learning. In support of this notion, simultaneous exposure to pairs of difficult-to-discriminate faces has been shown to result in superior discrimination performance, relative to successive exposure (Mundy et al., 2007). Participants were shown two faces, presented side-by-side, in both the simultaneous and the successive exposure condition. In the simultaneous exposure condition the two images were a pair of similar faces, whereas in the successive condition two copies of the same face were shown together. Both exposure conditions were presented in the same block, according to the following schedule: A-B (simultaneous pair), CC (first member of successive pair), DD (second member of successive pair), A-B (simultaneous pair), and so on. The main difference between this exposure schedule and intermixed exposure, is that each successive presentation of a stimulus pair was interspersed with a presentation of the simultaneous pair. The finding of better performance following simultaneous than sequential exposure was replicated with checkerboard patterns (Mundy, Honey, & Dwyer, 2009).

Relevance and limitations

The overall aim of this thesis to understand the effect of comparison on the perceived similarity of unfamiliar faces. A comparison-related decrease in perceived similarity, operationalised as performance on same–different judgement, and categorization tasks, has been observed for pairs of highly similar faces, and for other visual stimuli that are difficult to tell apart (Lavis et al., 2011; Lavis & Mitchell, 2006; Mundy et al., 2007, 2014; Mundy, Honey, & Dwyer, 2009). These finding suggest that comparison will decrease the perceived similarity of unfamiliar faces. However, one caveat that limits the degree to which these findings can be extrapolated is that the stimuli used in discrimination learning studies are highly similar, both in terms of the inherent similarity of the faces and at the pictorial level. As can be seen in Figure 3, the face stimuli used in discrimination learning studies are identical at the pictorial level, with only some slight phenotypic variation between the faces. Another issue is that discrimination learning studies measure perceived similarity indirectly, through performance. It is possible that direct measures of similarity, such as similarity ratings, and indirect measures, such as response accuracy, reflect different underlying similarity constructs. This question will be addressed in section 1.2.3.

1.2.1.3 The effect of comparison on the perceived similarity of similar visual stimuli that are easy to tell apart

The effect of comparison on the perceived similarity of similar visual stimuli that are easy to discriminate has not been extensively studied. However, there is evidence to suggest that comparing similar visual objects that are easy to discriminate results in an increase in their perceived similarity (Boroditsky, 2007; Hassin, 2001). An experiment in which participants were asked to compare labelled line drawings of pairs of similar animals—for example, a horse and a goat—found that comparison resulted in higher similarity ratings, relative to a no-comparison control (Boroditsky, 2007). For half the participants, the comparison task involved writing down three similarities between the two animals, the other half was required to write down three differences. Both similarity and difference comparisons led to an increase in perceived similarity, compared to a no-comparison control. To test whether comparison also affects the perceived similarity of similar objects for which participants had no pre-existing semantic knowledge, participants were asked to compare labelled, novel 3D objects (see Figure 4 for an example). The comparison task required participants to circle three differences on the experimental booklet, before describing them. The experiment did not include a similarity-focused comparison condition. The experiment found a comparison related increase for the novel pairs.

A slightly different pattern of results was obtained by Hassin (2001) who used the Ebbinghaus illusion to assess the effects of comparison on the perceived similarity of similar visual stimuli. Examples of the Ebbinghaus stimuli, and a summary of the methodology used in Hassin's (2001) study are provided in Figure 5. The study found that similarity comparisons reduced the Ebbinghaus illusion, suggesting that similarity-focused comparisons caused similar objects to appear more similar. This result supports Boroditsky's (2007) conclusion that comparison increases the perceived similarity of similar items. The study further found that difference-focused comparisons resulted in a slight reduction in perceived similarity (i.e. they amplified the illusion); however, this effect was not statistically significant. Hassin (2001) concluded that "similarity and difference judgments are sometimes not complementary simply because similarity judgments enhance perceptual similarity, whereas difference judgments enhance perceptual difference" (p. 730). However, because the decrease in perceived similarity following difference-focused comparison did not reach statistical significance, this result needs to be replicated before it can be deployed as evidence against structural alignment theory.

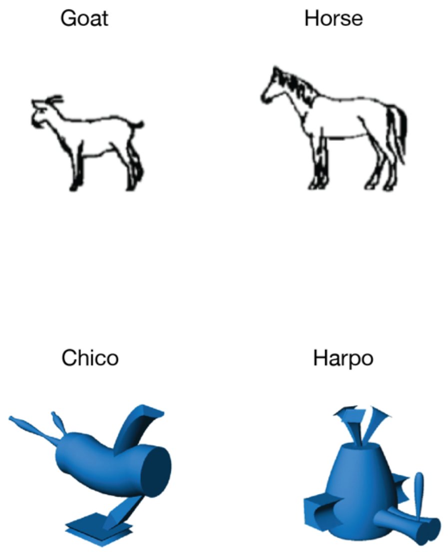


Figure 4: *Examples of the stimuli used in Boroditsky's (2007) study. The top panel shows two labelled line drawing of familiar animals. The bottom panel shows two labelled novel 3D objects.*

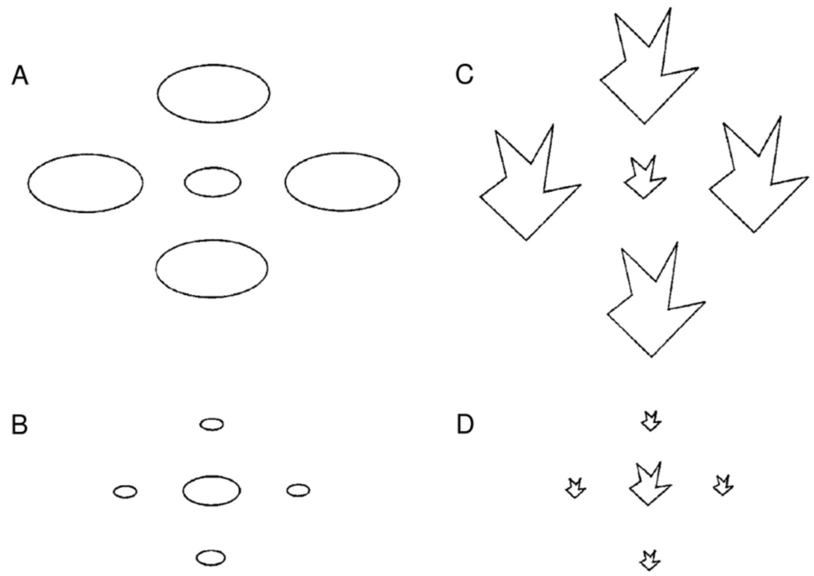


Figure 5: *Examples of the Ebbinghaus stimuli used in Hassin's (2001) study. The Ebbinghaus illusion is an optical illusion that distorts relative size perception. For example, the two central circles in A and B are the same size, yet the central circle surrounded by large circles (A) appears smaller than the central circle surrounded by small circles (B). In Hassin's (2001) study participants in the comparison conditions were ask to rate how similar (or different) the target (central) and context (surrounding) shapes are on a rating scale of 1-9. Perceived similarity was operationalised as target size estimates, which were made by marking the end-point of the target shape on a horizontal line. Comparison-related changes in perceived similarity were measured by comparing size estimates made after similarity or difference comparisons to a no-comparison control.*

Relevance and limitations

A comparison-related increase in perceived similarity has been found for novel 3D shapes, line drawings of animals, and for components of the Ebbinghaus illusion (Boroditsky, 2007; Hassin, 2001). The stimuli used in Hassin's (2001) study consisted of patterns of the same shape. The only difference between the shapes was their size. Consequently, the comparison process could only reveal one difference between the stimuli. As can be seen in Figure 4, the range of perceptual differences between the novel stimuli used in Boroditsky's (2007) study included features that were unique to each stimulus. Together, the finding that comparison-related increases in perceived similarity can be found for both easy-to-discriminate stimuli that vary along a single dimension, and for easy-to-discriminate novel objects that vary along several dimensions and have unique features, suggests that comparison might also increase the perceived similarity of unfamiliar faces that are easy to discriminate.

One limitation that the studies presented here share with the discrimination learning studies reviewed in the previous section is that the stimuli used do not vary pictorially, except for the novel 3D shapes, which appear to be facing in different directions (see Figure 4 for an example). It is therefore not possible to make any predictions about the role of pictorial similarities in mediating the relationship between comparison and changes in perceived similarity based on these findings.

Another caveat, which is unique to Boroditsky's (2007) study, is the use of labels. A study conducted by Wisniewski and Medin (1994) found that when participants were given meaningless category labels (i.e. group 1 and group 2) they tended to categorise drawings based on perceptual features. When the categories were meaningful (i.e. creative children versus non-creative children) they chose descriptions that focused on abstract properties (Wisniewski & Medin, 1994). Similarly, the use of labels in Boroditsky's (2007) study may have biased participants towards conceptual comparisons. This is particularly likely for the animal drawings, because participants have pre-existing semantic knowledge about the animals depicted. The increases in perceived similarity observed for these stimuli may therefore be attributable, at least in part, to changes in the participants' conceptual representations of the animals depicted. However, this alternative account does not apply to the novel objects, both because the labels used were meaningless, and because the participants did not have any pre-existing knowledge about the objects.

A methodological difference between the two studies reported here, and the discrimination learning studies reviewed in the previous section, is the comparison task used. Whereas participants in Boroditsky's (2007) and Hassin's (2001) study were asked to perform explicit comparison tasks, participants in discrimination learning studies were not required to perform a task during pre-exposure, beyond trying to find differences. The potential effect of comparison task on performance is further discussed in the next section.

1.2.2 Operationalising comparison

A common manipulation of comparison that has been deployed in studies of structural alignment is to ask participants to list commonalities and differences between stimuli (Boroditsky, 2007; Gentner & Gunn, 2001; Markman & Gentner, 1996). Participants in studies on discrimination learning, on the other hand, are not asked to list commonalities or differences during the stimulus pre-exposure phase. Nevertheless, the stimulus pre-exposure manipulation is likely to engage comparative processes. First, in several of the discrimination learning studies reviewed in section 1.2.1 participants were explicitly instructed to discover differences between stimuli during pre-exposure. For example: “the grids are very similar but some of them have small differences. Please try to find these differences” (Lavis & Mitchell, 2006, p.2087). It has been argued that these “instructions can be expected to enhance the rewarding properties of detecting a unique feature and might thus promote the development of an appropriate attentional response and of a perceptual learning effect” (Recio et al., 2016, p. 360). Second, even when participants were not instructed to look for differences, other task instructions nevertheless drew attention to the fact that the two stimuli were different. For example, task instructions which mention that the participants will be shown several sets of lookalikes (Mundy et al., 2007), or viral strain variations (Mundy, Honey, & Dwyer, 2009) are likely to highlight that subtle differences between the stimuli have diagnostic value, and are therefore worth attending to. Finally, it has been argued that even if participants are not explicitly instructed to search for differences during the pre-exposure phase, they will inevitably try to detect them when presented with highly similar stimuli (Mitchell & Hall, 2014).

There is evidence to challenge the suggestion that mere exposure is sufficient to engage comparative processes. Two studies failed to replicate the intermixed pre-exposure advantage in discrimination accuracy when participants were instructed either to simply observe the stimuli, or to look at them carefully (Navarro, Arriola, & Alonso, 2016; Recio et al., 2016). It is important to note that the stimuli used in these two studies were checkerboard patterns. Faces, are important social stimuli that attract and engage visual attention from early infancy onwards (Bakhshi, Shamma, & Gilbert, 2014; DeNicola, Holt, Lambert, & Cashon, 2013; Gliga, Elsabbagh, Andravizou, & Johnson, 2009; Langton, Law, Burton, & Schweinberger, 2008). It is therefore possible that the absence of an intermixed exposure schedule advantage is due to an interaction between the exposure schedule and the stimuli used. Whereas the checkerboard patterns did not engage attention in the no-instructions conditions, it is plausible that the faces would have still been compared. Furthermore, the evidence that mere exposure is not sufficient to engage comparative processes is limited intermixed exposure schedules. Simultaneous presentation schedules are more conducive to comparison (because the two stimuli are attended to at the same time), and are therefore more likely to be compared, even in the absence of explicit task instructions.

Another difference between the two experimental paradigms is the exposure schedule used. Whereas pre-exposure in discrimination learning consists of multiple trials, the similarity and difference listing tasks consist of single trials. It could therefore be argued that any differential effects of comparison are attributable to differences in the extent to which the different paradigms facilitate stimulus exposure. However, it has been shown that presenting face pairs five times for 2 s is sufficient for optimal discrimination learning (Mundy et al., 2007, experiment 4). Considering that the comparison task deployed in structural alignment research is more involved—for example, participants in Boroditsky's (2007) study had to write down three features—the overall amount of stimulus engagement is arguably comparable. A related difference between the two experimental paradigms is that, whereas the time available to make a comparison was not restricted in the two studies that have found comparison-related increases in perceived similarity, individual stimulus presentation in the exposure phase of discrimination learning studies are timed. The studies that have used face stimuli have used presentation intervals that ranged from 500ms to 2s (Dwyer et al., 2009; Mundy et al., 2007, 2014). The potential influence of short presentation intervals on information processing during comparison is discussed in section 1.2.3.

Relevance to the aims of this thesis

One of the aims of this thesis is to examine the effect of comparison-task on both the occurrence and the direction of comparison-related changes in perceived similarity. The above discussion has highlighted the variety of tasks that have been deployed to operationalise comparison, ranging from presenting participants with pairs of lookalikes, to asking them to perform an explicit similarity-listing task. The above discussion has also shown that both the number of comparison trials for a given stimulus pair, and the duration of individual comparison trials varied across studies. These considerations have informed the design of the experiments presented in the empirical chapters of this thesis. They are further discussed in section 1.3, which provides an overview of the empirical chapters.

1.2.3 Measuring perceived similarity

An important methodological difference between the discrimination learning literature and Boroditsky's (2007) study of the effects of comparison on perceived similarity is the dependent measure deployed. Whereas studies on discrimination learning measure performance on same–different tasks, participants in Boroditsky's study were asked to judge perceived similarity on a rating scale. Similarity and difference ratings are direct measures of similarity, whereas performance accuracy and response times on a same–different discrimination task are indirect measures. It is possible that similarity ratings and same–different judgments measure different underlying constructs. If this were the case it would be theoretically possible for comparison to both increase perceived similarity, and improve performance on discrimination tasks. However, the findings from studies that have investigated the relationship between similarity ratings and performance on same–different discrimination tasks suggest that response times, discrimination accuracy, and similarity judgements measure the same underlying construct.

Research on the relationship between similarity ratings and performance on same–different discrimination tasks has shown that stimuli that are rated to be more similar are also more difficult to discriminate, and vice versa. In particular, a strong positive correlation was found between similarity ratings and error rates on discrimination tasks (Palmer, 1978; see also Tversky & Gati, 1982). More discrimination errors were made for stimuli with high similarity ratings than for stimuli with low similarity ratings. Similarly, a study that measured perceived similarity, by asking participants to provide dissimilarity ratings, found that the dissimilarity ratings predicted performance accuracy on a discrimination task (Monahan & Lockhead, 1977). Studies that tested the relationship between similarity ratings and reaction times on a discrimination task found a strong positive correlation between the two measures (Goldstone, 1994; Palmer, 1978; Podgorny & Garner, 1979). High similarity ratings were associated with slow response times on the discrimination task, and vice versa.

Two studies that have investigated the relationship between similarity ratings and performance on face-discrimination tasks have also found that high perceived similarity is associated with poor performance, and vice versa (Martini et al., 2015; Sergent, 1984). In the first study participants were asked to rate the dissimilarity between pairs of faces (Sergent, 1984). The study found a negative correlation between response times and dissimilarity ratings. The more dissimilar two faces were perceived to be, the faster participants responded that they were different (Sergent, 1984). Finally, a recent study that has examined the relationship between twins' perceptions of the degree to which they resemble their co-twin (measured on a rating scale from 0 "does not at all resemble my face" to 9 "fully resembles my face") and performance on a face identification task, found that the higher a twin's perceived similarity to their co-twin, the lower their ability to identify their own face (Martini et al., 2015).

The strength of the correlation between similarity ratings and indirect measures of similarity, including performance on discrimination tasks, has been found to vary across studies (see Medin et al., 1993). At least some of this variability is likely to be attributable to task constraints. For example, whereas similarity rating tasks are generally performed under unrestricted viewing conditions, studies that assess the discriminability of stimuli sometimes deploy very short presentation durations to avoid ceiling effects. Short stimulus presentations have been shown to affect both the amount and type of information that is processed (Sergent & Takane, 1987; Stanovich, 1979). Specifically, there is evidence to suggest that global differences and differences in texture are processed sooner than local differences (Lockhead, 1972, 1979; Love, Rouder, & Wisniewski, 1999; Nothdurft, 1992). It follows that same–different judgements made after short stimulus pre-exposures might be more strongly influenced by global differences than either similarity ratings, or same-difference judgements made under longer viewing conditions.

A further difference between the two paradigms is that, same–different judgments are often made under time constraints whereas similarity ratings are not. As a result, the comparison process is more likely to get cut short for same–different judgments than for similarity ratings. Another reason the comparison process might get cut short during same–different judgments is that only one difference needs to be detected to decide that two stimuli are not the same. It follows that the most salient features will have a stronger impact on perceived similarity when it is operationalised as performance on a same–different task, than when it is operationalised as similarity ratings (Sergent, 1984). Evidence in support of this claim is provided by the findings from a study on face processing (Sergent, 1984). The study reported reaction time, response accuracy and dissimilarity rating data that converged to show that the face contour is the most salient feature in determining the similarity of line drawings of faces. However, analysis of the reaction time data showed that "64 percent of the variance was explained by the sameness or difference of the contour, which suggests that raw latency data may overestimate the salience of the most salient feature" (Sergent, 1984, p. 237).

Relevance to this thesis

In sum, studies that have investigated the relationship between direct and indirect measures of similarity have consistently found a positive correlation between perceived similarity, response times, and error rates. It therefore seems unlikely that the decrease in perceived similarity observed in discrimination learning studies, and the increase in perceived similarity observed in Boroditsky's (2007) study are attributable to differences in the measures of similarity deployed. In this thesis both similarity ratings, and same–different judgements were therefore deployed to measure the effect of comparison on perceived similarity.

One important caveat to the notion that response times, error rates, and similarity ratings all measure the same underlying similarity construct is that differences in task constraints can impact on the relative contribution of stimulus features on the perceived similarity of the compared stimuli. Under time constraints global differences and salient features are likely to have a greater impact on similarity than in the absence of time constraints. Another reason why salient features are likely to play a bigger role in in similarity when it is operationalised as performance on a same–different task, relative to when it is operationalised as a similarity rating, is that just one difference is sufficient to determine that two stimuli are not the same. To control for these effects, the experiments reported in the empirical chapters of this thesis did not impose any time constraints on either the similarity rating task, or the same–different judgment task.

1.2.4 Theoretical accounts of comparison

1.2.4.1 Comparison as a process of structural alignment

The central tenet of structural alignment theory is that comparison amounts to the alignment of common representational structures (Markman & Gentner, 1993b, 1993c). To establish structurally consistent mappings between items, the process of structural alignment needs to satisfy two constraints: one-to-one mapping and parallel connectivity (Markman & Gentner, 1993a, 1993b). One-to-one mapping is satisfied when each element (e.g. attribute, relation or object) in one representation matches to at most one element in the other representation. Parallel connectivity is satisfied when corresponding relations have matching objects. To illustrate, if the relation 'repairs' in 'man repairs a computer' is matched with the relation 'fixes' in 'robot fixes a car' then car and computer need to be aligned to satisfy the parallel connectivity constraint. It follows that "object mappings are determined not only on the basis of their intrinsic similarities, but also on the basis of their playing similar roles in like relational structures" (Markman & Gentner, 1993b, p. 518). A third constraint is systematicity, according to which relations that belong to systems of interconnected relations will be favoured over isolated relational matches (Gentner, 1983; Gentner & Toupin, 1986; Markman & Gentner, 1993b, 1993a).

The hypothesis that comparison involves the alignment of structured representations was tested in a study that used pairs of causal scenes (Markman & Gentner, 1993c). Participants were required to match an object in one scene with a corresponding object in the second scene. A mapping was considered an 'object' mapping if a perceptually similar object was chosen, and a 'relational' mapping if objects were matched based on their common role in the scene. Participants in the similarity-first condition were asked to rate the similarity of the two scenes before performing the matching task. Consistent with the prediction that similarity comparison is a process of structural alignment, participants in the similarity-first condition chose the relational match more frequently than participants who did not rate the similarity of the scenes prior to mapping. Furthermore, participants in the similarity-first condition chose the relational match more frequently than participants who rated the artistic merit of the scenes prior to mapping. The latter finding discounts alternative interpretations that attribute the comparison-related increase in relational mapping to stimulus pre-exposure.

The structural alignment process generates a structurally consistent mapping which is used to establish the commonalities (matching elements) and differences (mismatches) of a pair (Markman & Gentner, 1993a, 1993b). Consistent with feature-based accounts of similarity (Tversky, 1977), structural alignment theory posits that similarity increases as a function of commonalities and decreases as a function of differences (Markman & Gentner, 1993b). Tversky (1977) explored the relative contribution of common and distinctive features to perceived similarity in a series of experiments. In the first experiment similarity ratings as well as feature lists for vehicles were elicited. The experiment found a positive correlation of .84 between rated similarity and shared features, and a negative correlation of -.64 for distinctive features. This finding suggests that commonalities play a greater role than differences in determining similarity. To further test this claim participants were asked to list commonalities and differences between word pairs (Markman & Gentner, 1993b). Similarity ratings were found to increase with the number of commonalities listed, and to decrease with the number of differences. Furthermore, the study replicated the finding that commonalities are more important to perceived similarity than are differences.

Limitations

Structural alignment theory can account for the comparison-related increases in perceived similarity that have been found for similar objects (Boroditsky, 2007; Hassin, 2001). According to structural alignment theory comparison increases the perceived similarity of alignable objects, because the commonalities that are highlighted by the alignment process are weighted more highly in perceived similarity than the differences (Markman & Gentner, 1993b, 1996). However, the theory cannot explain why comparison decreases the perceived similarity of visual stimuli that are difficult to discriminate (Mundy et al., 2007, 2014; Mundy, Honey, & Dwyer, 2009). To increase the explanatory power of structural alignment theory, it therefore needs to be considered alongside other theoretical accounts of comparison. One potential candidate is representational distortion theory.

Representational distortion theory posits that the degree of similarity between two entities is shaped by the ‘transformational distance’ between them (Hahn et al., 2003). “Transformational distance refers to the complexity required to ‘distort’ the representation of one object into the representation of another” (Hodgetts et al., 2009, p. 64). In other words, perceived similarity is dependent on the amount of cognitive effort required to transform mental representations of objects. According to a hybrid model the transformation between two objects necessitates the specification of an alignment between object representations (Hodgetts et al., 2009). Structural alignment facilitates the transformation process by exposing components shared between representations that can “be transformed as a whole, rather than piecemeal” (Hodgetts et al., 2009, p.64). In another hybrid model representational distortion is used to constrain the structural alignment process (Grimm, Rein, & Markman, 2012). More specifically, representational distortion is used to determine the optimal mapping between two items in comparisons where several mappings are possible. The best mapping is the one that requires the shortest transformation distance. However, whilst these two hybrid models provide a more detailed account of comparison they still fail to explain comparison related increases in perceived similarity.

To gain a full understanding of the relationship between comparison and perceived similarity, structural alignment theory needs to be considered alongside a theory of similarity that can explain comparison-related decreases in perceived similarity. One such account is habituation, which is discussed in the next section.

1.2.4.2 Comparison as a process of habituation

The term ‘perceptual learning’ has been used to refer to relatively long-lasting as well as short-term changes in perception brought about by experience (Bedford, 1999; Goldstone, 1998). One area of perceptual learning that has been extensively studied is the effect of stimulus pre-exposure on discrimination performance (e.g. Gibson, 1969; Honey, Bateson, & Horn, 1994; Mundy, Honey, & Dwyer, 2007, 2009; Wang & Mitchell, 2011). This research has shown that pre-exposure to similar stimuli can facilitate later discrimination between them (for reviews, see Goldstone, 1998; Mitchell & Hall, 2014). According to an early theoretical account of perceptual learning developed by Eleanor Gibson in the 1960s “the opportunity to compare two similar stimuli allows a process of stimulus differentiation to operate wherein attention to the unique features of the stimuli is increased relative to their common features” (Mundy, Honey, & Dwyer, 2009, p. 24).

Supporting evidence for the notion that comparison plays a causal role in discrimination learning is provided by research on the effects of different exposure schedules on discrimination performance. These studies consistently show that those exposure schedules that facilitate comparison of the to-be-discriminated stimuli, such as intermixed exposure, result in better discrimination performance. The superiority of intermixed over blocked pre-exposure has been demonstrated for a variety of stimuli, including unfamiliar faces (Mundy et al., 2014, 2007), checkerboard patterns (Lavis & Mitchell, 2006), dot patterns (Mundy et al., 2014), and geometric shapes (Nelson & Sanjuan, 2009). Direct evidence that these performance gains are attributable to comparison is provided by the finding that interrupting comparison at the exposure stage (by inserting visual masks between stimuli) diminishes the benefit of intermixed exposure (Dwyer, Mundy, & Honey, 2011). Further evidence that comparison facilitates discrimination learning is provided by the finding that simultaneous exposure to a stimulus pair (a manipulation that is especially conducive to comparison because it eliminates memory load) is the most beneficial exposure schedule for discrimination learning (Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009).

Gibson's second claim that the opportunity to compare two stimuli during pre-exposure increases the attentional weighting of the unique stimulus attributes, is also well supported. A study conducted by Wang and Mitchell (2011, experiment 4) used eye tracking to test whether the superior performance following inter-mixed over blocked pre-exposure would be reflected in an increase in attention to the unique elements. The experiment replicated the finding that intermixed pre-exposure results in superior discrimination performance compared to blocked pre-exposure. In line with the hypothesis that intermixed pre-exposure helps guide attention to unique stimulus features, the eye tracking data showed that participants spend more time looking at unique stimulus features following intermixed than blocked pre-exposure (Wang & Mitchell, 2011; see also Wang, Lavis, Hall, & Mitchell, 2012). Further supporting evidence is provided by the finding that participants who have learned about the attributes and location of discriminating features during pre-training, can transfer that learning to novel backgrounds and discriminating features (Jones & Dwyer, 2013; Moreno-Fernández, Salleh, & Prados, 2015; Wang et al., 2012; see also Angulo & Alonso, 2013). These findings show that viewers learn both where to look and what to look for during pre-exposure, and that this learning drives their attention at test.

Several explanations have been put forward for the effect of pre-exposure schedule on discrimination learning (for a review, see Mitchell & Hall, 2014). One explanation is that selective habituation to the common stimulus features directs attention to the unique stimulus features (Honey & Bateson, 1996; see also Hall, 2003). According to this account, successive presentations of stimuli with shared (X) and unique (A and B) elements results in greater habituation to the shared elements (X) than to the unique elements (A and B), because the shared elements are seen more often. Evidence consistent with this claim is provided by the finding that mere exposure to the shared element between two checkerboard patterns resulted in superior discrimination performance compared to a no pre-exposure control (Wang &

Mitchell, 2011, experiment 2). Eye tracking data showed that the superior discrimination performance was associated with greater visual attention to the unique stimulus features. This finding lends further support to the argument that habituation to shared elements during pre-exposure enhances attention to the unique elements during the discrimination task.

The habituation model was further developed by Mundy and his colleagues (Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009) who argue that short-term habituation effects that operate during simultaneous comparisons, have a lasting effect on the relative attentional weightings that are assigned to different object features. They argue that during simultaneous exposure participants habituate to the common features as they shift their gaze back and forth between the two stimuli. This habituation to the common features in turn results "in some long-lasting change, wherein the unique elements become better represented and available to be learned about subsequently" (Mundy et al., 2007, p. 136). Finally, the authors state that the habituation process is less likely to operate for stimuli pairs that share few common elements.

Combining structural alignment and habituation

Neither structural alignment theory nor the habituation account can explain the range of comparison-related changes in perceived similarity that have been observed for visual stimuli. Whilst the habituation account can explain the comparison-related decreases in perceived similarity that have been found for homogenous stimuli that are difficult to tell apart, it cannot account for comparison-related increases in perceived similarity. The opposite holds true for structural alignment theory. Another structural account that might explain the range of comparison-related changes is representational distortion. Representational distortion theory posits that the similarity of a given stimulus pair reflects the ease of transforming one stimulus representation into another (Hahn et al., 2003; Hodgetts et al., 2009). It follows that prior comparison should increase perceived similarity because the transformation process should become easier with repetition. However, like structural alignment, representational distortion cannot account for comparison-related decreases in perceived similarity. For this reason, and because it shares many assumptions with structural alignment (e.g. both accounts presume that comparison involves mental transformation of stimuli, either to align them or to morph them into each other), representational distortion theory will not be further considered.

It has been argued that structural alignment might operate alongside a second mode of similarity processing (Markman & Gentner, 2005). The second mode provides fast similarity judgments based on feature and surface similarities, but does not take the structure of the compared objects into account. Structural alignment is a slower mode of comparison that yields a more detailed, and nuanced similarity assessment. According to this view, an initial assessment of similarity, which is provided by the fast mode, guides the allocation of attentional resources (Markman & Gentner, 2005). It is possible that structural alignment and habituation act in opposition to each other to guide attention.

Relevance to the aims of this thesis

The overarching aim of this thesis is to understand how the interplay between comparative processes, and stimulus attributes affects perceived similarity. Theoretical accounts of comparison further us towards this goal by defining the cognitive mechanisms that underpin performance. Empirical findings in turn test the boundary conditions of theoretical accounts. Neither structural alignment theory nor the habituation account of comparison can explain the range of comparison-related changes that have been found for similar visual stimuli. They will therefore be considered alongside each other in this thesis.

1.3 Summary, and overview of empirical chapters

Unfamiliar face matching—the process of verifying the equivalence of two or more faces—has been shown to be surprisingly error prone, even when only two faces need to be matched, and high quality face stimuli are used (Henderson et al., 2001; Megreya & Bindemann, 2009; Megreya & Burton, 2008; Megreya et al., 2013). Understanding why these errors occur is of both applied and theoretical interest. On the applied front, understanding the causal factors that give rise to poor performance on face matching tasks can be used to predict, and possibly improve, performance in a variety of operational contexts. On the theoretical front, it can be used to refine theories of comparison. This thesis assumes that the likelihood that two unfamiliar faces will be judged to be the same person, is largely determined by the perceived similarity of the stimulus pair.

It follows that an understanding of the causal factors that underlie poor performance on face matching needs to start with an understanding of the variables that affect perceived similarity. This thesis examines the effect of comparison on the perceived similarity of unfamiliar faces.

The face stimuli that are matched in applied settings vary widely in the degree to which they resemble each other, both because of phenotypic variation in human faces, and due to pictorial differences. There is evidence to suggest that these variations in similarity interact with the comparison process, to affect perceived similarity. As illustrated in Figure 6, comparison has been shown to increase the perceived similarity of visual stimuli that are similar, but easy to tell apart (Boroditsky, 2007; Hassin, 2001), and to decrease the perceived similarity of similar stimuli that are difficult to tell apart (Mundy et al., 2007, 2014; Mundy, Honey, & Dwyer, 2009). Extrapolating to face stimuli these findings suggest that comparison has diametrically opposed effects on the perceived similarity of similar face stimuli that are difficult to discriminate, and those that are easy to tell apart. However, as can be seen in Figure 6, the stimuli used in these studies are not representative of the range of faces that are encountered in applied settings. In fact, the only face stimuli that were deployed are phenotypically highly similar, and pictorially identical face morphs. The first aim of this thesis is to address this gap by assessing the effect of comparison on the perceived similarity of a wider range of unfamiliar face stimuli.

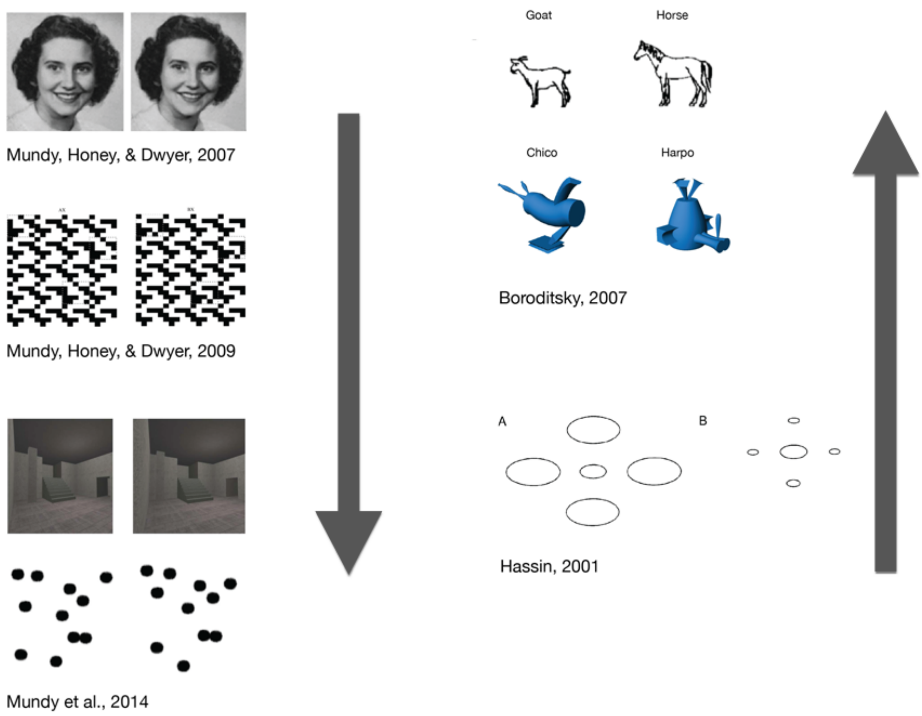


Figure 6: *The effects of comparison on perceived similarity. The opportunity to compare highly similar stimuli results in a decrease in similarity (i.e. improves performance on a stimulus discrimination task). Comparing similar stimuli that are members of related categories results in an increase in perceived similarity. For dissimilar stimuli that belong to unrelated categories comparison results in a decrease in perceived similarity (but only for difference-focused comparisons).*

The second aim of this thesis is to understand whether, and in what manner, the nature of the comparison task influences perceived similarity. There is evidence that forensic facial identification examiners outperform members of the public on face matching tasks (Norell et al., 2015; White, Dunn, Schmid, & Kemp, 2015; White, Phillips, Hahn, Hill, & O'Toole, 2015). Morphological image analysis—a comparison technique that is commonly used by facial identification examiners—focusses on the analysis of individual features (Wilkinson & Evans, 2009). Considering that morphological image analysis, and other face matching training methods, are focussed on features, it has been argued that the superior performance of these experts might be attributable to increased selective attention to facial features (White, Phillips, et al., 2015). The finding that similarity-focussed comparisons increase perceived similarity, whereas difference-focussed comparisons may decrease it (Hassin, 2001), suggests that the focus of the comparison task might affect the outcome of the comparison. In the experiments reported in this thesis, the effect of the focus of the comparison task on perceived similarity was explored by requiring participants to complete both a similarity- and a difference-listing tasks. The boundary conditions of any comparison-related influences on perceived similarity were further assessed by asking participants to perform an implicit comparison task, and by asking them to simply compare observed faces without writing down any commonalities and differences.

Face matching is routinely performed in a variety of operational settings, including passport checks at airports, identity verification in banks, proof of age to buy alcohol, and deciding whether a suspect and the culprit captured are a match. This thesis will focus on two applied contexts: deciding whether a face matches a passport photo at border control, and deciding whether a suspect's face matches that of the perpetrator captured on Close-Circuit-Television (CCTV). A comparison-related increase in perceived similarity would increase the likelihood of same responses on face matching tasks. In other words, observers would become more accurate at identifying that face stimuli are the same person, but less accurate at identifying that two face stimuli depict different people. The opposite pattern of results would be expected for comparison-related decreases in similarity. Yet, despite their importance the conditions that mediates them are not well understood. The current thesis contains three empirical chapters that examine how the interplay between stimulus attributes, and comparison task affects the perceived similarity of unfamiliar face pairs.

Chapter 3 explores the effect of comparison on the perceived similarity of unfamiliar faces that should be relatively easy to tell apart, both because the face stimuli are of good quality with relatively few pictorial differences, and because there are clear phenotypic differences between the faces. Understanding the effects of comparison on the perceived similarity of faces that should be relatively easy to distinguish is of particular relevance within the context of passport checks at airports, because performing repeated comparisons has been shown to increase the likelihood that two unfamiliar faces that share some phenotypic resemblance are judged to be same person (Alenezi & Bindemann, 2013; Alenezi et al., 2015; Bindemann, Fysh, Cross, & Watts, 2016; Fysh & Bindemann, 2017). It also offers a useful test of the boundary conditions of the different theories of comparison, because the predictions of structural alignment theory have never been tested with face stimuli, and the habituation account has only been tested with faces that are pictorially identical and phenotypically difficult to discriminate.

Chapter 3 presents three experiments in which participants were asked to compare two faces before rating the perceived similarity between them on a scale from 1 (not similar) to 7 (very similar). Comparison-related changes in perceived similarity were measured by comparing similarity ratings obtained for face pairs that were subjected to a comparison task prior to the similarity rating task, to similarity ratings obtained for a no-comparison control face pair. There were two comparison tasks: a similarity-focussed comparison task (in which participants were asked to write down three similarities between two faces), and a difference-focussed comparison task (in which participants were asked to write down three differences). The primary aims of Experiment 1 were to assess whether comparison will affect the perceived similarity of unfamiliar faces that are relatively easy to tell apart, and to test whether any comparison-related changes are affected by the relative similarity of the compared faces. Experiment 2 tested an alternative account for comparison-related changes in perceived similarity, namely, that they are attributable to increased familiarity with the compared faces. To address this question, participants were asked to describe the two faces in a pair before rating their perceived similarity. In the first two experiments the two faces in a pair were shown from different views: frontal and ¾. In Experiment 3 participants were asked to compare upright to inverted face to establish whether any comparison-related changes in perceived similarity would extend to a different pictorial manipulation.

Chapter 4 explores two boundary conditions of the effect of comparison on the perceived similarity of unfamiliar faces that should be relatively easy to tell apart, which are of particular relevance to the context of passport identity checks. The first boundary condition is the view from which the compared faces are presented. Whereas the face pairs in the first three experiments were shown from different angles, the stringent requirements for passport photographs to be accepted mean that border control agents compare two faces from the same angle: a passport portrait photo, and a person standing in front of them. Experiment 4 assessed the effect of comparison on the perceived similarity of unfamiliar faces shown from the same frontal view. There were two phenotypic similarity conditions: half of the participants were asked to compare faces that were relatively similar, the other half were asked to compare faces pairs that were relatively dissimilar. Experiment 4 also further tested the boundary conditions of the comparison-related changes in perceived similarity, by asking participants to perform a task that is likely to invoke comparison, without explicitly needing to verbalise the comparison. The second boundary condition that is of relevance in the applied context of border control is the time available to perform a comparison. Border control agents are under pressure to meet performance targets (Toynbee, 2016), and therefore only have a limited time available to decide whether a given face-photo pairing is the same person or not. Participants in Experiment 5 were shown unfamiliar face pairs for a fixed duration of time, before rating the similarity of the pair. They were asked to compare the faces for the duration of the presentation but were not required to write down any information about them.

The task of matching the face of a suspect (from a photo or in person) to CCTV footage is common in forensic and criminal justice contexts (Norell et al., 2015; Wilkinson & Evans, 2009). Both system specific, and situational variables can affect the quality of the face-specific information that is available in CCTV footage. Chapter 5 explores the effect of these pictorial differences on perceived similarity. The stimuli used in Experiment 6 consisted of two images of unfamiliar faces. One of the images was degraded to simulate the effects of poor quality video footage. Participants in Experiment 7 were asked to compare stills of faces to video footage. The amount, and quality of face-specific information available from the video footage was degraded, partially because the face took up a relatively small amount of the visual field. In both experiments, participants were asked to decide whether two faces were the same person or not. Perceived similarity was measured by comparing the number of ‘same’ and ‘different’ responses obtained after completing a comparison task, to the number of ‘same’ and ‘different’ responses for a no-comparison control pair. Both experiments deployed a similarity- and a difference-listing task. In addition, participants in Experiment 7 were also asked to perform a face-description task, to assess whether any comparison-related changes for the stimuli might be attributable to familiarity.

Chapter 2: Methodology

This chapter provides a rationale for the experimental method used in the research reported in this thesis, and outlines its limitations. Section 2.1 commences with a summary of the hypotheses that were tested. It then offers an overview of the experimental approach used, and how the chosen method differs from the approach commonly used in psychophysics experiments. The section then outlines how the independent variables (comparison and stimulus similarity) and the dependent variable (perceived similarity) were operationalised. Section 2.1 concludes with a discussion of how both similarity ratings and same–different judgments can be conceptualised as reflecting distances within face space. Section 2.2 explores whether the findings of the experiments reported in this thesis can be generalised to non-face stimuli. To address this issue the literature on whether face processing is qualitatively different from other forms of object processing is evaluated. The section concludes with a discussion of the extent to which the cognitive processes that underlie face recognition are likely to impact on simultaneous unfamiliar face matching performance.

Section 2.3 evaluates the ecological validity of the research presented in this thesis. It reviews face comparison practices deployed during legal proceedings, and in the context of border security. It outlines how various cognitive biases may impact on human performance in these applied contexts. Finally, the section also discusses how the use of automated face recognition systems affect human face matching performance. Section 2.4 addresses the question of population validity. Many practitioners who perform face comparisons as part of their work have received specialist training in unfamiliar face matching. The findings from research with novice participants, including the research reported in the empirical chapters of this thesis, might not generalise to these experienced, trained individuals. Similarly, any effects of face comparison on perceived similarity observed in a sample drawn from a normal distribution of face-recognition abilities, such as the individuals who took part in Experiments 1–7, might be limited to individuals with normal face recognition abilities. Section 2.4 concludes with a review of the literature on super-recognisers—a small subsection of the general population who possess exceptional face recognition, and matching abilities.

2.1 Method and hypotheses

2.1.1 Summary of hypotheses

The seven experiments reported in this thesis were designed to examine the effect of comparison on the perceived similarity of unfamiliar face stimuli. As discussed in section 1.2.1.1 there are two stimulus attributes that can contribute towards the perceived similarity of a given unfamiliar face pair. The first attribute is phenotypic similarity—the degree to which the depicted individuals look alike. There is some evidence to suggest that comparison will decrease the perceived similarity of face pairs with a high degree of phenotypic similarity (Honey & Bateson, 1996; Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009), and increase the perceived similarity of face pairs with a low degree of phenotypic similarity (Boroditsky, 2007). However, because these findings were obtained in studies that have either used homogenous artificial faces or non-face objects as stimuli, their generalisability to natural face stimuli is limited. A further limitation is that the two bodies of research, and their theoretical underpinnings, have not been integrated. As a result, the interplay between the different cognitive processes that have been put forward to explain comparison-related increases and decreases in perceived similarity has not been subjected to empirical scrutiny. It is therefore not possible to pinpoint whether, and at what degree of phenotypic stimulus similarity, the effect of comparison on perceived similarity is likely to change direction. The first hypothesis that was tested in this thesis states that comparison will change the perceived similarity of phenotypically similar and dissimilar face pairs relative to a no-comparison control. The interaction between phenotypic similarity and comparison was assessed in Experiment 1, and in Experiments 4–6.

Pictorial differences between face stimuli include variations in pose, facial expression, lighting, and distance from camera, differences in image quality, and the presence of disguises such as hats and sunglasses. There is evidence that pictorial differences play a role in the perceived similarity of unfamiliar faces (Bindemann & Sandford, 2011; Jenkins et al., 2011; Kramer & Ritchie, 2016; Megreya et al., 2013; Noyes & Jenkins, 2017; Redfern & Benton, 2017). For example, a study that assessed the impact of within-person variability on face perception found that viewers sorted photos (sourced from the results of an internet image search to capture within-person variability) of two unfamiliar individuals into eight identities on average (Jenkins et al., 2011; see also Redfern & Benton, 2017). In other words, they believed that eight, not two, individuals were depicted in the photos. Similarly, a study that investigated the effect of differences in camera distance on unfamiliar face matching accuracy found that the pictorial manipulation resulted in a decrease in accuracy for same-identity pairs, but not for different-identity pairs (Noyes & Jenkins, 2017). These finding suggest that the presence of pictorial variations decreases the perceived phenotypic similarity of the depicted faces. However, simultaneous face matching studies have shown that the negative impact of pictorial differences on accuracy is not limited to match (i.e. same identity) trials, but also decreases performance on mismatch (i.e. different identity) trials (Bindemann et al., 2013; Lee et al., 2009; Megreya et al., 2013). Arguably the effect of a given pictorial difference on the perceived similarity of two unfamiliar faces depends on whether it obscures salient phenotypic difference between two faces (for example, a distinctive nose shape), or creates the appearance of non-existent phenotypic differences (for example, when the camera angle and distance alters the shape of a face). The role of pictorial differences in mediating the effect of comparison on the perceived similarity of face stimuli is currently not well understood. The second hypothesis that was explored in this thesis states that comparison will alter the perceived similarity of face stimuli across a range of pictorial differences relative to a no-comparison control. It is important to note that the interaction between pictorial differences and comparison was not directly tested in any of the experiments. Instead, pictorial differences were varied between experiments, ranging from non-existent (the homogenous face morphs used in Experiments 4 and 5) to extensive (the videos and stills used Experiment 7).

Although comparison has been shown to play a crucial part in a variety of cognitive processes, its effect on the perceived similarity of naturalistic visual stimuli such as faces remains relatively unexplored. To help close this gap the research reported in this thesis concentrated on three aspects of the comparison task. The first aspect was whether the comparison is centred on finding commonalities or differences between the compared faces. There is some evidence to suggest that whilst similarity-focussed comparisons increase perceived similarity, difference-focussed comparisons decrease it (Hassin, 2001). However, there is also evidence to suggest that similarity- and difference-focussed comparisons can both result in an increase in the perceived similarity of the compared stimuli (Boroditsky, 2007). The second aspect was whether comparison was an explicit task, or whether the need to compare the faces in a pair was an implicit requirement. The third hypothesis that was tested in this thesis states that a variety of different comparison tasks will result in a change in perceived similarity relative to a no-comparison control. The effect of different comparison tasks on the perceived similarity of unfamiliar faces was assessed in Experiments 1–4, and in Experiments 6 and 7. The final aspect of comparison that was assessed was comparison duration. Research on face matching under time-pressure suggests that the time available to make a comparison can affect the outcome of the comparative process (Bindemann et al., 2016; Fysh & Bindemann, 2017; Wirth & Carbon, 2017). The fourth hypothesis states that the effect of comparison on the perceived similarity of unfamiliar face stimuli will change as a function of comparison duration. The effect of comparison duration on perceived similarity was assessed in Experiment 5. The four hypotheses that were explored in this thesis are summarised in Table 1.

	Hypotheses
1	Comparison will change the perceived similarity of phenotypically similar and dissimilar face pairs relative to a no-comparison control.
2	Comparison will alter the perceived similarity of face stimuli across a range of pictorial differences relative to a no-comparison control.
3	A variety of different comparison tasks will result in a change in perceived similarity relative to a no-comparison control.
4	The effect of comparison on the perceived similarity of unfamiliar face stimuli will change as a function of comparison duration.

Table 1: *The four hypotheses assessed in this thesis.*

2.1.2 Method overview

The general method deployed in Experiments 1–7 is summarised in Figure 7. As can be seen comparison was a within-subjects variable, and stimulus similarity was a between-subjects variable. In Experiment 5, which was designed to assess the time-course of comparison, the within-subjects variable was presentation duration. In all experiments except for Experiment 5 participants completed at least three comparison conditions. For example, participants in Experiment 1 completed a similarity-focussed comparison (in which they were asked to describe three similarities between two faces), a difference-focussed comparison (in which they were asked to describe three differences), and a no-comparison control condition (in which they did not perform a comparison task). The two images that were subject to a comparison were always presented simultaneously, without any time constraints. After completion of the similarity-focussed comparison task the face pair that was the subject of the comparison was presented a second time, and participants were asked to rate the similarity of the faces on a scale of 1 (not similar) to 7 (very similar). The face pair that was compared in the difference-focussed comparison task, was also subjected to a similarity rating. The face-pair in the no-comparison control condition, as the name implies, was not subjected to a comparison. It was only presented once, to elicit a similarity rating. The comparison tasks are further discussed in section 2.1.4.

In Experiments 1–5 the dependent variable consisted of similarity rating task. In Experiments 6 and 7 the dependent variable consisted of a same–different judgement. In Experiments 1, 4, 5, and 6 phenotypic stimulus similarity (i.e. the degree to which the two faces in a pair resembled each other) was varied between participants. Approximately half of the participants were shown relatively similar face pairs and the other half were shown relatively dissimilar face pairs. In addition to the phenotypic differences some of the experimental stimuli also had pictorial differences. Whereas phenotypic similarity was a between-subjects variable, the pictorial differences existed within face pairs. For example, as can be seen in Table 2, one of the faces in each of the face pairs used in Experiments 1 and 2 was shown from a frontal view, the other from a ¾ view. Because of this experimental design decision it was not possible to assess the interaction between different pictorial difference manipulations and the comparison task. The stimulus sets and similarity manipulations are outlined in section 2.1.3.

In each experiment except for Experiment 5, there was only one experimental trial per experimental condition. This approach was chosen to address a concern that performing multiple trials would introduce response biases. For example, task repetition has been shown to increase the number of ‘same’ responses on a simultaneous face matching task (Alenezi & Bindemann, 2013; Alenezi et al., 2015). The approach of administering one trial per experimental condition differs from psychophysics experiments in which large numbers of trials (ranging from dozens to hundreds) are administered. One drawback of the approach adopted here relative to the approach adopted in psychophysics experiments is that it is not possible to analyse datasets from single trails with signal detection procedures. As a result, it is not possible to discriminate between changes in response thresholds (e.g. how similar two faces need to be before the observer responds that they ae the same person) and changes in perception (e.g. how similar two faces appear to the viewer). The phrase 'perceived similarity'—as it is used in this thesis—does not distinguish between perceptual changes in similarity, and changes in response thresholds.

Another difference between the method deployed in the experiments reported here, and psychophysics experiments is that the experiments were untimed. Except for Experiment 5 (which investigated the time-course of comparison) participants were free to examine the faces for as long as they wished in the different treatment conditions. The similarity rating task (Experiments 1–5), and the same–different judgment task (Experiments 6 and 7) were also untimed. One motivation for the untimed administration is that time-pressure has been shown to affect perceived similarity as measured by performance on a same-different judgement task (Bindemann et al., 2016; Fysh & Bindemann, 2017; Wirth & Carbon, 2017). Timed stimulus presentations are used in psychophysics experiments to avoid of ceiling effects. However, ceiling effects were not a concern in the experiments reported here. In Experiments 1–5 perceived similarity was assessed using a similarity rating task. Any ceiling effects on this task are more likely to be attributable to stimulus similarity (i.e. near identical stimuli might receive very high similarity ratings) than to stimulus exposure duration. Experiments 6 and 7 did deploy a same–different task. However, they were meant to mimic real world forensic applications which are not subject to strict time limits.

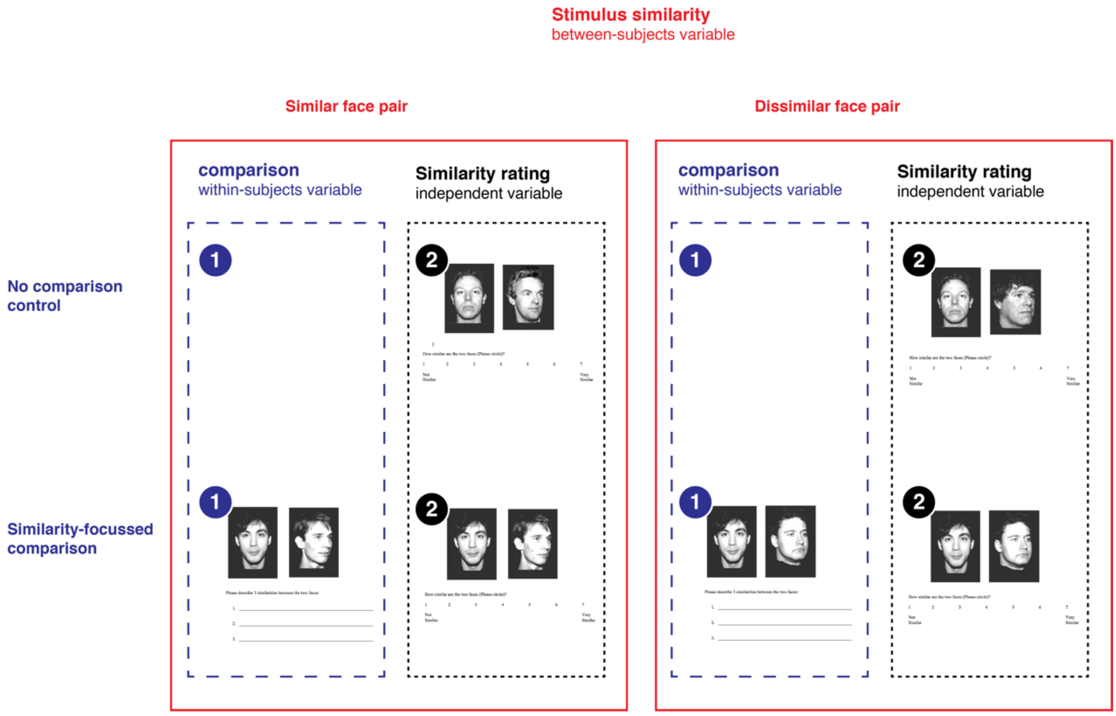


Figure 7: *The general experimental design. In most of the experiments participants performed at least three comparison conditions including a similarity-focussed comparison, and a no-comparison control. Face pairs in the similarity-focussed comparison condition were presented twice. Once during the comparison task and once during the similarity rating task. Face pairs in the no-comparison control were only presented once to elicit a similarity rating. Comparison task was always a within-subjects variable. Phenotypic similarity was a between-subjects variable in a subset of experiments.*

2.1.3 The stimuli used in Experiments 1–7

It has been argued that problems associated with face recognition can be usefully divided into two types: "imaging problems, such as variations in viewpoint and lighting; and problems inherent in the structure of faces, such as their configuration or distinctiveness" (Hancock et al., 2000, p. 330). In this thesis variations between unfamiliar face stimuli that are attributable to imaging are referred to as pictorial differences, and variations that are attributable to differences in facial structure are referred to as phenotypic differences. The range of unfamiliar face pairs that were used as stimuli in Experiments 1–7 is shown in Figure 8. As can be seen phenotypic similarity varied substantially between face pairs—ranging from two images of the same person (Experiment 6, same faces condition) to two dissimilar looking male faces (Experiment 5, different faces condition). It is important to note that although the face pairs used in Experiments 1–7 varied in phenotypic similarity, they all consisted of Caucasian male faces with an approximate age range of 18 to 30 years. As a result, the impact of differences in appearance that are attributable to differences in gender, age, and ethnicity was not captured in the research reported in this thesis. Same category face-pairs (i.e. faces that belong to the same gender, ethnicity, and age group) were used to control for the potential impact of category knowledge on similarity assessments. Figure 8 also showcases the range of pictorial differences between stimuli. As can be seen the number of pictorial differences between faces ranged from none (the similar faces in Experiment 5) to many (the video and still image pairs used in Experiment 7). Figure 8 also showcases the types of pictorial differences including the differences in pose (for example, full face versus ¾ view) and differences in image quality (for example, pixelated versus non-pixelated).

As illustrated in Figure 9 the unfamiliar face stimuli that were used can be subdivided into those that are artificial (i.e. stimuli that were created using face morphing software) and those that are natural (i.e. photos, and video footage). It has been argued that the use of artificial stimuli in face research limits the degree to which findings can be generalised to real world contexts (Burton, 2013). This begs the question—why use stimuli created with face morphing software at all? Face morphs were used in some of the experiments reported here because they allow for the systematic manipulation of phenotypic stimulus similarity. It is important to note that Burton's (2013) criticism is not restricted to artificial face stimuli. It also applies to stimulus sets that were generated as a resource for research on face processing. Burton (2013) has argued that the use of these constrained stimulus sets (which control for pose, facial expression, and lighting conditions) has obscured the effects of within-face variability on performance. Except for Experiment 7 this criticism can be applied to the natural face stimuli used in the experiments reported here. However, as has been proposed in section 1.2.1.1 both phenotypic similarity (i.e. the degree to which the depicted faces resemble each other) and pictorial similarity (i.e. the degree to which the faces are captured under the same conditions) contribute towards the perceived similarity of a given stimulus pair. The use of stimuli taken from face databases allows for the examination of the role of phenotypic similarity whilst controlling for the influence of pictorial differences.

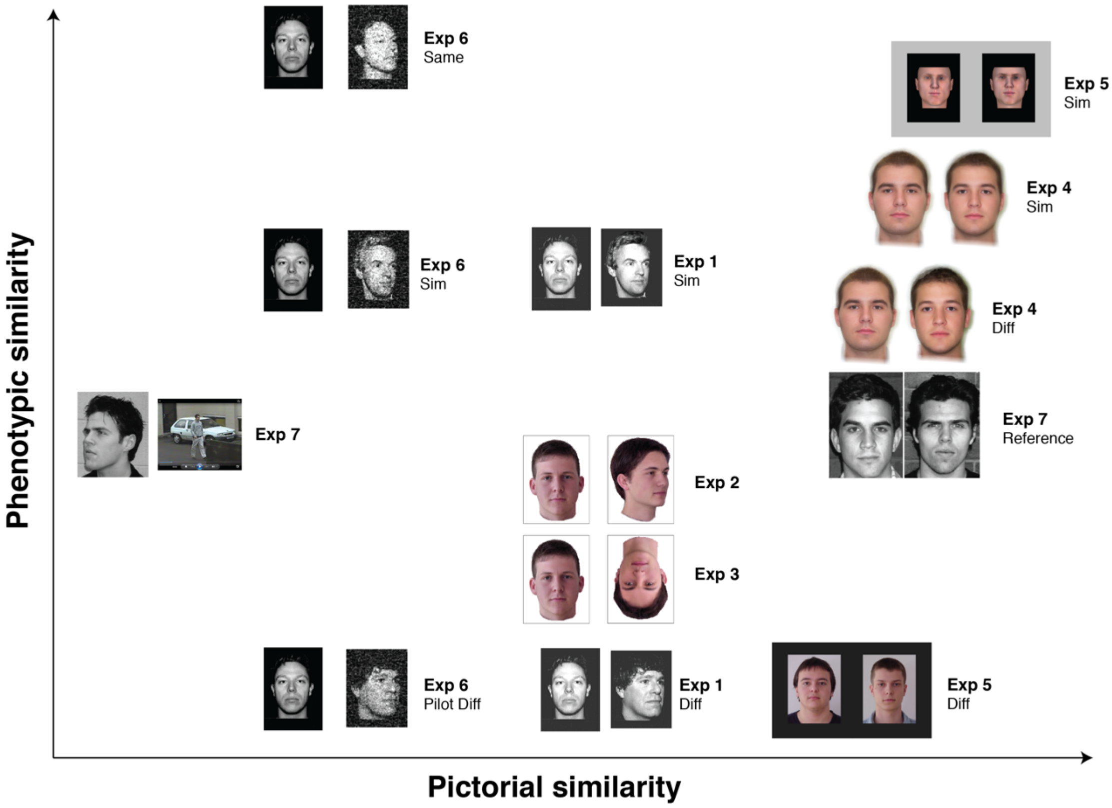


Figure 8: *The range of stimuli used in Experiments 1-7. The face pair labelled 'Exp 7 Reference' has been taken from Davis and Valentine (2009) to illustrate the phenotypic resemblance between the faces used in Experiment 7 under optimised viewing conditions. The face pair labelled 'Exp 6 Pilot Diff' was used in a pilot study discussed in the introduction of Experiment 6. The remaining face pairs are the experimental stimuli used.*

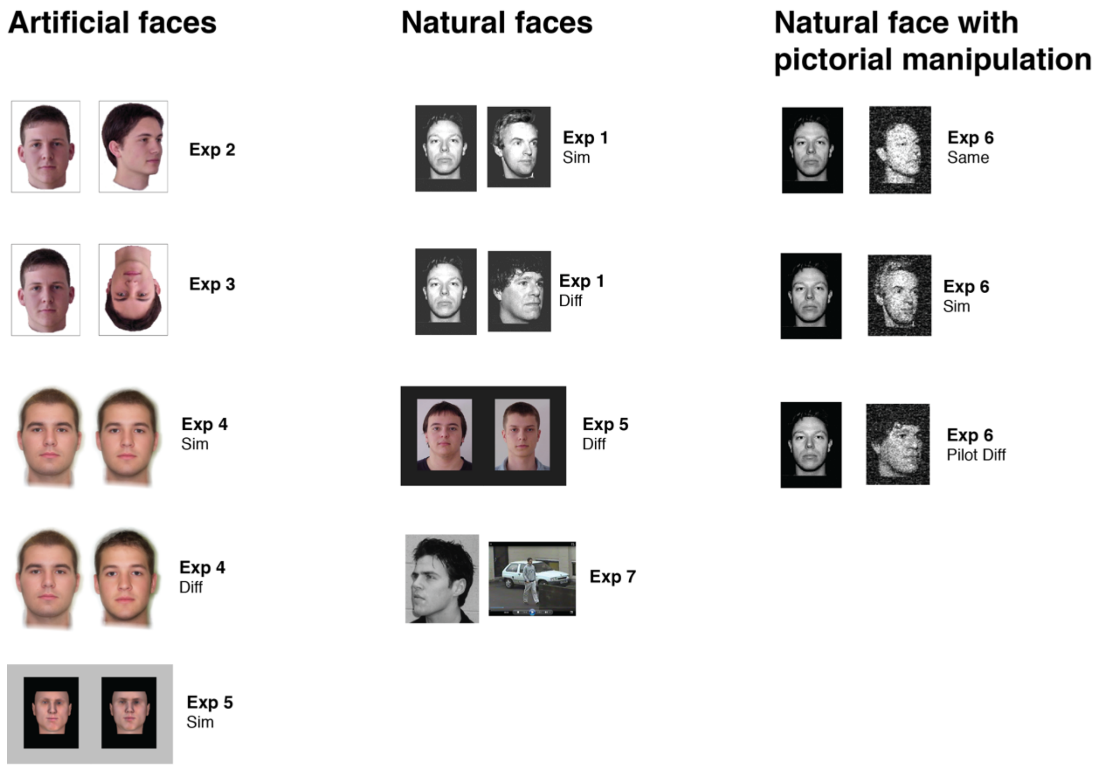


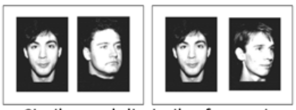
Figure 9: *The face stimuli used in Experiments 1–7. The stimuli can be divided into two categories. The first category, shown in the leftmost column, consists of artificial faces that were generated with face morphing software. They were created to systematically manipulate stimulus similarity. The second category consists of natural faces that were sampled from stimulus sets that have been created for face research. The rightmost column shows a pictorial manipulation (the addition of visual noise) that was used in Experiment 6 to simulate low quality CCTV footage.*

2.1.3.1 Phenotypic similarity

Three approaches were used to both match faces into pairs, and to create sets of face pairs that have a comparable degree of phenotypic similarity. The specific approach that was used in each experiment is listed in Table 2. The first approach was to use subjective similarity assessments. For example, the similar and dissimilar face pairs used in Experiment 1 were created by asking participants in a pilot study to rate the perceived similarity of face pairs on a scale of 1 (not similar) to 7 (very similar). The pairings with the highest similarity ratings were used in the similar faces condition and those with the lowest similarity ratings were used in the dissimilar faces condition. Similarity ratings constitute a direct measure of similarity. Indirect measures of similarity were also used to match faces into pairs. For example, in Experiment 5 a grouping task was used to match faces into dissimilar pairs. The advantage of using subjective similarity assessments is that the commonalities and differences between the matched faces reflect the facial characteristics that drive human perceptions of similarity. The main disadvantage is that this approach makes it difficult to create sets of face pairs that are equal in their degree of between-pair similarity because it relies on the range of faces available in face databases.

In Experiments 2 and 3, and in the dissimilar faces condition in Experiment 4, faces were matched into pairs based on similarity assessments made by the experimenter. For example, in Experiments 2 and 3 faces were matched into pairs based on face-shape, eye colour, and hair colour. To control for phenotypic similarity between stimulus pairs the same set of criteria were applied to each face pair. The main disadvantage of this approach is that that the criteria the experimenter used to match faces may not be the most pertinent to observers when assessing similarity. However, the subjective impressions of the experimenter were not the only means by which stimulus heterogeneity was controlled for. In Experiments 2–4 face morphs and face averages were used to reduce the heterogeneity of the stimuli. The use of artificial faces that are created with face morphing software is further discussed below.

The second approach that was used to pair faces was to use identity as a proxy for phenotypic similarity. In Experiment 6, accuracy for similar face pairs was compared to accuracy for face pairs that had the same identity (i.e. two images of the same unfamiliar face). This method rests on the assumption that face pairs that consist of different photos of the same individual are more similar than face pairs that consist of images of different people. There is evidence that pictorial differences can reduce the perceived phenotypic similarity of same-identity face stimuli (Jenkins et al., 2011). However, this is unlikely to have affected the phenotypic similarity manipulation in Experiment 6, because the same pictorial manipulations were applied to similar and same-identity face pairs. The main disadvantage of using identity as a proxy for similarity is that it can only be used to create high-similarity face pairs. For example, in Experiment 6 a different face matching approach was still needed to create face pairs for the similar faces condition.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Face pair	 Similar and dissimilar face pair			 Similar and dissimilar face pairs
Approach used to match faces into pairs	Similarity ratings obtained in a pilot study Similar face pairs = pairs with high similarity ratings dissimilar face pairs = pairs with low similarity ratings	One face in each pair had a square or long face shape, the other had an oval face shape (as judged by the experimenter). Brown eyes vs green / blue eyes Same hair colour		Similar face pairs = face average A + face morph (50% average A / 50% average B) dissimilar face pairs = face average A + face average B (The face averages were matched into pairs based on the experimenter's subjective impression of similarity)
Approach used to control for variance in phenotypic similarity between face pairs	Face pairs that were grouped together had comparable similarity ratings	Each face was a morph of two faces to reduce stimulus heterogeneity.		Face averages were used to reduce stimulus heterogeneity
Pictorial differences within face pairs	Frontal view vs 3/4 view	Frontal view vs 3/4 view	Upright vs upside down	none
Approach used to control for pictorial differences between face pairs	The monocrome photos were captured under the same photographic conditions All faces had a neutral facial expression	The colour images were captured under uniform lighting conditions with the same camera All faces had a neutral expression		The faces that were used to create the face morphs were captured under the same photographic conditions All face morphs had the same neutral expression and were shown from a frontal view

	Experiment 5	Experiment 6	Experiment 7
Face pair	 Similar and dissimilar face pair	 Same and similar face pairs	
Approach used to match faces into pairs	Similar face pairs = 3D face model A + 3D face model A with 10% random variance added dissimilar face pairs = faces that were clustered together at most once in a sorting task performed by eight pilot study participants	Same face pairs = same identity pairs similar face pairs = pairs with high similarity ratings (obtained in pilot study)	Faces were matched into pairs based on the likelihood that they would be mistaken for each other (see Davis & Valentine, 2009).
Approach used to control for variance in phenotypic similarity between face pairs	Face pairs in the similar faces condition were artificially created to be highly similar	Face pairs that were grouped together in the similar faces condition had comparable similarity ratings	Both face pairs generated a high number of incorrect same responses on a same-different task (see Davis & Valentine, 2009)
Pictorial differences within face pairs	none	Frontal view vs 3/4 view non-pixelated vs pixelated	Multiple, including image quality and medium (image vs video clip)
Approach used to control for pictorial differences between face pairs	The photos used in the dissimilar faces condition were captured under the same photographic conditions All faces (photos and 3D models) had a neutral facial expression and were shown from the same frontal view	The monocrome photos were captured under the same photographic conditions All faces had a neutral facial expression	Both face pairs consisted of one high quality photograph, and one short video clip showing a set sequence of movements

Table 2: *The approaches used to match faces into pairs, and to create stimulus sets.*

The third approach that was used to create sets of face pairs was to use face morphing software. As mentioned above, the main advantage of this approach is that it allows for the close control of phenotypic similarity. Consider for example the similar face pairs used in Experiment 5. As can be seen in Figure 9 these faces are highly similar. This is because the second face in each pair was created by adding 10% of random variation to the first. The main disadvantage of using face morphing software for stimulus creation is that the resulting face stimuli are artificial. The use of artificial faces limits the extent to which findings can be generalised to natural face stimuli (i.e. photos, and video stills). This issue is particularly pertinent in Experiment 5 for two reasons. First the face pairs used in the dissimilar faces condition were not artificial. This limits the degree to which it is possible to conclude that any differences in perceived similarity between similar and dissimilar faces are attributable to differences in phenotypic similarity. Any differences may instead be attributable, at least in part, to the fact that the dissimilar faces consisted of photos whilst the similar faces were artificial stimuli. Second, the face morphs used where cropped at the top. This reduced the availability of some external features of the face—the face outline, and hair. This may have influenced the perceived similarity of the similar face pairs because external features have been shown to play an important role in unfamiliar face processing (Bonner, Burton, & Bruce, 2003; Bruce et al., 1999). However, these shortcomings are limited to the similar face pairs used in Experiment 5. As can be seen in Figure 9, the artificial faces used in Experiments 2–4 did have hair and a complete face outline. Furthermore, research on the use of fraudulent ID shows that face morphs like those used in Experiments 2 and 3 (i.e. artificial faces that were created by morphing two photographs) can be mistaken for authentic photographs (Robertson, Kramer, & Burton, 2017). This finding illustrates that face morphs can be very realistic face stimuli.

The software used to create artificial face stimuli used in Experiments 2–5

The two types of software applications commonly used to create artificial faces are composite creation, and face morphing software. The artificial stimuli used in Experiments 2–5 were created with face morphing software. Composite creation applications are nevertheless briefly reviewed here to provide a fuller understanding of the different applications available to create an objective or subjective facial likeness. Composite creation software is used in police investigations to construct an image of the perpetrator's face based on the memory of an eyewitness. The resulting composite is based on an individual's memory of a face, and therefore a subjective likeness to the perpetrator. Traditional mechanical composite systems, which were introduced in the 1970s, often resulted in a poor likeness of the target face because the composite was constructed from a limited range of facial features (for a review of facial composite systems see Frowd, 2015, 2017). Correct identifications from composites created with mechanical systems were further hampered by the presence of demarcation lines between different facial elements. Feature-based composite creation software including E-FIT and FACES was designed to overcome these limitations (Frowd, 2015). Feature-based applications contain a much greater range of features, which can be blended to create yet more features. The realism of the faces created with these systems is enhanced by the absence of demarcation lines, and by the fact that features can be resized and positioned freely within the composite. Evidence that facial composites created with feature-based systems are better target likenesses is provided by the finding that their use resulted in an 18% correct naming rate, whilst the correct naming rate for the mechanical Photofit system was found to be only 6% (Frowd et al., 2005).

The holistic composite systems EFIT-V, EvoFIT and ID have been developed to further improve on the capability to create an accurate likeness of a given individual. Their development was informed by an understanding that faces are processed more holistically than other objects (see section 2.2.1). Holistic systems generate a set of artificial faces from which a witness can select a subset of faces that resemble the target face (Frowd, 2015). The artificial faces are generated using principal component analysis (PCA). Some of the parameters that are randomly varied by the software when generating faces (including face width and weight) can be described by an observer with relative ease. However, the effect of most parameters on facial appearance is difficult to convey verbally (Frowd, 2017). The subset of faces selected by a witness is used to generate a new set of faces, which is likely to contain instances that resemble the target face more closely. One method by which holistic composite systems create new faces from a sample is by selecting two faces and combining them—with each face contributing 50% towards the new 'child' face. Alternatively, a genetic mutation technique can be used in which random variation is added to a given face. The creation of a new set of faces based on a previous subset can be repeated several times until a close likeness is found. To further increase the likeness of the final composite EvoFIT and EFIT-V both contain holistic scales that can be used to manipulate holistic attributes such as age, attractiveness, health and honesty. Mean identification accuracy using EvoFIT was found to be between 25% and 46% in lab-based experiments (Frowd, Pitchford, et al., 2011; Frowd, Skelton, et al., 2012). Similarly, studies that have assessed success rates in applied contexts have found that composites created with EvoFit and EFIT-V resulted in identifications of suspects in 25% to 60% of cases (Frowd, Hancock, et al., 2011; Frowd, Pitchford, et al., 2012; Solomon, Gibson, & Mist, 2013).

The second type of application commonly used to create artificial faces is face morphing software such as Psychomorph (Tiddeman, Burt, & Perrett, 2001), Abrosoft FantaMorph, and Morpheus Photo Morpher. The first step when using Psychomorph and similar software packages is to delineate each photograph that will contribute towards a given artificial face. Delineation involves the positioning of points to mark the features of the face (including the eyes, the mouth, the nose and the face outline). The next step depends on the desired outcome. Psychomorph can be used to create face averages, caricatures, face transforms, and face morphs. Face averages are created by computing a midpoint for each marker to define the features and shape of the average face, and by blending together pixel brightness and colour values to create an average texture and colour. This technique was used to create the dissimilar face pairs used in Experiment 4. Face morphs are created by blending one image into another by a predefined degree. For example, to create a 50% morph between face A and face B the face morphing software moves the markers that delineate face A 50% towards the markers that delineate face B. Pixel colour and brightness are also blended to create values that lie halfway between those of stimuli A and B. This technique was used to create the second stimulus in the similar faces condition in Experiment 4, which consisted of a 50% morph between two face averages. Similarly, each stimulus used in Experiments 2 and 3 was a 50% morph between two photos of male faces. Morpheus Photo Morpher was used to generate the face stimuli because it allowed for the creation of ¾ view face morphs. Both face morphing and face averaging can be considered objective similarity manipulations, because the same manipulation is consistently applied to all facial features. For example, when creating an average face all features are moved towards the average position computed from all the input faces.

Most face morphing software applications including those discussed above operate on two-dimensional images. FaceGen Modeller, on the other hand, operates on 3D face models. The face space that is used to generate the 3D models was created by performing a principal components analysis on a data set of 272 high-resolution 3D face scans. One-hundred-and-nine of the scanned faces were female, 163 were male. Eight of the scanned faces were classified as South Asian, 29 as East Asian, 183 as European, 26 as African, and 26 as other. The individuals whose faces were scanned ranged in age from 12 to 67 years, with a median age of 28 years. FaceGen Modeller was used to create the similar face pairs in Experiment 5. The face pairs were created in two stages. In the first stage 3D face models were generated from photos used in the dissimilar faces condition. These 3D face models constituted the first face in each similar face pair. The second face in each pair was created by adding 10% random variation to the first. The process by which the random variation was added is akin to the genetic mutation technique used in holistic composite creation software. In addition to adding random variation to face models the FaceGen Modeller can also be used to create entire 3D face Models at random.

Creating similar faces by adding random variation to a 3D model created with FaceGen Modeller can be considered an objective similarity manipulation because the random variation that is added is based on an analysis of the variance in appearance found in a relatively large sample of 3D face scans. However, there are two caveats with this conclusion. First, principal components do not explain an equal amount of variance. It is not clear how FaceGen Modeller's random mutation function manipulates the different principal components, and how a random 10% mutation is balanced across the principal components. Intuitively, a 10% mutation along the first principal component, which reflects the dimension with the greatest variance, should have a stronger impact on the appearance of a face than a 10% mutation along the 50th component. Singular Inversions (the company that has developed FaceGen Modeller) provide no documentation on how the random mutation manipulation has been implemented. The second caveat relates to the range of faces that were analysed. A total of 101 European male face scans have contributed towards the creation of the face space that is used to generate the 3D face models. Whist it seems plausible that this constitutes a sufficient sample, it is not clear whether it represents the full range of natural variance in Caucasian male faces. More concerningly, all other ethnic categories are underrepresented. For example, only eight South Asian faces were included. The biases that are introduced by the overrepresentation of Caucasian male faces in image sets that have been used both to create face modelling algorithms, and to train face-recognition algorithms is further discussed in section 2.3.

2.1.3.2 Pictorial differences

As can be seen in Figure 8, in all experiments, except for Experiments 4 and 5, there were pictorial differences between the two faces in a pair. In Experiments 1 and 2 there was a single pictorial difference: one face in each pair was shown from a frontal view, the other from a ¾ view. This variation in pose might seem like a trivial manipulation. However, there is evidence that changes in facial appearance that are introduced by varying camera-to-subject distance, or by capturing images months apart, are sufficient to have a negative impact on face matching accuracy (Megreya et al., 2013; Noyes & Jenkins, 2017). Furthermore, differences in pose between learning and test (for example, presenting a face from a frontal view in the learning phase and from a profile view in the test phase) have been shown to have a negative effect on face recognition accuracy (Hill et al., 1997; Liu & Chaudhuri, 2002; O'Toole, Vetter, & Blanz, 1999; Troje & Bülthoff, 1996; Wallraven, Schwaninger, Schuhmacher, & Bülthoff, 2002). Finally, and most importantly, differences in pose have also been shown to affect accuracy on sequential (Burke, Taubert, & Higman, 2007; Favelle, Palmisano, & Avery, 2011; Newell, Chiroro, & Valentine, 1999), and simultaneous (Bruce et al., 1999) face matching tasks. It therefore seems plausible that variations in pose could impact on the perceived similarity of a face pair.

Experiment 3 examined whether holistic processing plays a role in mediating the effect of comparison on the perceived similarity of unfamiliar face stimuli. There is evidence to suggest that face recognition is heavily reliant on holistic processing, which is disrupted by inversion (Collishaw & Hole, 2002; Farah, Tanaka, & Drain, 1995; Farah, Wilson, Drain, & Tanaka, 1998; Freire, Lee, & Symons, 2000). An ideal assessment of whether holistic processing plays a role in mediating the effect of comparison on perceived similarity would be to compare performance in an upright faces condition to performance in an inverted faces condition, using the same face stimuli. However, it was not prudent to present two inverted faces side-by-side in Experiment 3, because the experiment (like most of the seven experiments reported in this thesis) was administered on paper, to groups of students. It would therefore have been easy for students to sidestep the experimental manipulation (and make the comparison easier to perform) by turning the experimental booklet upside-down. In Experiment 3, one face in each pair was presented upright, the other upside-down. This pictorial manipulation was introduced to disrupt holistic processing. The underlying assumption is that comparison of an upright to an inverted face requires mental rotation. There is evidence that the mental rotation of complex visual stimuli, including faces, is a piecemeal (feature-based) process (Ashworth, Vuong, Rossion, & Tarr, 2008; Collishaw & Hole, 2002; Förster, Gebhardt, Lindlar, Siemann, & Delius, 1996; Liesefeld, Fu, & Zimmer, 2015; Lin & Hsieh, 2012; Bruno Rossion, 2009; Bruno Rossion & Boremanse, 2008). It follows that participants will be unable (or at least impaired in their ability) to utilise holistic processing when comparing an upright to an inverted face.

In Experiment 6 there were two pictorial manipulations. The first manipulation was a difference in pose, the second manipulation was a difference in image quality. In Experiment 6 image quality was manipulated by pixelating one face in each pair. This image quality manipulation was designed to simulate poor quality CCTV footage. In Experiment 7 participants were required to decide whether a face depicted in a high-quality image was the same than a face shown in a relatively low-quality video clip. Previous research has shown that the use of pixelated images, and of stills taken from poor quality CCTV footage, are detrimental to performance accuracy in a simultaneous, unfamiliar face matching task (Bindemann et al., 2013; Henderson et al., 2001). There is also evidence to suggest that the number of pictorial differences can have a cumulative effect on performance. A study that assessed the combined effect of two pictorial manipulations found that for pitch rotations, differences in pitch (i.e. differences in the degree to which a face looks up or down) and lighting direction (i.e. whether the face was lit from the top, the bottom or the front) both affected performance, and that there was an interaction between the two variables (Favelle, Hill, & Claes, 2017). However, for yaw rotations no interaction was found between the yaw rotation manipulation (i.e. the degree to which a face looks to the left or right) and the corresponding lighting direction manipulation (i.e. whether the face was lit from the left, the right, or the front). Whilst differences in yaw rotation were found to have a strong effect on performance, the lighting direction manipulation was found to have no impact (Favelle, Hill, & Claes, 2017). Experiments 6 and 7, assessed whether the effect of comparison on the perceived phenotypic similarity of unfamiliar faces is altered by the presence of multiple pictorial differences.

The main shortcoming of the approach used in Experiments 1–7 to investigate the role of pictorial differences is that pictorial differences were not manipulated within experiments. This approach was adopted to focus on the role of phenotypic similarity, which was a between-subjects variable in Experiments 4–6, and in Experiment 1. At the time the experiments were run, there was no existing research on the role that phenotypic similarity plays in mediating the effect of comparison on the perceived similarity of unfamiliar faces. And whilst there is evidence that pictorial differences play an important role in determining the similarity of unfamiliar face stimuli, it was decided to prioritise variations in phenotypic similarity—which are inherent to the stimulus—before assessing the role of pictorial differences. To assess the impact of a given pictorial difference on performance the difference needs to be varied whilst holding phenotypic similarity constant. For example, a hypothetical study that deploys a mixed-design, with comparison as the within-subjects variable, could be run to assess the role that variations in viewing angle play in mediating the effect of comparison on phenotypic similarity. In this experiment, the two faces in a pair are shown from different views (e.g. frontal and profile) in one between-subjects condition, and from the same view in the second between-subjects condition. The same identities are used in both conditions to control for the influence of difference in phenotypic similarity.

A second shortcoming of the implementation of the pictorial difference variable in Experiments 1–7 is that it was conflated with the dependent variable. In Experiments 1–5 a rating task was used to measure perceived similarity, and there was either a single pictorial difference, or no pictorial differences between the two faces in a pair. In Experiments 6 and 7 a same–different rating task was used to assess perceived similarity, and there were either two, or multiple, pictorial differences between the stimuli. This conflation arose because one of the aims of Experiments 6 and 7 was to assess whether any comparison related changes in perceived similarity could have an impact on performance in forensic contexts. Whilst the images that are used for identifications in many applied contexts are standardised, the images that are compared in forensic settings often vary in both viewing angle and image quality. Furthermore, whilst a similarity assessment underlies both similarity ratings and same–different judgements, decisions in applied contexts are often binary. For example, a police officer who is assessing the similarities between an image of a potential suspect and an image of the perpetrator may need to decide whether to bring the suspect in for questioning.

2.1.4 Assessing the impact of comparison on the perceived similarity of faces

A variety of different comparison and control tasks were deployed to examine the influence of comparison on perceived similarity. They are summarised in Table 3. In each experiment (except for Experiment 5, which examined the effect of varying comparison duration) there was a no-comparison control, a similarity-focussed comparison, and a difference-focussed comparison condition. Participants were required to write down three similarities in the similarity-focussed comparison condition, and three differences in the difference-focussed comparison condition. Participants could choose whether to focus on individual features (e.g. differences in nose shape) or on more holistic attributes of the face (e.g. the configuration of facial features, or perceived honesty). Arguably participants were more likely to focus on features when listing commonalities (or differences) between faces, because features are easier to describe than holistic attributes. For example, before listing a holistic difference between two faces that makes one appear more approachable than the other, the observer first needs to retrieve an appropriate adjective (e.g. approachable, friendly, or amiable). The process of finding an appropriate term is presumably easier when referring to an individual feature such as the nose, or the mouth.

To assess whether any effect of comparison on perceived similarity is limited to contexts in which commonalities and differences are explicitly listed participants in Experiment 4 were asked to rate the perceived health of each of the two faces in a pair. It was assumed that participants would compare the faces to come up with an appropriate metric of health. Unlike face pairs in the comparison conditions, face pairs in the no-comparison control condition were only presented once. To assess whether any effects of comparison might be attributable to stimulus pre-exposure (i.e. the fact that stimulus pairs in the comparison conditions were presented twice), participants in Experiment 4 were asked to perform a simple binary age decision that was presumed to not necessitate a comparison. Similarly, participants in Experiments 2 and 7 were asked to describe each individual face in a pair before performing a similarity rating or same–different judgement task.

Whilst there is evidence to suggest that comparison affects the perceived similarity of faces (Boroditsky, 2007; Mundy et al., 2014, 2007; Mundy, Honey, & Dwyer, 2009), it is expected that the perceived similarity between two faces is to a large degree stimulus driven. That is, whilst two faces that share few commonalities might appear more similar following comparison, they are unlikely to appear as similar as two faces that are objectively highly similar, such as identical twins. To control for the influence of differences in stimulus similarity the face pairs used within a given experiment (for experiments with a within-subjects design), or between-subjects condition were matched to be of comparable similarity. The different approaches that were used to match face pairs are summarised in Table 2. To further control for the influence of differences in phenotypic similarity between stimulus pairs, the use of stimulus pairs was counterbalanced across the different within-subjects conditions.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
No-comparison control	✓	✓	✓	✓		✓	✓
Description control		✓					✓
Pre-exposure control				✓			
Similarity-focussed comparison	✓	✓	✓	✓		✓	✓
Difference-focussed comparison	✓	✓	✓	✓		✓	✓
Implicit comparison				✓			
Timed comparison					✓		

Table 3: *The different control and comparison conditions used in the seven experiments presented in the empirical chapters of this thesis.*

Individual differences in the ability to recognise and match faces is another factor that can influence perceived phenotypic similarity (Suchow, 2018). According to this view there is an interdependency between the observed and the observer. On one end of the spectrum are observers with prosopagnosia who have difficulty recognising familiar faces. To individuals with prosopagnosia there are many lookalikes (Suchow, 2018). On the other end are super-recognisers (discussed in section 2.4) who have exceptional face recognition abilities. Super-recognisers are unlikely to encounter doppelgängers, because to them every face is unique (Suchow, 2018). A related argument is that the range of faces a person has experienced can influence the perceived similarity of a given pair of faces. It is an underlying assumption of the face-space framework (discussed in section 2.1.5) that the dimensions of face-space are optimised to discriminate the population of faces that the observer experiences (Valentine, Lewis, & Hills, 2016). When a person is required "to make fine distinctions within a narrow region of face space, the person's visual system adapts, resulting in a warping of face space" (Suchow, 2018 para. 7). The effects of individual differences in ability and exposure history, are likely to interact. In people with prosopagnosia the process of face-space optimisation is arguably less efficient. It follows that relative to a person with normal face-recognition and matching abilities, a person with prosopagnosia's face-discrimination abilities might be less affected by the range of faces they encounter. In Experiments 1–6 the effect of individual differences in ability and experience was controlled for by implementing comparison as a within-subjects variable. This argument does not hold for Experiment 7, because each participant completed only one experimental condition. However, considering the relatively large sample size (there were 45 participants in each of the 3 treatment conditions), it is expected that the range of face-recognition and matching abilities in each treatment condition is likely to reflect the distribution for the population from which the participant sample was drawn.

2.1.5 Measuring perceived similarity

Two dependent measures were used to assess perceived similarity: similarity ratings (Experiments 1–5), and same–different judgements (Experiments 6 and 7). In the similarity rating task participants were asked to rate the perceived similarity between two faces on a scale of 1 (not similar) to 7 (very similar). In the same–different judgment task participants had to decide whether two images depict the same face. It has been argued in section 1.2.3 that similarity ratings and same–different judgments reflect the same underlying similarity construct. Face space constitutes a framework that can be used to model the similarity between faces captured by direct (i.e. similarity ratings) and indirect (e.g. same–different judgments) measures of similarity.

Face space is a multi-dimensional similarity space that represents the phenotypic variation of faces (Valentine, 1991; Valentine et al., 2016). Faces are represented as points in the multidimensional space. The similarity between two faces can be described by their proximity to each other in that space: similar faces are represented close together, dissimilar faces are represented far apart (Valentine, 1991; Valentine et al., 2016). The first publication on the face space framework outlines two variants of face space (Valentine, 1991). In the norm-based coding variant faces are encoded in relation to an abstracted norm, which forms the origin of the multi-dimensional space. The similarity metric (i.e. how similarity maps onto distance in face-space) for the norm-based coding model was postulated to be based on vector similarity (Valentine, 1991). Byatt and Rhodes (1998) defined the vector-based similarity metric for the norm-based model as “the cosine of the angle between the vectors’ representations of two faces (relative to a norm face) divided by the simple distance between the two faces” (Valentine et al., 2016, p. 2002). In the exemplar-based coding model faces are stored as exemplars, and the origin represents the point of maximum exemplar density (Valentine, 1991). The similarity metric consists of the Euclidian distance between exemplars.

Although the two variants differ conceptually they make similar predictions with regards to various phenomena in face perception. For example, both versions of the framework can account for the well-established finding that faces rated as distinctive are better recognised than faces that are rated to be typical (Cohen & Carr, 1975; Ellis, 1991; Going & Read, 1974; Light, Kayra-Stuart, & Hollander, 1979; Valentine & Endo, 1992; Wickham, Morris, & Fritz, 2000). Both models assume that the ease with which a given face can be discriminated depends on the number and the proximity of neighbouring faces in face space (Byatt & Rhodes, 1998). It follows that recognition memory should be superior for distinctive faces because they occupy less densely populated areas of face space than typical faces, and are therefore less likely to be confused for other faces. In addition to accounting for the effects of distinctiveness on face recognition, the face space framework offers a unified account of the effects of inversion, ethnicity, adaptation, and caricature on face recognition (for a review see Valentine et al., 2016).

Face space has mainly been used to explain performance on tasks that have a memory component (for example, recognising a previously seen face). It can be argued that the simultaneous face comparisons are likely to recruit the same stimulus dimensions than those that are involved when matching a memory representation to a face. One criticism of the face space framework is that neither the nature nor the number of dimensions that constitute face space are defined (Johnston, Milne, Williams, & Hosie, 1997). There is evidence to suggest that the dimensions of face space include face shape, hair length, and age (Shepard, Davies and Ellis, 1981, as cited in Davies & Young, 2017). Other potential dimensions are face width, inter-ocular distance, and gender (Valentine et al., 2016). A study that investigated which features are most critical to face identification found that the five (out of a possible 20) face space dimensions with the highest discriminative power are—lip thickness, eye colour, eye shape, eyebrow thickness and hair colour (Abudarham & Yovel, 2016). The main shortcoming of this study is that only focussed on features, other potentially critical dimensions such as the overall friendliness of a face were not considered. Despite these efforts to identify the nature of the underlying dimension the predictive powers of face space are limited because it remains under-defined (Davies & Young, 2017).

As discussed in section 2.1.4 the main limitation of the measures used to assess perceived similarity in Experiments 1–7 is that they were conflated with the pictorial difference manipulation. The perceived similarity of stimulus pairs that had either one pictorial difference or no pictorial differences was measured using similarity ratings. The perceived similarity of stimulus pairs that had two or more pictorial differences was measured using same–different judgements. It is possible that the two tasks may have led participants to focus on somewhat different dimensions of face space. Potential future studies that address these issues are discussed in the final chapter of this thesis.

2.2 Can the findings obtained in Experiments 1–7 be generalised to comparisons that do not involve faces?

It has been argued that "the cognitive demands of face perception differ from most other forms of non-face object recognition" because face recognition necessitates the individuation of many visually similar exemplars under a wide range of viewing conditions (Behrmann, Richler, Avidan, & Kimchi, 2015, p. 758). This begs the question: To what degree can the findings from studies that deploy face stimuli be generalised to other complex visual stimuli? The answer to this question depends on several interrelated factors. The first factor is whether, and to what degree, face processing is unique. If faces are processed just like any other visual object then the findings from research that uses face stimuli can be extrapolated to other types of visual stimuli. However, there is evidence to suggest that faces are processed more holistically than other visual objects. The notion that face stimuli constitute a special visual category because faces are processed holistically, and how it impacts on research that uses unfamiliar faces is examined in Section 2.2.1. The second factor that determines the degree to which the findings obtained in Experiments 1–7 can be extrapolated to non-face stimuli is whether holistic processing plays a central role in simultaneous face matching. That is, whether, and to what degree, performance on a simultaneous face matching task is dependent on holistic processing. Section 2.2.2 reviews studies that have examined the relationship between individual differences in holistic processing abilities and performance on face recognition and simultaneous face matching tasks. The third factor, which is examined in Section 2.2.3, is the degree to which the unique properties of face processing are limited to familiar faces. It is possible, that whilst familiar faces constitute a special stimulus category, images of unfamiliar faces are processed like any other complex visual stimulus.

2.2.1 Are there unique visual mechanisms for processing faces?

It has been suggested that the question whether faces are special essentially refers "to whether there are unique visual mechanisms for processing identity-related information in faces as compared to other objects" (McKone & Robbins, 2012, p. 149). According to this view, the existence of a distributed face-individuation neural network that encompasses several face-selective brain areas (Behrmann, Scherf, & Avidan, 2016; Nestor, Plaut, & Behrmann, 2011; Zhao, Zhen, Liu, Song, & Liu, 2018)—including the fusiform face area, and the occipital face area (Huang et al., 2014; Kanwisher & Yovel, 2006; Pitcher, Walsh, & Duchaine, 2011; Weibert & Andrews, 2015)—constitutes evidence that faces are special. However, studies that have deployed high-resolution functional magnetic resonance imaging (fMRI) have found that the fusiform face area consists of heterogeneous sub-regions, some of which are highly selective to non-face object categories, including cars and birds (Grill-Spector, Sayres, & Ress, 2006; McGugin, Gatenby, Gore, & Gauthier, 2012;

McGugin, Newton, Gore, & Gauthier, 2014; Schwarzlose, 2005; see also Xu, 2005). These findings lend support to the argument that processing in face-selective areas extends to other object categories that are difficult to individuate (Gauthier, Skudlarski, Gore, & Anderson, 2000; Harley et al., 2009; McGugin et al., 2014).

A second argument in support of the notion that faces constitute a special visual category posits that face recognition relies on a unique processing style. According to this view faces, unlike other objects, are not processed as a collection of discrete features, but holistically (Farah et al., 1998; McKone & Robbins, 2012). For simplicity, the term holistic processing is used here as an umbrella term for holistic, second-order relational, and configural processing. However, it is important to note that these terms are not strictly synonymous, but reflect different theoretical accounts (Richler, Palmeri, & Gauthier, 2012). Holistic processing has been used to refer to the notion that faces are processed as undifferentiated wholes (Farah et al., 1998; Young, Hellawell, & Hay, 1987). According to this view face recognition requires little part decomposition (Farah et al., 1998). Another definition posits that holistic processing operates on representations that capture the spatial relationships between facial features (Diamond & Carey, 1986; Leder & Bruce, 2000). According to this view both features, and the spatial relationships between them (e.g. nose-mouth distance), are used in face recognition. What distinguishes face from object recognition is that face recognition is much more reliant on relational information (Diamond & Carey, 1986). Yet another interpretation argues that holistic processing emerges from the interactive processing of featural and relational information (Kimchi & Amishav, 2010). One important commonality between these definitions is that the processing of individual features is not considered sufficient for effective face recognition. Furthermore, it has been argued that "most researchers would probably agree with a definition of holistic face processing as the simultaneous integration of the multiple features of a face into a single perceptual representation" (Bruno Rossion, 2008, p. 275).

There are four lines of evidence to suggest that faces are processed holistically. First, an individual facial feature, for example the mouth, is recognised more accurately when it is presented in the context of a face, than when it is presented in isolation (Tanaka & Farah, 1993; for a review see Tanaka & Simonyi, 2016). This part-whole effect has been interpreted as showing that face representations correspond to entire faces, not individual features (Behrmann et al., 2015). A similar part-whole effect is not seen for non-face objects, such as houses, or for inverted faces (Tanaka & Farah, 1993). The second line of evidence is supplied by studies in which participants were asked to decide whether the top (or bottom) halves of two sequentially presented composite faces are the same or not (Behrmann et al., 2014; for reviews see Murphy, Gray, and Cook, 2017, and Rossion, 2013). Research on composite task performance has shown that naming latencies and same–different judgments are slower and more error prone when the two face-halves are aligned than when they are misaligned (Hole, 1994; Young et al., 1987). This difference in performance has been attributed to the processing of the task-irrelevant half of the aligned composite faces (Bruno Rossion & Boremanse, 2008). According to this interpretation the task-irrelevant half interferes with performance in the aligned composite condition because its presence alters the appearance of the task-relevant half (Murphy et al., 2017). For misaligned composite faces the task-irrelevant half has little effect on task performance, because misaligned face halves are not processed as whole faces.

The third line of evidence that face processing is likely to be predominantly holistic is provided by the finding that the detrimental effect of inversion on recognition performance (Jolicoeur, 1985; Tarr & Pinker, 1989), commonly attributed to the disruption of holistic processing (Farah et al., 1995; Van Belle, De Graef, Verfaillie, Rossion, & Lefèvre, 2010), is disproportionally greater for faces than for other objects (Robbins & McKone, 2007; B Rossion & Curran, 2010; Valentine & Bruce, 1986; Yin, 1969; Yovel & Kanwisher, 2004). However, a study in which participants were trained to name upright faces and ‘greebles’ (a category of novel, homogenous objects that was designed to be similar to faces) found a comparable inversion effect for greebles and faces during a subsequent naming test (Ashworth et al., 2008). The authors conclude that “the inversion effect is largely driven by properties of the stimulus categories and not the stimulus category per se” (Ashworth et al., 2008, p. 780). The relationship between holistic processing and expertise with visually homogenous categories is further discussed below.

The fourth line of evidence suggesting that faces are processed holistically is that the selective impairment in face processing seen in prosopagnosia—“a visual agnosia characterised by the inability to recognise familiar faces” (Barton, 2009, p. 242)—is associated with an impairment in holistic processing (Avidan, Tanzer, & Behrmann, 2011; Kimchi, Behrmann, Avidan, & Amishav, 2012; T. T. Liu & Behrmann, 2014; Lobmaier, Bölte, Mast, & Dobel, 2010; Palermo et al., 2011;

Ramon, Busigny, & Rossion, 2010; Rivest, Moscovitch, & Black, 2009; Saumier, Arguin, & Lassonde, 2001). For example, individuals with prosopagnosia show a small inverse of the part-whole face advantage—unlike controls they are somewhat better at recognising facial features in isolation than as part of a face (Busigny, Joubert, Felician, Ceccaldi, & Rossion, 2010; Ramon et al., 2010). Similarly, individuals with prosopagnosia show no performance advantage for misaligned over aligned faces on the composite task (Avidan et al., 2011; Busigny et al., 2010; Palermo et al., 2011; Ramon et al., 2010), nor do they show the inversion effect (Behrmann, Avidan, Marotta, & Kimchi, 2005; Busigny et al., 2010; Busigny & Rossion, 2010). However, some individuals with developmental (Biotti et al., 2017; Le Grand et al., 2006; Susilo et al., 2010) and acquired (Finzi, Susilo, Barton, & Duchaine, 2016) prosopagnosia have normal holistic processing abilities. These findings have been interpreted as evidence that prosopagnosia is a heterogeneous condition which can be caused by impairments in feature-based processing (Le Grand et al., 2006; Susilo et al., 2010).

It has been shown that the holistic processing deficit associated with prosopagnosia can also impact on the processing of non-face objects (Barton, 2009; Barton & Cherkasova, 2005; De Gelder, Bachoud-Lévi, & Degos, 1998; De Gelder & Rouw, 2000; Tanzer, Freud, Ganel, & Avidan, 2013). Further support for the notion that rather than being a unique attribute of faces, holistic processing also affects other objects is provided by the finding of a part-whole effect for non-face objects (Tanaka & Farah, 1993; Tanaka & Gauthier, 1997; but see Richler, Mack, Palmeri, and Gauthier, 2011). Similarly, there is evidence to suggest that faces are not the only stimulus class that is disproportionately affected by inversion. An inversion effect comparable to that for faces has been found when dog experts were asked to identify dogs (Diamond and Carey, 1986), for budgerigar experts (Campbell & Tanaka, 2018), for handwriting experts (Bruyer & Crispeels, 1992), and for expert radiographers who were asked to identify mammograms (Chin, Evans, Wolfe, & Tanaka, 2017). Based on these findings it has been argued that rather than being unique to faces, the inversion effect results from expertise with visually homogenous objects that require individuation at the identity level (Ashworth et al., 2008; Campbell & Tanaka, 2018; Gauthier & Tarr, 1997). To test this argument, Gauthier and her colleagues conducted a series of studies in which participants were trained to identify greebles—novel objects that share a common relational structure (Gauthier & Tarr, 1997, 2002;

Gauthier, Williams, Tarr, & Tanaka, 1998). The studies found that becoming a greeble expert entails a shift in processing from feature-based to more holistic. Further support for the notion that holistic processing reflects expertise comes from studies on the own-race bias. These studies have found that other-race faces—with which participants have less familiarity—are processed less holistically than own-race faces (Michel, Caldara, & Rossion, 2006; Rhodes, Brake, Taylor, & Tan, 1989; Tanaka, Kiefer, & Bukach, 2004).

In sum, the findings reviewed above suggest that face processing is somewhat special in that it relies more heavily on holistic processing relative to other objects. However, it is important to note that the difference between face and object processing are likely to be quantitative rather than qualitative. First, although holistic processing is considered a hallmark of face processing, there is evidence to suggest that feature-based processing plays an equally important role (Amishav & Kimchi, 2010; Cabeza & Kato, 2000; Hayward, Rhodes, & Schwaninger, 2008; Schwaninger, Lobmaier, Wallraven, & Collishaw, 2009; Schwarzer & Massaro, 2001). Second, it has been argued that holistic representations "emerge spontaneously in self-organizing competitive systems in regions of the input space where many exemplars are seen" (Wallis, 2013, p. 19/21). It follows that holistic representations are likely to arise for faces, and for other homogenous objects that require visual expertise to discriminate (Bruyer & Crispeels, 1992; Diamond & Carey, 1986; Gauthier & Tarr, 1997, 2002; Tanaka & Gauthier, 1997).

2.2.2 Are individual differences in holistic processing likely to impact upon simultaneous face matching performance?

Research on the impact of individual differences in holistic processing on face perception has mainly focussed on face recognition. The majority of studies that have investigated the relationship between holistic processing and face-recognition abilities have found a significant positive correlation between the two variables (Avidan et al., 2011; DeGutis, Mercado, Wilmer, & Rosenblatt, 2013; DeGutis, Wilmer, Mercado, & Cohan, 2013; Engfors, Jeffery, Gignac, & Palermo, 2017; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012; Richler, Cheung, & Gauthier, 2011; Wang, Li, Fang, Tian, & Liu, 2012; but see Konar, Bennett, and Sekuler, 2010, Richler, Floyd, and Gauthier, 2015, Sunday, Richler, and Gauthier, 2017, and Verhallen et al., 2016). These findings have been interpreted as showing that there is a predictive link between holistic processing and face recognition abilities (DeGutis, Wilmer, et al., 2013; Yovel, Wilmer, & Duchaine, 2014).

There is little evidence for a similar predictive link for simultaneous face matching performance. A study that used both the face inversion effect, and performance on the part-whole task to measure holistic processing found a positive correlation between holistic processing and face matching abilities (Rezlescu, Susilo, Wilmer, & Caramazza, 2017). However, it is important to note that the tests used to measure the face inversion effect and simultaneous face matching abilities both used inverted faces. It is likely that the moderate positive correlation between the inversion effect and simultaneous face matching performance is attributable, at least in part, to this commonality. Two studies that measured holistic processing using the composite task found no correlation between the two variables (Konar et al., 2010; Rezlescu et al., 2017). These findings suggest that variations within the normal range of face matching performance are determined, at least in part, by individual differences that are not specific to faces. Another factor that is likely to contribute towards performance on face matching tasks is the fact that the faces used in matching studies are generally not familiar to the participants. The differences between familiar and unfamiliar face processing, and how they might impact on face matching performance is discussed in the next section.

2.2.3 The relationship between face matching performance and face familiarity

People can recognise familiar faces with ease, despite variations in expression, pose, lighting, hairstyle, and the presence or absence of glasses (Hancock et al., 2000). Furthermore, we can recognise familiar faces even from low quality images or video footage, and when facial features are obscured by beards, sunglasses, hats and other disguises (for reviews see Bruce, 2012, and Johnston and Edmonds, 2009). Unfamiliar face identification, on the other hand, has been shown to be error prone (for reviews, see Hancock et al., 2000, and Johnston and Edmonds, 2009). Performance differences between familiar and unfamiliar face processing are not limited to contexts that involve a memory component. For example, a series of experiments on face learning has shown that face matching speed increases with familiarity (Ruth Clutterbuck & Johnston, 2002, 2004, 2005). However, whilst there is ample experimental evidence that simultaneous unfamiliar face matching is error prone (Bruce et al., 1999; Bruce, Henderson, Newman, & Burton, 2001; Davis & Valentine, 2009; Henderson et al., 2001; Kemp et al., 1997; Megreya & Burton, 2006, 2008; Megreya et al., 2013), few studies have directly compared familiar and unfamiliar face matching performance for faces that are personally familiar to the participants. The scarcity of research comparing familiar and unfamiliar face matching is possibly attributable to the fact that familiar face matching is perceived to be an easy task that is likely to generate ceiling effects.

A study that deployed low-quality video targets to increase the difficulty of the face matching task found that whilst individuals who were familiar with the target faces performed at 90% accuracy, individuals who were not familiar with the target faces performed at 70% accuracy (Bruce, Henderson, Newman, & Burton, 2001, experiment 1). Perhaps the most compelling illustration of the difference between familiar and unfamiliar face matching performance comes from a study that asked participants to cluster photos of two individuals (Jenkins et al., 2011). The images used for each of the individuals consisted of a representative sample, which was collated to reflect the variability of photos of the same person. Participants who were unfamiliar with the individuals depicted clustered the 40 images into eight distinct groups on average (Jenkins et al., 2011, experiment 1; see also Andrews et al., 2015). Participants who were familiar with the individuals depicted, on the other hand, correctly sorted the 40 images into two groups (Jenkins et al., 2011, experiment 2).

One possible explanation for the differences between unfamiliar and familiar face matching performance is provided by an influential theoretical account of face recognition, which argues that unfamiliar and familiar face processing rely on different mental representations (Bruce & Young, 1986). The theory posits that the mental representations that are recruited during familiar face recognition consist of structural codes (Bruce & Young, 1986), which incorporate multiple visual impressions of a given face (Burton, Jenkins, & Schweinberger, 2011). It is currently unclear whether these structural codes consist of a refined average of a given face (Etchells, Brooks, & Johnston, 2017; Kramer, Ritchie, & Burton, 2015), or whether the different visual impressions of a face are encoded separately (Longmore et al., 2008). According to Bruce and Young's (1986) theory of face recognition, unfamiliar face processing operates on pictorial representations. These pictorial codes reflect the visual properties of the experience, which can include image specific information such as lighting and pose (Bruce & Young, 1986). An extreme interpretation of this view posits that unfamiliar faces are processed as visual patterns, and are matched just like any other visual object (Burton, White, & McNeill, 2010; Hancock et al., 2000; Megreya & Burton, 2006). The ability to accommodate differences between the pictorial representations of a face emerges with familiarity (Jenkins et al., 2011).

The process of face familiarisation "involves a transition during learning from relying largely on pictorial codes, to relying largely on structural codes" (Burton et al., 2011, p. 946). An illustration of this gradual transition from unfamiliar (view-dependent, pictorial) to familiar (view-independent) face processing is provided by studies that have compared event-related brain potentials (ERPs) elicited by the repeated presentation of familiar and unfamiliar faces. Studies that have measured the N250r ('r' for repetition) component in event-related brain potentials have found that whilst repetition of both familiar and unfamiliar faces can evoke a N250 response only familiar faces generate a robust N250r ERP signal when the two sequentially presented images differ in lighting, camera angle, expression, or are visually distorted (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Kaufmann, Schweinberger, & Burton, 2009; Zimmermann & Eimer, 2013). The view-independence of the N250r is related to person recognition, and emerges over time as a given face becomes more familiar (Bindemann et al., 2008; Zimmermann & Eimer, 2013). This finding suggest that unfamiliar face matching is based on view-dependent pictorial codes (Zimmermann & Eimer, 2013).

In sum, there is evidence to suggest that there are qualitative differences between familiar and unfamiliar face processing (Megreya & Burton, 2006; for a review see Johnston & Edmonds, 2009). An influential theoretical account of face-processing suggests that whereas familiar faces are encoded structurally, the representations of unfamiliar faces are pictorial (Bruce & Young, 1986). An implication for the current research is that the findings from Experiments 1–7 are likely to generalise to other complex visual objects that are processed pictorially.

2.2.4 Conclusion

The findings obtained in Experiments 1–7 for unfamiliar face stimuli are likely to generalise to other visual stimuli for the following three reasons. First, the research reviewed above suggests that the holistic processing bias that has been observed for face stimuli constitutes a quantitative rather than a qualitative difference between face and object processing. Second, the fact that there is little evidence for a correlation between unfamiliar face matching accuracy and holistic processing suggests that unfamiliar face matching performance depends on cognitive and perceptual processes that are not unique to faces. Finally, unlike familiar face processing, which has been shown to be largely unaffected by changes in view and lighting, unfamiliar face processing has been shown to be sensitive to pictorial differences.

2.3 Ecological validity

2.3.1 Face matching practices in applied contexts

According to Edmond and Wortley (2016) “the ability to identify individuals from images seems to be increasingly important for policing and national security, especially border control” (p. 486). This is a concerning development considering that there is ample experimental evidence that simultaneous unfamiliar face matching is error prone (Bruce et al., 1999, 2001; Davis & Valentine, 2009; Henderson et al., 2001; Kemp et al., 1997; Megreya & Burton, 2006, 2008; Megreya et al., 2013). For example, studies that have deployed a 1-in-10 face matching task have found error rates between 20% and 30% when participants were asked to match faces with neutral expressions shown from a frontal view (Bruce et al., 1999; Megreya & Burton, 2008; Megreya et al., 2013). A further drop in accuracy was found when the target face wore a hat (Henderson et al., 2001 experiment 3), or was presented at a different angle to the faces in the array (Bruce et al., 1999). Finally, when the difficulty of the face matching task was increased by using target faces captured from poor quality CCTV footage, accuracy on a 1-to-8 face matching task was found to be only slightly above chance (Henderson et al., 2001).

Simultaneous face matching is a critical task in a variety of professional contexts (see Davis & Valentine, 2015). Passport officers routinely verify identities by matching the face of a person standing in front of them to their photo ID. Police officers identify Persons Of Interest (POIs) by comparing images taken from a crime scene with those of known suspects. CCTV camera operators match images of a target under surveillance to live CCTV footage. Most comparisons in these applied contexts are limited to single identity matches. In an experiment designed to mimic the setup commonly encountered during passport checks, participants were asked to compare a photograph depicting a face with a neutral expression shown from a frontal view (i.e. resembling a passport photo), to either a second photograph of a neutral, frontal-view face, or to a live-person (Megreya & Burton, 2008, experiment 3). The study found an error rate of approximately 15% in both experimental conditions. Unfamiliar face matching accuracy on paired-matching tasks is negatively affected by pictorial variations between passport-style photos that were captured at different points in time (Megreya et al., 2013). Accuracy on a paired matching task dropped from 90% when the two photographs were captured on the same day, to 70% when the two images were taken months apart (Megreya et al., 2013, experiment 2).

The findings reported above are unlikely to represent an accurate picture of the prevalence of face matching errors in applied contexts, because they were obtained under experimental conditions that controlled for the influence of extraneous variables. A wide variety of additional factors are likely to impact on face matching performance in real world scenarios, not all of them negatively. For example, professionals who perform comparisons as part of their jobs are arguably more likely to be motivated to perform the task well, than research participants. A study that tested whether the motivation to perform well can improve face matching accuracy, found that offering a food incentive for above average performance improved accuracy on mismatch trials relative to a no-reward control condition (Moore & Johnston, 2013). Factors that may impact negatively on performance include demands that are introduced through task repetition, time pressure, and the need to perform multiple tasks simultaneously (Alenezi & Bindemann, 2013; Alenezi et al., 2015; Wirth & Carbon, 2017). An individual's emotional state can also affect their performance (Attwood et al., 2013; Megreya & Bindemann, 2013). For example, a study that manipulated anxiety levels by administering CO2 found that acute anxiety decreased performance on match trials, relative to a no anxiety control (Attwood et al., 2013).

The availability of contextual information introduces a range of cognitive and social biases which may affect the outcome of a face matching decision. The term forensic confirmation bias has been used to "summarize the class of effects through which an individual's pre-existing beliefs, expectations, motives, and situational context influence the collection, perception, and interpretation of evidence during the course of a criminal case" (Kassin, Dror, & Kukucka, 2013, p. 45). These biases can result in a form of tunnel vision that is focussed on confirmatory evidence (Findley & Scott, 2006). For example, participants who were asked to form an initial hypothesis in the presence of weak circumstantial evidence were found to focus on finding, and interpreting additional evidence in line with their hypothesis (O'Brien, 2009; see also Hill, Memon, and McGeorge, 2008, Kassin, Goldstein, and Savitsky, 2003). Similarly, police officers' initial beliefs have been shown to affect their interpretation of evidence (Charman, Kavetski, & Mueller, 2017). The belief that two images depict the same person might lead police officers and other professionals who regularly perform face comparisons to focus on commonalities between the two faces, and to overlook or discount differences.

The effect of the forensic confirmation bias on the interpretation of photographic evidence for person identification is discussed in section 2.3.2. The section also explores how common practices in a variety of jurisdictions impact on the reliability of image-based person identifications. The potential impact of biases (e.g. the own-ethnicity bias) and work demands (e.g. the need to perform multiple checks simultaneously) on the accuracy of identity verifications performed at passport control are discussed in section 2.3.3. Automatic face recognition software is used to assist facial examiners in a variety of tasks, including the detection of fraudulent passport applications. Section 2.3.4. evaluates the capabilities of face recognition algorithms, and explores how the use of automatic face identification systems impacts on human face matching performance. Implications for the ecological validity of the research presented in the empirical chapters of this thesis are outlined in section 2.3.5.

2.3.2 Using images to identify persons of interest in legal and forensic contexts

English and Welsh courts permit a variety of witnesses—including police officers, members of the public, members of the jury and experts—to identify persons of interest (POIs) in images (Edmond & Wortley, 2016). Some of these witnesses are not familiar with the POI. It is, for example, permissible for a witness who does not know the defendant to give evidence of recognition, provided the witness has studied images of the culprit over an extended period (for example, by repeatedly watching a CCTV recording), and compared them to a contemporary photograph of the defendant

(Edmond & Wortley, 2016). There is evidence to suggest that the availability of multiple images of the same person can improve unfamiliar face matching accuracy (Bindemann & Sandford, 2011; Dowsett, Sandford, & Burton, 2016; Matthews & Mondloch, 2017; Menon, White, & Kemp, 2015). For example, in an experiment in which participants were asked to match a target identity to one of 30 faces, performance was found to improve when the participants had access to multiple images of the target identity (Dowsett et al., 2016). Access to multiple images is particularly beneficial if the images showcase high within-person variability (Baker, Laurence, & Mondloch, 2016; Menon et al., 2015; Ritchie & Burton, 2017). For example, participants were found to be more accurate on a face matching task when they had access to two high-variability images, relative to both a single image condition, and a low-variability image pair condition (Menon et al., 2015).

It has been argued that the transition from unfamiliar to familiar face processing is facilitated when individuals have access to multiple images of the same face (Andrews et al., 2015). Yet, it is important to note that although the availability of multiple images improves face matching performance, neither familiarisation prior to the matching task, nor the availability of multiple images during the task eliminates errors. Participants who were exposed to high within-person variability in a name-learning phase outperformed those who were shown low variability images on a subsequent face matching task that used high quality images (Ritchie & Burton, 2017, experiment 2). However, accuracy on match trials in the high variability group was still lower than the above 90% accuracy rate that has been observed for familiar, low-quality targets (Bruce, Henderson, Newman, & Burton, 2001, experiment 1). In sum, although the availability of multiple images can improve matching performance, accuracy rates will be lower than for individuals that are personally familiar to the observer. Furthermore, accuracy rates are likely to be even lower in applied settings, because the person performing the matching task is likely to be privy to incriminating information about the target. The impact of contextual knowledge and biases on face-matching accuracy is further discussed below.

There is evidence to suggest that contextual information can bias jurors' perceptions of visual information (Dror et al., 2005; Jones, Crozier, & Strange, 2017; Kukucka & Kassin, 2014; Smalarz, Madon, Yang, Guyll, & Buck, 2016). Participants who were given incriminating information about the existence of a confession were found to be more likely to erroneously conclude that two handwriting samples were written by the same person, than participants who were not exposed to the biasing information (Kukucka & Kassin, 2014). Similarly, providing participants with contextual information about a violent crime has been shown to increase the likelihood that two ambiguous fingerprints are judged to be a match, relative to a more neutral context, in which the crime did not result in any injury or physical harm (Dror et al., 2005). Finally, a study that investigated the role of criminal stereotypes on fingerprint matching accuracy found that the presence of stereotypical information (e.g. information about gender, age, and education level that conforms to stereotypes about a given crime profile) increased the likelihood that two fingerprints were erroneously judged to be a match (Smalarz et al., 2016).

Jurors in Australia are permitted to compare images of the culprit to the person in the dock (Davis & Valentine, 2015). Similarly, provided the images of the culprit are considered sufficiently clear, jurors in England and Wales are permitted to compare them to the defendant (Edmond & Wortley, 2016). However, there is evidence to suggest that irrespective of image clarity, the presence of incriminating evidence can bias the jurors' perceptions of the likeness between the defendant in the dock and the perpetrator captured on camera are the same person (Bressan & Dal Martello, 2002; Charman, Gregory, & Carlucci, 2009). Facial similarity ratings of adult-child pairs were found to be higher when participants believed that the pairs were genetically related (i.e. that they were parent and child), than when they believed that they were unrelated. Belief about relatedness had a stronger effect on perceived similarity than actual relatedness (Bressan & Dal Martello, 2002). Similarly, participants in a mock investigation who were told that two witnesses had identified the suspect assigned higher similarity ratings to the suspect and a composite image of the culprit, than participants who were not given any additional information (Charman et al., 2009). Conversely, participants who were told that the witnesses had identified a different suspect provided lower similarity ratings, than the no-contextual information control (Charman et al., 2009). The images deployed in both studies were of a high quality, and the faces was clearly visible.

In Australia, police officers are only permitted to make image based identifications if they are sufficiently familiar with the individual (Edmond, Davis, & Valentine, 2015). This is a sensible precaution considering that police officers, judges, and other forensic and legal experts are subject to the same cognitive biases as members of the public. An archival analysis of DNA exonerations found that false confessions were frequently associated with additional errors in forensic science (Kassin, Bogart, & Kerner, 2012). The notion that the presence of a confession played a causal role in generating these errors is supported by the observation that false confessions were significantly more likely to precede than to succeed forensic-science errors (Kassin et al., 2012). Similarly, a study that examined real world cases in which fingerprints were misidentified found that violent crimes (i.e. murders and rapes) were overrepresented (Cole, 2005). The overrepresentation was not attributable to an increased likelihood to use fingerprint evidence in these cases (Cole, 2005). Finally, the presence of incriminating information, and the belief that a suspect is guilty have both been shown to bias expert decision making. A study that provided misleading contextual information to fingerprint experts about fingerprints they had previously identified as a match, found that most of the experts reversed their initial decision (Dror, Charlton, & Péron, 2006). Similarly, police officers' initial beliefs about a suspect's guilt has been shown to affect how they evaluate additional ambiguous evidence (Charman et al., 2017). "The more likely they were to believe the suspect was guilty, the more incriminating they perceived subsequent ambiguous evidence to be" (Charman et al., 2017, p. 198).

The Australian police sometimes deploys experts to perform image-based identifications (Edmond & Wortley, 2016). In general, "the opinions of those presented as experts is only admissible if the expertise 'fits' with the task of comparing faces or body features and remains restricted to the description of similarities" (Edmond & Wortley, 2016, p. 500). In England and Wales experts with facial mapping skills are permitted to give opinion evidence of identification (Edmond & Wortley, 2016). They are generally recruited in cases where the images of the scene are of a low quality, or the POI is not known to either the police or eye witnesses (Edmond & Wortley, 2016). Morphological comparison is the most commonly deployed facial mapping technique. It involves a side-by-side comparison between images of the culprit taken from the scene (which may be enhanced), and a contemporary photograph of the POI. Sometimes composite images and superimposition techniques (such fading one face into another, or wiping between two faces) are used to facilitate the analysis (Edmond et al., 2015). However, there is evidence to suggest that these techniques can bias the outcome of the comparison towards 'same' responses (Strathie & McNeill, 2016; Strathie, Mcneill, & White, 2012).

The accuracy of expert identifications could be improved by asking several examiners to perform a given judgement. Studies that have compared individual scores to aggregate scores that combine the data of several individuals, have shown that combining individual responses improves face matching accuracy (White, Burton, Kemp, & Jenkins, 2013; White, Phillips, et al., 2015; see also Jeckeln, Hahn, Noyes, Cavazos, and O'Toole, 2018). Near perfect face matching performance on the Glasgow Face Matching Test (GFMT) was found when the ratings of eight or more individuals were combined (White et al., 2013). However, aggregate performance on face matching tasks that were challenging because they used inverted, or negated versions of difficult-to-discriminate face stimuli, was found to be substantially below 100% (White et al., 2013). An upper limit of crowd accuracy was reached when the scores of 10 individuals were combined. "Therefore, it appears that while performance on unfamiliar face matching tasks is improved by combining responses of non-expert populations, there is little benefit to combining responses of larger groups" (White et al., 2013, p. 774). The findings obtained for negated and inverted faces suggest that it might be preferable to combine the aggregate the scores of expert examiners because they often evaluate images in which both the quality, and the amount of available phenotypic information is low. Combining the positive effects of response aggregation and expertise might result in a significant improvement in the accuracy of expert identifications.

A study that assessed the benefits of data aggregation for expert facial examiners, and two groups of non-experts (students, and professional controls) found that aggregation improved performance in all three groups (White, Phillips, et al., 2015). Near perfect performance was found in all three groups for group sizes of eight or more. However, the aggregate scores for facial examiners were higher than those for the two control groups at all sample sizes. Furthermore, the performance of facial examiners plateaued sooner than that of the other two participant groups (White, Phillips, et al., 2015). An examination of the aggregated judgment data for the different groups (see figure 4 in White, Phillips, et al., 2015, p. 5) suggest that the reliability, and value of identifications made by forensic facial examiners could be significantly improved by combining the scores of two or three facial examiners.

2.3.3 Using images to verify identity for border control

Passport photos are taken under uniform lighting, against a neutral background, with the subject facing the camera and adopting a neutral facial expression. However, face matching performance is error prone, even for faces that were photographed on the same day, from the same angle, with the same facial expression (Bindemann, Avetisyan, & Blackwell, 2010; Burton et al., 2010; Megreya & Bindemann, 2009; Megreya & Burton, 2008). Furthermore, because passports only need to be renewed once every 10 years, time-related changes in appearance are likely to have a negative impact on face matching accuracy in border control settings. Supporting evidence is provided by the findings of a study that compared face matching accuracy for passport-style images that were taken between 10 months and two years apart, to face matching accuracy for images that were captured on the same day (Megreya et al., 2013). The study found that accuracy was significantly lower when the compared images were captured months apart (Megreya et al., 2013).

Border Force agents match photographic ID to a person standing in front of them. This task has been shown to be error prone (Davis & Valentine, 2009; Kemp et al., 1997; Megreya & Burton, 2008). In particular, the presence of a live person has been shown to increase the likelihood that two faces are erroneously judged to be the same (Megreya & Burton, 2008). This bias towards ‘same’ responses is likely to be amplified in the context of border control for two reasons. First, repetitively performing a face matching task, as Border Force Agents are required to do, has been shown to decrease performance accuracy on mismatch trials (Alenezi & Bindemann, 2013; Alenezi et al., 2015; Fysh & Bindemann, 2017). Second, unlike participants in experiments that assess face matching abilities, Border Force agents need to check multiple pieces of information when performing identity checks, and interact with the person whose identity they are verifying. A study that assessed whether embedding a passport style photograph in an ID document affects face matching accuracy found that participants were more likely to make ‘same’ responses in the passport frame condition, relative to a no-frame control condition (McCaffery & Burton, 2016). The study further found that participants were poorer at spotting inaccurate biographical details when the two faces were a match. This finding suggests that laboratory-based face matching experiments may overestimate the degree to which fraudulent IDs are successfully detected in applied contexts (McCaffery & Burton, 2016).

The accuracy of face matching performance in verifying identity for border control may be affected by the own-ethnicity bias. The own-ethnicity bias refers to the finding that the faces of individuals who have the same ethnicity than the observer tend to be better remembered (i.e. they generate more hits and fewer false alarms) than the faces of individuals who belong to a different ethnic group (for a review see Meissner & Brigham, 2001). The two main causal explanations for the own-ethnicity bias are insufficient exposure to, and category-based encoding of, other-ethnicity faces (for a review see Wilson, Hugenberg, & Bernstein, 2013). The own-ethnicity bias is not limited to face-recognition tasks. Several studies have demonstrate that it also affects performance on simultaneous face matching tasks (Megreya, White, & Burton, 2011; Meissner, Susa, & Ross, 2013; Sporer, Trinkl, & Guberova, 2007). These finding suggest that Border Force agents might perform less accurately when verifying the identities of individuals who belong to a different ethnic group to them. However, there is evidence to suggest that the performance of passport officers could be improved through targeted individuation training (McGugin, Tanaka, Lebrecht, Tarr, & Gauthier, 2011; Tanaka & Pierce, 2009). In particular, it has been shown that training individuals to learn the names of other-ethnicity faces improved performance on a two-alternative forced-choice discrimination task for novel faces of the trained ethnicity (McGugin et al., 2011). Feedback on correct performance is crucial for individuation training to be effective (Yovel et al., 2012). It follows that mere exposure to faces from different ethnicities during day-to-day working is unlikely to be sufficient to enhance performance.

2.3.4 Use of automated systems in facial recognition

A common use for automated face recognition systems is to search for a face in a large database of facial images to detect duplicates (White, Dunn, et al., 2015). These many-to-one applications have been successfully deployed for fraud detection in the context of issuing passports and driving licenses. For example, between 2010 and 213 the New York Department of Motor Vehicles (DMV) investigated 13,000 cases of suspected identity fraud (O'Toole & Phillips, 2015). The suspected cases were raised by an automated face recognition system that searched for potential matches in a database containing more than 20 million photos. Of the 13,000 suspected cases, 2,500 resulted in arrests, and 5000 resulted in another form of administrative action (O'Toole & Phillips, 2015). Automated face recognition systems are also used to search databases of known offenders (Towler, Kemp, & White, 2017; White, Dunn, et al., 2015).

Since 2002 the performance of face-recognition algorithms that drive automated face recognition systems has been assessed by the US National Institute of standards and Technology (NIST) (O'Toole & Phillips, 2015; Phillips & O'Toole, 2014). The NIST challenges follow a standard protocol. The first step in any challenge is to collate a large dataset of images that match the test criteria (O'Toole & Phillips, 2015). In the second step each algorithm that was entered into the challenge computes similarity scores for the face pairs contained in the test dataset (O'Toole & Phillips, 2015; Phillips & O'Toole, 2014). An algorithm's performance is assessed by comparing its similarity scores for 'same' identity image pairs to its similarity scores for 'different' identity image pairs. A high performing algorithm consistently attributes higher similarity scores to 'same' identity face pairs, than to 'different' identity pairs. A threshold needs to be set to translate similarity scores into same-different judgements. Image pairs above the threshold are considered matches, and images below the threshold are treated as mismatches (O'Toole & Phillips, 2015; Phillips & O'Toole, 2014).

The accuracy of automated face-recognition systems is continuously improving, even under challenging conditions were images vary in illumination, facial expression, distance to the camera and resolution (O'Toole & Phillips, 2015). The best algorithms either perform comparable or outperform humans in identifying frontal images that were captured under different illumination conditions, including outdoor lighting, and that depict variations in hairstyle and clothing (O'Toole et al., 2007; O'Toole, An, Dunlop, Nato, & Phillips, 2012). However, humans outperform algorithms for difficult face pairs (for example, pairs sampled from unconstrained sets that show a wide range of pose variations) and for video (Blanton, Allen, Miller, Kalka, & Jain, 2016; for a review see Phillips & O'Toole, 2014). One contributing factor to the superiority of human performance for these stimuli is that human observers, when confronted with difficult to discriminate stimuli, effectively utilize non-face identity cues including information contained in the body (O'Toole et al., 2011; Rice, Phillips, Natu, An, & O'Toole, 2013).

One critical shortcoming current face-recognition algorithms share with humans is that they are biased. For example, a study that used an image set designed to reflect the breadth of human skin tones to assess the gender classification function of three facial analysis algorithms found that whilst the algorithms were highly accurate at identifying the gender of white males, with error rates ranging from 0% to 0.8%, they struggled to correctly categorise women with darker skin as female, with error rates ranging between 21% and 35% (Buolamwini, 2018). The accuracy of face-recognition algorithms has been shown to be biased towards the dominant ethnic group of the country in which they were developed (Phillips, Jiang, Narvekar, Ayyad, & O'Toole, 2011). It follows that the biases seen in face-recognition algorithms are likely to be an artefact of the image sets that were used to train them (O'Toole & Phillips, 2015).

Whilst the accuracy of face-recognition algorithms is benchmarked without accounting for human error, the operational performance of automated face recognition systems is determined by the system and its users (White, Dunn, et al., 2015). To illustrate, when screening passport applications Australian passport issuance officers compare each applicant's photo to a candidate list of eight potential matches that were retrieved by an algorithm from a large national database. Successful detection of fraudulent applications is therefore equally driven by the effectiveness of the automated face-recognition system in identifying potential matches in the database, and by the ability of the passport issuance officer to correctly identify matches, and successfully reject mismatches. A study that assessed human performance on this task found an error rate of over 50% for adult faces, and over 60% for child faces (White, Dunn, et al., 2015). Trained reviewers and untrained student operators were found to perform equally poorly. When reviewing applications passport issuance officers sometimes refer potential matches to a facial examiner for further evaluation. A group of these specialist facial examiners was found to perform somewhat better on the task. However, even this high performing group made errors on 30% of trails (White, Dunn, et al., 2015).

One potential explanation for the poor accuracy is that the task of matching targets to candidate lists that are drawn from very large data sets is simply very difficult (White, Dunn, et al., 2015). The population from which the candidate list is sample consists of a very large national database, which contains millions of passport photos. At least some of the photos that are returned are therefore likely to bear a very close resemblance to the target. Another potential contributing factor is that the system returns a rank ordered candidate list of potential matches. A study that has investigated how fingerprint experts process rank-ordered candidate lists of fingerprints that are returned by automated finger print identification systems found that the experts spent less time examining samples at the bottom of the list (Dror, Wertheim, Fraser-Mackenzie, & Walajtys, 2012). Critically, the study also found that more false positives identifications were made for fingerprints that were presented towards the top of the list, and more targets were missed if they were presented at the end of the list. It is possible that similar factors are at play when passport issuance officers review the candidate lists that are returned by automated face recognition systems.

2.3.5 Conclusion

Professionals who perform face-matching tasks in applied settings are subject to a variety of influences. These include contextual information and beliefs, time pressure, distractions, monotony and the need to perform multiple tasks at once. The prevalence of these variables in applied settings limits the ecological validity of face matching research. To illustrate, facial comparison is only one component of the identity verification process that takes place at passport control. Border Force agents run various data base and information checks, as well as talking to the traveller whose identity is being verified in order to establish whether they are who they say they are, and that they are eligible to enter the country. Research that explores how these factors affect face-matching is still in its infancy. One limitation of this body of work as it currently stands is that it is mainly focussed on identifying changes in face-matching accuracy, as well as sometimes exploring differences in the patterns of match and mismatch errors. Whilst this is an important first step, it is nevertheless crucial that the underlying cognitive mechanisms that drive these changes are understood.

The influences outlined above were treated as extraneous variables in Experiments 1–7. For example, whilst professionals are exposed to potentially biasing contextual information, the face-comparison in Experiments 1–7 were performed in a context-free vacuum. This limits the ecological validity of the findings reported in the empirical chapters of this thesis. However, an understanding of the effects of comparison on the perceived similarity of unfamiliar face pairs, and the role that variations and phenotypic similarity and the presence of pictorial differences play can help to elucidate the cognitive mechanisms that underlie face-matching performance. Understanding these mechanisms in turn will make it easier to understand how, when and why the extraneous influences that are prevalent in applied settings might influence the outcome of a given comparison. This argument is further explored in the final chapter of this thesis, which examines how the confirmation bias might interact with the process of structural alignment, and how the need for task switching might affect the process of habituation.

2.4 Population validity

Performance on face processing tasks is driven by various factors that are intrinsic to an individual, including—natural ability, motivation, training and experience (Noyes, Phillips, & O'Toole, 2017). Motivation was discussed in the previous section, and will not be further considered here. Natural face processing abilities range widely in the general population, from individuals with prosopagnosia who are severely impaired in their ability to recognise faces, to super-recognisers who have exceptional face-recognition abilities. The abilities of super-recognisers and the cognitive processes that underpin those abilities are examined in section 2.4.2. There is little support for the notion that professional experience in isolation improves performance on face identification tasks (Noyes et al., 2017). However, there is a growing body of evidence suggesting that the training that forensic facial examiners receive improves performance. The effectiveness of various training approaches is reviewed in section 2.4.1.

2.4.1 Training

There is evidence to suggest that expertise in performing face-comparisons can improve face matching accuracy (Norell et al., 2015; White, Dunn, et al., 2015; White, Phillips, et al., 2015; Wilkinson & Evans, 2009). For example, relative to untrained controls, forensic experts were found to be more accurate at matching image pairs, and more cautious when interpreting low quality images (Norell et al., 2015). Similarly, a study that deployed the Glasgow Face Matching Test (GFMT) to assess face matching accuracy found that forensic facial examiners outperformed professionals who are not experts in facial comparisons (White, Phillips, et al., 2015). Expertise was found to be particularly beneficial at longer exposure durations that allow for considered judgments. Although it is difficult to disentangle the effects of training from those of experience and motivation (Noyes et al., 2017), this finding suggests that the feature-by-feature comparison methods taught in professional training might be effective in improving performance (White, Phillips, et al., 2015).

A study that tested the effectiveness of a face-recognition training module on face matching accuracy, found that the training did not improve face matching performance (Woodhead, Baddeley, & Simmonds, 1979). However, the experimental task that was used to assess whether the training improved face matching accuracy may have contributed towards the null result. Participants were asked to match either four or 16 search targets (which were visible to them throughout the test), to a series of 240 sequentially presented faces (Woodhead et al., 1979). Each sequentially presented face was shown for 10 s. The authors acknowledge that the experimental task did not allow for enough time to deploy the time-consuming approach of classifying faces by their features that was taught in the face-recognition training module (Woodhead et al., 1979). A further potential contributing factor to the null result is the fact that the training was focussed on enhancing face-memory, not face matching performance. Trainees were encouraged to exaggerate a distinguishing facial feature (wide-set eyes, thin lips) in the mind's eye. Whilst there is evidence to suggest that caricaturing can aid recognition (Itz, Schweinberger, & Kaufmann, 2017; Lee, Byatt, & Rhodes, 2000; Mauro & Kubovy, 1992), its effect on unfamiliar face matching accuracy is more nuanced. A study that compared simultaneous unfamiliar face matching performance for caricatured and non-caricatured faces found that caricaturing improved performance on target absent trials, and impaired performance on target present trials (Mcintyre, Hancock, Kittler, & Langton, 2013). This finding suggests that any improvements that arise from mentally caricaturing a face are only likely to be seen on target-absent trials, and may disrupt performance on target present trials.

Alenezi and Bindemann (2013) conducted a series of six experiments to assess whether the provision of performance feedback would improve unfamiliar face matching accuracy. Summative outcome feedback at the end of a block of trials was not found to benefit performance (Alenezi & Bindemann, 2013). However, the provision of performance feedback on a trial-by-trial basis ameliorated a decline in accuracy on mismatch trials that occurred in the absence of feedback (Alenezi & Bindemann, 2013). The authors conclude that trial-by-trial feedback might help to maintain participants' engagement with the task (Alenezi & Bindemann, 2013). A related study found that trial-by-trial feedback training can improve performance beyond counteracting the detrimental effects of task repetition (White, Kemp, Jenkins, & Burton, 2014). The study found that performance on the GFMT in individuals who received trial-by-trial feedback improved across trial blocks, whereas the no-feedback control group showed no improvement. The study further found that these improvements were mainly driven by individuals who are relatively poor at face matching. Following trial-by-trial feedback training individuals in the low-aptitude group were found to perform as well as those in a high-aptitude group (White, Kemp, Jenkins, & Burton, 2014).

It has been argued that "trial-by-trial feedback can serve both to maintain accuracy and to improve performance, depending on the point in time at which this information is provided" (Alenezi & Bindemann, 2013, p. 751). According to this argument the performance improvements seen in White et al are attributable to the fact that participants were able to study their matching errors (Alenezi & Bindemann, 2013). This argument might also help to explain why pair-based working can improve performance. A study that compared individual and pair performance on the GFMT found that individuals outperformed pairs (see also Bruce et al., 2001, Experiment 3; Dowsett & Burton, 2015). For participants who were relatively poor at face matching the improvement in performance persisted when they subsequently performed the task alone. This finding suggests that pair-based training could be used to improve the unfamiliar face matching abilities of low performing individuals (Dowsett & Burton, 2015).

The Facial Identification Scientific Working Group (FISWG) provides a set of recommendations for facial comparison training programs (Facial Identification Scientific Working Group FISWG, 2012b). According to these recommendations forensic examiners should be trained on how to use facial comparison methods, tools and technologies, and on how to select a suitable approach. The FISWG recommends the use of morphological analysis as the primary method of comparison (Facial Identification Scientific Working Group FISWG, 2012a). Morphological analysis is a facial comparison method in which the similarities and differences between facial features (for example, eyes, ears, cheek area, blemishes, and face shape) are assessed (Facial Identification Scientific Working Group FISWG, 2012a). Research has shown that not all features that are considered in morphological comparisons are equally useful. Specifically, training individuals to deploy a face-shape strategy did not improve their performance on the Glasgow Face Matching Test (Alice Towler, White,

& Kemp, 2014). Furthermore, face-shape classifications were shown to have low within-rater and within-identity consistency (Alice Towler et al., 2014). Based on these findings the authors conclude that face shape should be excluded from the list of features that are examined during morphological facial comparisons (Alice Towler et al., 2014).

Evidence that one-to-one feature comparisons can improve unfamiliar face matching accuracy is provided by a study that deployed the Glasgow Face Matching Task (Towler, White, & Kemp, 2017). The study found that participants who rated the similarity of 11 features were more accurate than controls in judging that two faces were the same person (Towler et al., 2017). A follow up experiment that compared the performance of students and forensic facial examiners on the feature rating task found that the examiners were more accurate than the students (Towler et al., 2017, experiment 3). The experiment also found that whilst the student group showed an inversion effect on mismatch trials, the forensic facial examiners did not. Similarly, another study that compared the performance of facial examiners with that of untrained students also found a stronger face-inversion effect in the student group (White, Phillips, et al., 2015). The dissociation between the face-inversion effect and face matching performance suggests that forensic facial examiners superior face-matching performance is underpinned by feature-based processing. This is a marked difference to super-recognisers, who show enhanced holistic processing (Bobak, Bennetts, Parris, Jansari, & Bate, 2016; Russell, Duchaine, & Nakayama, 2009). Super-recognisers are discussed in the next section.

2.4.2 Individual differences

There are large individual differences in face matching and face-recognition abilities (Megreya & Burton, 2006; Turano, Marzi, & Viggiano, 2016; Wilmer, 2017). Whilst training can have a positive effect on identification accuracy, these individual differences are likely to play a greater role. Evidence in support of this claim is provided by a study that tested whether experienced passport officers—most of whom had completed a training module on identify verification from photographs—would outperform untrained student controls on a face matching task (White, Kemp, Jenkins, Matheson, et al., 2014). The study found no difference in face matching accuracy between the passport officers and the controls. However, there were large individual differences in performance in both groups, with some individuals performing at 100%

accuracy. For the passport officer group, individual differences in face matching accuracy were unrelated to length of service. Together these finding suggest that neither training nor experience on the task had a significant impact on face matching performance.

The term super-recogniser was coined to refer to individuals who perform two standard deviations above the mean on standardised face-identification tasks (Edmond & Wortley, 2016; Russell et al., 2009), but several different inclusion criteria have been used (for a review see Noyes et al., 2017). It has been argued that in the absence of any evidence to the contrary, super-recognisers are “best understood as the top performers sampled from a distribution of normal facial-recognition skills, rather than as a distinct population of people with ‘superior recognition capacity” (Noyes et al., 2017, p. 1/29). A study that assessed the face-recognition and-face matching abilities of four individuals who claimed to be exceptional at recognising faces found that they outperformed controls on both tasks (Russell et al., 2009). This pattern of results was replicated in a follow-up study using different face-recognition and face matching tests (Russell, Chatterjee, & Nakayama, 2012). Super-recognisers have also been shown to outperform controls on face-processing tasks that resemble applied scenarios, including matching a target face to faces presented in a line-up, and recognising faces in poor quality video footage (Bobak, Hancock, et al., 2016).

Whilst super-recognisers have been shown to excel on a variety of face identification tasks, the evidence that they possess exceptional face matching skills is less compelling than the evidence for their superior face-recognition abilities. For example, whilst the super-recognisers in the study conducted by Russell and his colleagues (2009) outperformed controls on both face-recognition and face matching tasks, the effect size for the face matching task was smaller than for face recognition. Furthermore, visual inspection of individual scores suggests that only a subset of the super-recognisers tested by Russell and his colleagues (2009) performed better than the control mean on the face matching task (Bobak, Bennetts, et al., 2016).

The argument that not all super-recognisers have superior face matching abilities is supported by the findings of a study that performed single-case comparisons (Bobak, Bennetts, et al., 2016). The study found that only half of the super-recognisers outperformed controls on the face matching tasks, for the other half superior performance was restricted to face recognition (Bobak, Bennetts, et al., 2016). Similarly, a study that tested the performance of seven super-recognisers on the Glasgow Face Matching Test (GFMT) found that whilst the super-recognisers outperformed controls on a group level, only some case comparisons reached significance (Bobak, Dowsett, & Bate, 2016). To assess whether the absence of performance differences might be attributable to ceiling effects the study also administered the Models Face Matching Test (MFMT), which was designed to be more difficult than the GFMT. All but one super-recogniser outperformed control participants on the more difficult MFMT (Bobak, Dowsett, et al., 2016). Taken together these findings suggest that whilst the absence of performance differences might be partially attributable to limitations in the sensitivity of the face matching tasks (see also Bobak, Pampoulov, & Bate, 2016), not all individuals with superior face-recognition abilities necessarily also excel at face matching (Bobak, Bennetts, et al., 2016; see also Noyes et al., 2017).

It has been argued that super-recognisers' superior face identification abilities are attributable to proficiencies in holistic processing (Bobak, Bennetts, et al., 2016). As discussed in section 2.2, the inversion effect is one the main measures of holistic processing. A larger inversion effect—i.e. a greater difference in the efficiency with which upright and inverted objects are processed—is thought to indicate stronger holistic processing (Bobak, Bennetts, et al., 2016). When super-recognisers' holistic processing abilities were assessed using upright and inverted objects, the super-recognisers did not show a greater inversion effect than controls (Bobak, Bennetts, et al., 2016). However, this finding is somewhat unsurprising considering that super-recognisers superior memory and matching abilities appear to be domain specific (Bobak, Bennetts, et al., 2016; see also Davis, Lander, Evans, & Jansari, 2016). Super-recognisers have been found to show greater face-inversion effects than controls (Bobak, Bennetts, et al., 2016; Russell et al., 2009). For example, in a study conducted by Bobak and his colleagues (2016) five out of six super-recognisers showed evidence of an enhanced face-inversion effect. The authors acknowledge that this finding is compatible with the notion that super-recognisers may simply be exceptional at abstracting featural information from faces (Bobak, Bennetts, et al., 2016), which can also be affected by inversion (for a review see Mckone & Yovel, 2009). Nevertheless, they conclude that "heightened holistic processing may represent a common underpinning mechanism across even heterogeneous cases of super recognition" (Bobak, Bennetts, et al., 2016, p. 59; but see Noyes et al., 2017 for a discussion of limitations).

A study that tested the face matching abilities of four members of the Metropolitan Police super-recogniser team (which was set up in May 2015) found that they performed well above average on face matching tasks involving both high-quality and degraded images (Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016; see also Davis et al., 2016). This finding suggest that the creation of specialist police identification units, staffed with individuals who are known to have superior face-identification abilities is a fruitful endeavour. It is important to deploy objective tests of face identification abilities when recruiting individuals to these specialist units. This raises the issue of choosing an acceptance criterion. Should members of these units perform two standard-deviations above the population mean? Should it be more, just to be certain? Alternatively, is a performance level of one standard deviation above the mean good enough—especially when the results of multiple recognisers are pooled? A further caveat that impacts on the use of police-recognisers is that super-recognisers do make errors (Edmond & Wortley, 2016). For example, when super-recognisers were required to remember unfamiliar faces for a longer period prior to the identification task, their error rate was 33 per cent (Bobak, Hancock, et al., 2016). Difficulties can arise if identifications made by super-recognisers are treated as fact. A final risk with deploying police super-recognisers is that just like everybody else, they are not immune to biases.

2.4.3 Conclusion

This section has examined two populations with superior face processing skills—super-recognisers, and forensic facial examiners. The superior abilities of super-recognisers appear to be innate, whereas those of facial examiners are acquired through training. Furthermore, whilst the exceptional abilities of super-recognisers have been associated with enhanced holistic processing skills, those of forensic facial examiners are not. Supporting evidence is provided by the finding that forensic facial examiners show a reduced inversion effect (Towler et al., 2017; White, Phillips, et al., 2015). Instead, forensic facial examiners' superior performance appears to be attributable to the effective application of feature-by-feature comparisons.

The participants who took part in Experiments 1–7 were students, and members of the public who attended university open days. They are likely to represent a range of natural face-processing abilities. Except for trained forensic examiners, professionals have been found to perform comparable to student controls on a variety of face matching tasks. It follows that some generalisations can be made from the participant sample to professionals. However, the findings may not generalise to forensic examiners who have received training on how to perform feature-by-feature comparisons. Supporting evidence is provided by the finding that whilst performance of a feature-by-feature comparison task improved the face-matching accuracy of students, the improvement was not associated with a reduced inversion effect (Towler et al., 2017). Similarly, the findings of Experiments 1–7 may also not generalise to super-recognisers. The implications in relation to super-recognisers are further discussed in the final chapter of this thesis.

Chapter 3: The effect of comparison on the perceived similarity of unfamiliar face stimuli that are relatively easy to tell apart

Face comparison is a common task in several applied settings, including border control and criminal investigations. The face stimuli that are encountered in these contexts vary widely in their degree of resemblance, both because of phenotypic variation in human faces, and due to pictorial differences. To inform our understanding of human performance, it is important to understand how these variations in the similarity of face stimuli interact with the comparison task to affect perceived similarity. This chapter examines the effect of comparison on the perceived similarity of face stimuli that are relatively easy to discriminate, both because they are of good image quality and relatively similar pictorially, and because there are clear phenotypic differences between the faces. On the face of it, understanding how comparison affects the perceived similarity of faces that are easy to tell apart may not seem of relevance in applied settings. However, unfamiliar face matching performance has been shown to be error prone, even when high-quality video stills and photographs are used (Bruce et al., 1999; Henderson et al., 2001).

Performance on face matching tasks has been shown to be affected by factors that are likely to impact on accuracy in applied contexts. For example, performance on a same–different face matching task has been shown to be significantly less accurate for two photos taken a few of months apart, compared to two photos taken on the same day (Megreya et al., 2013; see also Bindemann & Sandford, 2011). The authors concluded that experimental studies might underestimate task difficulty in applied settings, such as border control, because passports are generally valid for 10 years (Megreya et al., 2013). Furthermore, there is a body of research that suggests that the discriminability of a given face pair can vary depending on the task context (Alenezi & Bindemann, 2013; Alenezi et al., 2015; Attwood et al., 2013; Beattie et al., 2016). It follows that two faces that seem relatively easy to discriminate in one context, may be difficult to tell apart in another.

Task repetition is one contextual factor that has been shown to affect observers' ability to discriminate face stimuli that they would be able to tell apart in other contexts (Alenezi & Bindemann, 2013; Alenezi et al., 2015). An experiment in which participants were asked to perform 250 face matching trials found that task repetition resulted in a decline in the participants' ability to correctly state that two faces are different (Alenezi & Bindemann, 2013). The decrease in performance on different trials was replicated for a range of face stimuli in a series of follow-up experiments (Alenezi & Bindemann, 2013; Alenezi et al., 2015). In one of these experiments participants were asked to perform 1000 face matching trials. In the first block of trials the participants identified mismatches with an accuracy of 80%. On the final test block their performance accuracy on mismatch trials had dropped to 50%. The authors concluded that with continuous exposure to a matching task, different faces increasingly look the same (Alenezi & Bindemann, 2013).

Understanding the effect of comparison on the perceived similarity of unfamiliar faces that are relatively easy to discriminate is also of theoretical interest. Whereas structural alignment theory predicts that comparison should increase the perceived similarity of unfamiliar faces (Boroditsky, 2007; Markman & Gentner, 1993b, 1996), the habituation account of comparison suggests the opposite (Honey & Bateson, 1996; Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009). There is empirical evidence to support both claims (Boroditsky, 2007; Hassin, 2001; Mundy et al., 2014, 2007). Yet, the degree to which either set of findings can be extrapolated to the range of naturally encountered face stimuli is limited by the stimuli that were used. Whilst comparison-related decreases in perceived similarity have been demonstrated with face stimuli, the stimuli used in these studies were pictorially homogenous, and could only be discriminated by a few, subtle phenotypic differences (Mundy et al., 2014, 2007). The stimuli for which comparison-related increases in perceived similarity have been observed, are line drawings of animals, novel 3D objects, and elements of the Ebbinghaus illusion (Boroditsky, 2007; Hassin, 2001). The predictions of structural alignment theory are yet to be directly tested with faces.

Understanding whether, and under what conditions, comparison of unfamiliar faces will alter their perceived similarity, will help to define the boundaries of both structural alignment theory, and the habituation account of comparison. This theoretical knowledge, in turn, can be used to inform best practices in applied settings. Knowing that an issue exists, and understanding the potential impact it has on performance, whilst being a crucial first step, does not provide sufficient information to rectify it. To improve face matching accuracy, it is therefore important to understand how the interplay between stimulus attributes, comparative processes, and situational variables affects performance. Theoretical accounts of comparison further us towards this goal by defining the cognitive mechanisms that underpin performance.

The current chapter consists of three experiments that investigated whether comparing two unfamiliar faces before rating their similarity affects how similar the compared faces are perceived to be. Two comparison manipulations were deployed: a similarity-listing task, and a difference-listing task. These two manipulations were chosen because there is some evidence to suggest that similarity- and difference-focused comparisons can have different effects on perceived similarity (Hassin, 2001). Experiment 1 tested whether, and how, comparison affects the perceived similarity of phenotypically similar and dissimilar faces that are relatively easy to tell apart. Experiment 2 is a replication of the first experiment with a different stimulus set. Furthermore, Experiment 2 also tested an alternative explanation for comparison-related changes in perceived similarity, namely, that they are attributable to greater familiarity with the compared faces. Experiment 3 assessed whether the effects of comparison on perceived similarity extend to face pairs that are difficult to compare, because comparing them requires mental rotation in the picture plane.

The face stimuli used in Experiments 1–3 were presented from different angles. In the first two experiments one face in a pair was shown from a frontal view and the other was shown from a ¾ view. The motivation for introducing this manipulation was twofold. First, the face stimuli that are compared in applied settings can vary widely in pictorial similarity, due to differences in pose, lighting, facial expression, and a variety of other factors. The faces were presented from different angles to reflect this variance and thereby improve the ecological validity of the stimuli. Second, the faces were presented from different angles to test the boundary conditions of the habituation account of comparison. Comparison-related decreases in perceived similarity have been observed for face stimuli that were pictorially identical (Mundy et al., 2014, 2007). The compared faces were shown from the same angle, in the same lighting, and with the same background. The only differences between the faces were subtle variations in their facial features. Presenting faces from different angles assesses whether habituation still plays a role for face pairs that are not pictorially homogenous.

3.1 Experiment 1: The Effect of comparison on the perceived similarity of similar and dissimilar face pairs

The first aim of this experiment was to understand the effect of comparison on the perceived similarity of unfamiliar faces that are easy to discriminate. To assess the effect of comparison on perceived similarity, participants were asked to compare black-and-white images of male faces shown from different angles. It was assumed that if comparison affects the perceptual similarity of faces that are easy to tell apart, then face-pair similarity ratings made after performing a comparison would differ significantly from similarity ratings that were not preceded by a comparison.

The second aim of this experiment was to explore whether any effect of comparison on the perceived similarity of faces that are easy to tell apart, is mediated by the phenotypic similarity of the compared faces. To assess the role of stimulus similarity, half of the face pairs were matched to be relatively similar, and half to be relatively dissimilar. It was assumed that if phenotypic similarity mediates the effect of comparison on perceived similarity, then there would be a significant interaction between stimulus similarity, and comparison task.

The third aim of this experiment was to assess the role of the comparison task in mediating the effect of comparison on the perceived similarity of unfamiliar faces. There is some evidence to suggest that whereas similarity-focussed comparisons increase the perceived similarity of the compared stimuli, difference-focussed comparisons result in a decrease in perceived similarity (Hassin, 2001). It was assumed that if similarity- and difference-focused comparisons have opposing effects on perceived similarity, then similarity ratings following difference-focused comparisons would be lower than the similarity ratings for the no-comparison control, and similarity ratings following similarity-focused comparisons would be higher.

3.1.1 Method

Participants

One hundred and sixty-two University of Southampton undergraduate students (131 females, 31 males), ranging in age from 18 to 42 years (*Mdn* = 19 years) volunteered to take part. They were randomly allocated to the two between-participant treatment conditions: 86 students participated in the similar faces condition; the remaining 76 participated in the dissimilar faces condition.

Materials

The stimuli consisted of pairs of black-and-white images of male faces obtained from the Psychological Image Collection at Stirling (http://pics.psych.stir.ac.uk). In a pilot study 51 volunteers were asked to judge the similarity of three target faces to 12 comparison faces on a scale of 1 (*not similar*) to 7 (*very similar*). Each of the 36 face pairs consisted of one frontal view, and one three-quarter view. Three face pairs with high similarity ratings (M = 4.2, SD = 1.7) were used in the similar faces condition; and three pairs of faces with low similarity ratings (M = 1.6 SD = 0.8) were used in the dissimilar faces condition. Examples of these face pairs are shown in Figure 10.

Procedure

The participants were asked to fill in a short, five-page booklet at the beginning of a lecture. The task took them approximately five minutes to complete. Depending on the between-participant treatment condition the questionnaire contained either three similar or three dissimilar face pairs. Each participant completed three within-participant trials: a control, a similarity comparison, and a difference comparison. In the control condition participants were asked to indicate how similar the two faces are on a scale of 1 (*not similar*) to 7 (*very similar*). In the similarity-comparison condition this similarity rating was preceded by the question “Please describe 3 similarities between the two faces”. In the difference-comparison conditions this similarity rating was preceded by the question “Please describe 3 similarities differences between the two faces”. The allocation of face pairs to the treatment conditions and the order of presentation were randomized across participants.

Figure 10: *An example of the dissimilar (left) and similar (right) face pairs used in Experiment 1.*

3.1.2 Results

The effect of comparison on the perceived similarity of similar and dissimilar face pairs is shown in Figure 11. A 3 (comparison: no comparison, similarity comparison, difference comparison) x 2 (stimulus: similar faces, dissimilar faces) mixed ANOVA revealed a significant main effect of the between-participant stimulus variable, $F(1, 160) = 122.16$, $p < .001$, $\eta_p^2 = .433$. This manipulation check confirms that participants perceived the similar face pairs ($M = 3.8$, $SD = 1.3$) to be more similar in appearance than the dissimilar face pairs ($M = 2.3$, $SD = 1$).

The main effect of the within-participant comparison variable was also significant, $F(2, 320) = 4.92$, $p < .01$, $\eta_p^2 = .03$. The interaction between stimulus type and comparison was not significant, $F(2, 320) = 1.22$, $p = .298$, $\eta_p^2 = .008$. Due to the absence of an interaction between stimulus type and comparison, planned contrasts were conducted on the combined data for similar and dissimilar face pairs (see Figure 11). A planned contrast revealed that faces whose similarities were assessed were subsequently perceived to be more similar to each other than face pairs in the no-comparison control condition, $F(1, 160) = 7.29$, $p < .01$, $\eta_p^2 = .044$. The same pattern of results was observed for faces whose differences were assessed prior to similarity judgments, $F(1, 160) = 8.38$, $p < .01$, $\eta_p^2 = .05$.

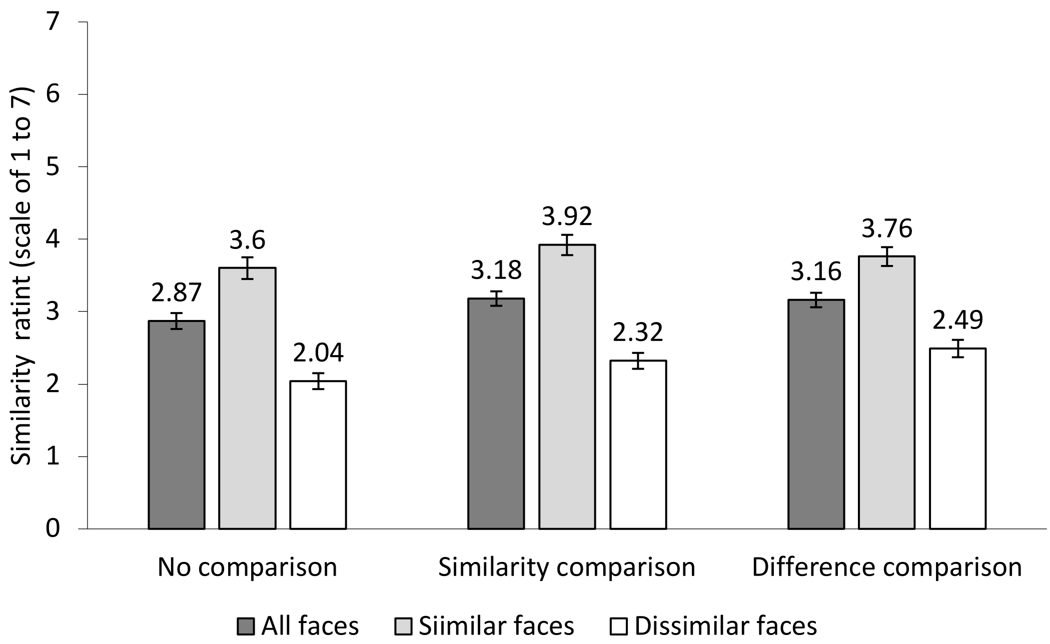


Figure 11: *The mean similarity ratings (and associated standard errors) for the different comparison conditions in Experiment 1.*

3.1.3 Discussion

A comparison-related increase in perceived similarity was found for both similar and dissimilar faces. This finding shows that prior comparison of face pairs that are easy to tell apart increases the perceived similarity of the compared faces. This result is consistent with structural alignment theory, which predicts that comparison should increase the perceived similarity of both similar and dissimilar face pairs, because the similarities that are revealed during alignment, are weighted more highly in perceived similarity than the differences (Boroditsky, 2007; Markman & Gentner, 1993b). Similarity- and difference-focused comparisons both resulted in a significant increase in perceived similarity, relative to the no-comparison control. This finding suggests that the comparison-related increase in perceived similarity was not dependent on the nature of the comparison task. This finding is also compatible with structural alignment theory, because the same structural alignment process underlies both tasks. It follows that both similarity- and difference-focused comparisons should result in the discovery of commonalities. The theoretical implications of these findings are further examined in the discussion section at the end of this chapter.

An alternative explanation for the observed effects of comparison on similarity perception is that they might be attributable to participants' greater familiarity with the compared faces. In other words, similar stimuli appeared more similar after the comparison task, not because they were compared, but because they were perceived together. Associative processes are one mechanism by which greater familiarity with stimuli could result in an increase in perceived similarity. Evidence from both animal (e.g. Honey & Hall, 1989, 1991; Kaiser, Sherburne, Steirn, & Zentall, 1997; Urcuioli & Lionello-DeNolf, 2005; Urcuioli, Zentall, Jackson-Smith, & Steirn, 1989; Ward-Robinson & Hall, 1999) and human (e.g. Delamater & Joseph, 2000; Ellis, Feuge, Long, & Pegram, 1964; Hall, Mitchell, Graham, & Lavis, 2003; Rossman & Goss, 1951) perceptual learning studies suggests that associations formed during pre-training affect the ease with which differential responses are acquired. The viability of stimulus familiarity as an alternative explanation was tested in Experiment 2. One possible limitation to the external validity of the current findings is the gender ratio of the participants. The issue of gender differences in face perception, and their potential impact on the findings of the experiments reported in this thesis are addressed in the chapter discussion.

3.2 Experiment 2: Is the effect of comparison on the perceived similarity of face pairs attributable to familiarity?

The first aim of this experiment was to test the reliability of the finding that comparison increases the perceived similarity of faces that are relatively easy tell apart with a different stimulus set. The stimuli used in this experiment were pairs of male faces shown from different angles. Unlike in the first experiment, the face pairs were not matched based on similarity ratings. Instead pairs of morphed faces were used as stimuli to reduce the overall phenotypic heterogeneity of the stimulus set. The use of morphed faces reduces the ecological validity of this study. This trade-off was deemed necessary because each participant only performed one trial per experimental condition. Consequently, any variations in the similarity of the stimulus pairs could significantly impact on the findings.

The second aim of this experiment was to test an alternative explanation for the increases in perceived similarity in the comparison conditions, namely that they are attributable to participants' greater familiarity with the stimuli. A face description task was used to test this alternative hypothesis. Previous research has shown that describing an unfamiliar face from memory can improve recognition memory for that face (Brown & Lloyd-Jones, 2006, 2005). The authors argue that verbal descriptions allow for "the formation of richer semantic associations with the described face, which benefits retrieval" (Brown & Lloyd-Jones, 2006, p. 283). In the context of the current experiment, the existence of richer memory representations for described faces may act as an additional commonality between them, thereby increasing their perceived similarity relative to face pairs that are encountered for the first time. Additionally, associations formed between two stimuli during their shared exposure history may also act as commonality. It was predicted that if greater familiarity with the compared face pairs is the cause of the increase in perceived similarity observed in Experiment 1, then describing faces prior to rating their similarity would result in an increase in perceived similarity.

3.2.1 Method

Participants

Seventy-three Goldsmiths, University of London undergraduate psychology students (53 females, 20 males), ranging in age from 18 to 33 years (*Mdn* = 19 years), participated in exchange for course credit.

Materials

Face morphs were used to reduce the heterogeneity between stimulus pairs. Eight morphed faces were generated using the Morpheus Photo Morpher software package (version 3.16). Each face consisted of a 50% morph between two individual male faces taken from the, Computer Vision Laboratory, University of Ljubljana, Face Database (http://www.lrv.fri.uni-lj.si/facedb.html). The two male faces that comprised each morph were matched on face shape and pose to facilitate the morphing process.

The morphed faces were grouped into four pairs. The first face-morph in each pair was shown in a portrait view, and the second was shown in a ¾ view (see Figure 12). To control for inter-item similarity one face in each pair had a relatively square or long shape and the other had a relatively oval shape. Furthermore, one face in each pair always had darker eyes.

Procedure

The participants were asked to fill in a short seven-page booklet during a research methods class. Each participant completed four within-participant trials: a control, a similarity comparison, a difference comparison, and a face description. In the control condition participants were asked to indicate how similar the two faces are on a scale of 1 (*not similar*) to 7 (*very similar*). In the similarity-comparison and difference-comparison conditions this similarity rating was preceded by the question "Please describe 3 similarities (differences) between the two faces". In the face-description condition, participants were asked to give a brief description of each face before rating the similarity of the two faces. The allocation of face pairs to the treatment conditions and the order of presentation were counter-balanced across participants.

Figure 12: *An example of the face pairs used as stimuli in Experiment 2. Each face in a pair consisted of a morph composed of two individual faces.*

3.2.2 Results

Mean similarity ratings are shown in Figure 13. A repeated measure ANOVA showed a significant main effect of face-pair pre-processing, $F(3,216) = 3.43$, $p < .05$, $\eta_p^2 = .045$. Planned contrasts revealed that similarity comparisons resulted in significantly higher similarity ratings than the no-comparison control, $F(1,72) = 5.65$, $p < .05$, $\eta_p^2 = .073$. Difference comparisons also resulted in an increase in perceived similarity compared to the control, $F(1,72) = 5.05$, $p < .05$, $\eta_p^2 = .066$. Similarity ratings in the describe condition did not differ significantly from the control, $F(1,72) = .03$, $p = .862$, $\eta_p^2 = .0004$.

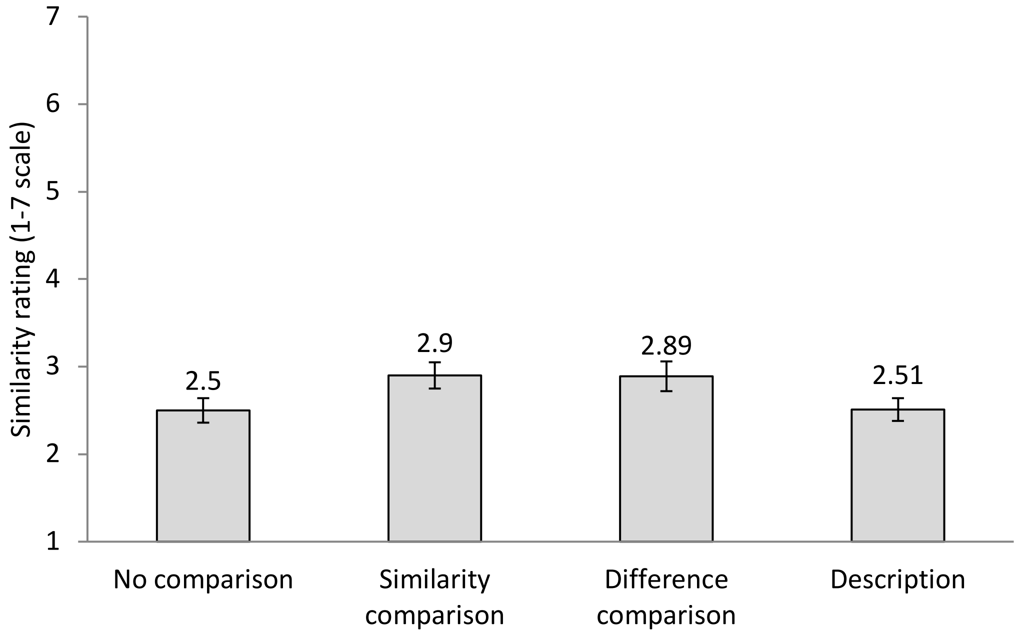


Figure 13: *The mean similarity ratings (and associated standard errors) for the different face-pair pre-processing conditions in Experiment 2.*

3.2.3 Discussion

A significant increase in perceived similarity of face pairs was found in both the similarity- and difference-focused comparison conditions, relative to the no-comparison control. This finding constitutes a replication of the pattern of results obtained in Experiment 1 with a different stimulus set. Describing the faces in a pair did not result in higher similarity ratings relative to the control. This finding suggests that stimulus familiarity is unlikely to be the cause of the higher similarity ratings in the comparison conditions. The reliability of this conclusion was tested in Experiment 4 with a different familiarity manipulation.

Together the findings from the first two experiments show that comparing two unfamiliar face stimuli that are relatively easy to discriminate increases the perceived similarity between them. In both experiments, participants were asked to rate the similarity of faces shown from different angles (frontal and ¾ view). These pictorial differences were introduced to increase the ecological validity of the experiments. The face stimuli that are compared in applied settings, such as during criminal investigations, can be very pictorially dissimilar. For example, CCTV footage often shows faces from unusual angles, because CCTV cameras are generally in high locations. These pictorial differences pose unique challenges to the comparison task. The face pairs used in the current and previous experiment, on the other hand, were relatively easy to compare, because they were captured from the same camera height, and because the angular difference between the two poses (frontal and ¾ view) is relatively small. To test whether the effect of comparison on perceived similarity is limited to face pairs that are relatively easy to compare, participants in Experiment 3 were asked to compare upright and inverted faces.

3.3 Experiment 3: The effect of comparison on the perceived similarity of upside-down faces

The aim of this experiment was to test whether the comparison-related increases in perceived similarity observed in Experiments 1 and 2 extend to face pairs that are difficult to compare. There is a large body of research which demonstrates that inversion has a detrimental effect on face recognition, and matching performance (Bruyer, Galvez, & Prairial, 1993; Collishaw & Hole, 2000; Freire et al., 2000; Goffaux & Rossion, 2007; Leder & Carbon, 2006; Valentine & Bruce, 1986, 1988; Yin, 1969). However, because the current experiment was administered as a paper booklet, it would have been possible for participants to simply turn the booklet around. Face inversion does not only impact performance when both faces are inverted. 'Mental rotation' studies using photos of faces have found a linear decline in matching performance with increasing angular distance (Bruyer et al., 1993; Collishaw & Hole, 2002; Valentine & Bruce, 1988). To render the comparison task more difficult, participants were therefore asked to compare upright to inverted faces.

3.3.1 Method

Participants

Seventy-five Goldsmiths, University of London psychology students (54 females, 21 males), ranging in age from 18 to 47 years (median age = 19 years), participated in exchange for course credit.

Materials

Eight morphed faces were generated using the Morpheus Photo Morpher software package (version 3.16). Each face consisted of a 50% morph between two individual male faces taken from the Computer Vision Laboratory, University of Ljubljana, Face Database (http://www.lrv.fri.uni-lj.si/facedb.html). The two male faces that comprised each morph were the same faces that were used to generate morphs in experiment two. The morphed faces were grouped into three pairs. The first face-morph in each pair was shown in an upright position and the second was shown upside down (see Figure 14). Both the upright and the inverted face morph in a pair were shown from a frontal view.

Procedure

The participants were asked to fill in a short five-page booklet during a research methods class. Each participant completed three within-participant trials: a control, a similarity comparison, and a difference comparison. In the control condition participants were asked to indicate how similar the two faces are on a scale of 1 (*not similar*) to 7 (*very similar*). In the similarity-comparison, and difference-comparison conditions this similarity rating was preceded by the question “Please describe 3 similarities (differences) between the two faces”. The allocation of face pairs to the treatment conditions, and the order of presentation were counterbalanced across participants.

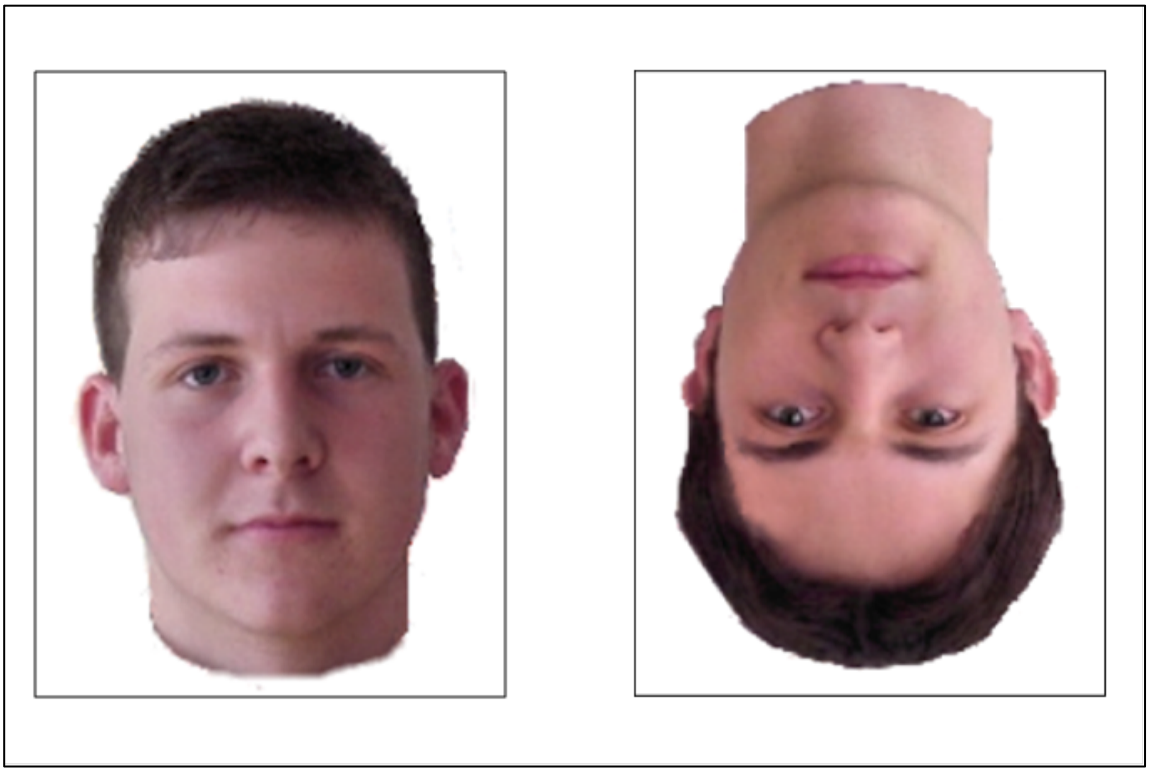


Figure 14: *An example of the face pairs used as stimuli in Experiment 3. Each face in a pair consisted of a morph composed of two individual faces.*

3.3.2 Results

Mean similarity ratings are shown in Figure 15. A repeated measure ANOVA showed a significant main effect of face-pair pre-processing, $F(2,148) = 6.15$, $p < .01$, $\eta_p^2 = .077$. Planned contrasts revealed that similarity comparisons resulted in significantly higher similarity ratings than no-comparison control, $F(1,74) = 14.48$, $p < .001$, $\eta_p^2 = .164$. Similarity ratings following difference comparisons did not differ significantly from the control, $F(1,74) = 2.67$, $p = .107$, $\eta_p^2 = .035$.

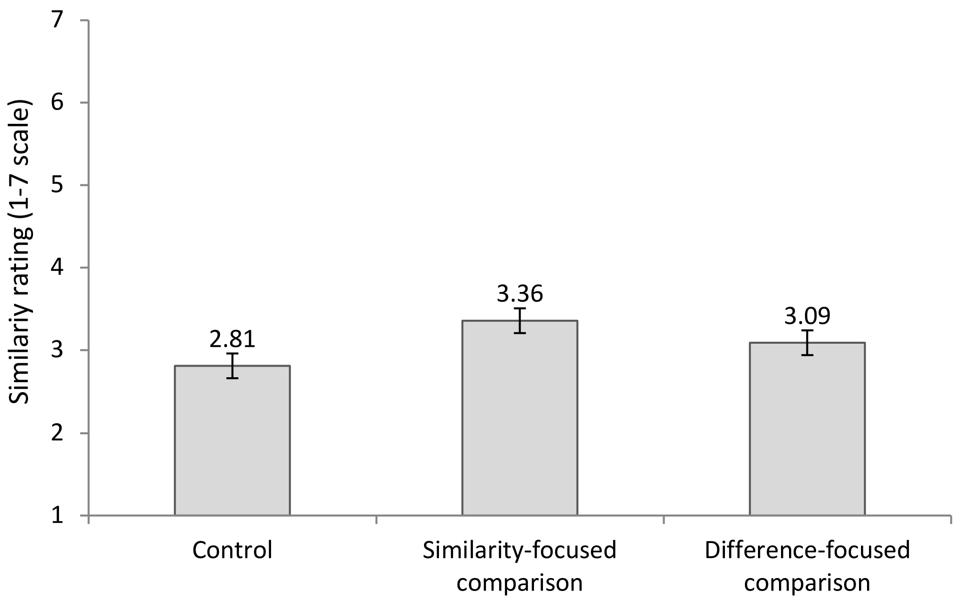


Figure 15: *The mean similarity ratings (and associated standard errors) for the different face-pair pre-processing conditions in Experiment 3.*

3.3.3 Discussion

To test whether the effect of comparison on perceived similarity is limited to face pairs that are easy to compare, participants in Experiment 3 were asked to compare upright and inverted faces. Similarity-focused comparison of upright and inverted faces was found to result in an increase in the perceived similarity of the compared faces. This finding suggests that comparison exerts an influence on perceived similarity under conditions where the comparison task is difficult to perform. Visual inspection of Figure 15 suggests that difference-focused comparison also resulted in an increase in perceived similarity compared to the no-comparison control. However, this difference was not statistically significant. One possible explanation for the lack of a significant finding in the difference-focused comparison condition is that the effect of comparison on perceived similarity is less pronounced when the comparison task is difficult to perform. Nevertheless, the finding of comparison-related increases in perceived similarity of inverted faces suggests that the findings reported in Experiments 1 and 2 have applicability to a range of forensics contexts, including those in which pictorial attributes of the stimuli make the comparison task difficult to perform. The reliability of this assumption was assessed in Experiment 6 using a different stimulus set. The stimuli used in Experiment 6 consisted of the face-pairs with added visual noise. This manipulation was chosen to simulate the poor image quality of some CCTV footage.

3.4 Chapter 3 discussion

This chapter examined the effect of comparison on the perceived similarity of unfamiliar faces that are relatively easy to tell apart. Experiment 1 found that comparing two similar or dissimilar faces shown from different angles increases the perceived similarity of the compared faces, relative to a control that was not subjected to a prior comparison. The comparison-related increase in perceived similarity was seen following both similarity- and difference-focussed comparisons. Experiment 2 replicated the finding that comparison increases the perceived similarity of unfamiliar faces that are relatively easy to tell apart, with a different stimulus set. Experiment 2 also showed that increased familiarity with the faces in a pair is not sufficient to increase their perceived similarity. Experiment 3 demonstrated that comparison-related increases in perceived similarity are not limited to stimuli that are easy to compare. These results have implications both for theories of comparisons, and in applied contexts.

3.4.1 Theoretical considerations

Experiments 1–3 have shown that comparing two unfamiliar faces that are relatively easy to tell apart increases their perceived similarity. Except for experiment 3—which did not find a significant increase in perceived similarity for the difference-focussed condition—perceived similarity increased following both similarity- and difference-focussed comparisons. This pattern of results conforms to the predictions of structural alignment theory. According to structural alignment theory comparison is a process in which two representations are brought into alignment (Markman & Gentner, 1993a, 1993b). During the structural alignment process similarities are established first, and related differences are noticed second (Gentner & Gunn, 2001). Thus, regardless of whether a comparison involves the identification of similarities or differences, the alignment process necessitates the perception of commonalities, which are weighted more highly in perceived similarity than differences (Markman & Gentner, 1993b). The findings of Experiments 1–3 constitute the first demonstration that the predictions of structural alignment theory extend to unfamiliar face stimuli.

The habituation account predicts that comparison will decrease the perceived similarity of unfamiliar faces, because viewers habituate to the commonalities between stimuli, which focusses their attention on the differences. However, the habituation process is less likely to occur for stimuli that share few common elements (Mundy et al., 2007, p. 136). The current findings suggest that the pictorial differences that are introduced when faces are shown from different angles in the 3D (Experiments 1 and 2) or 2D plane (Experiment 3), are sufficient to constitute a reduction in common elements. Alternatively, the absence of comparison-related decreases in perceived similarity may be attributable to the phenotypic differences between the compared faces. The phenotypic resemblance of the face stimuli used in studies on discrimination learning were designed to be “akin to that of identical twins” (Dwyer et al., 2009, p. 337). To assess the role of phenotypic similarity participants in Experiment 1 were asked to compare similar and dissimilar face pairs. A comparison-related increase in perceived similarity was found for both the similar and the dissimilar face pairs. However, both face pairs used in Experiment 1 were relatively easy to tell apart. It is therefore possible that comparison-related decreases in perceived similarity only occur for faces that that are difficult to distinguish phenotypically. These alternative interpretations were further tested in Experiment 4, in which participants were asked to compare similar and dissimilar face morphs shown at the same angle.

A final possibility is that comparison-related decreases in perceived similarity only occur for stimuli that are difficult to discriminate. Thus, rather than habituation only occurring when most of the pictorial information in the compared images is identical, it is more important that much of the pictorial information in each image is irrelevant to the task at hand. It then follows that habituation effects may also occur when there is a lot of visual noise or superfluous information in an image. This alternative was tested in Experiments 6 and 7.

3.4.2 Applied considerations

With more than 100 million passengers arriving in the UK in 2013, and 7,600 full-time equivalent border force staff to check their passports and travel documents (The Comptroller and Auditor General, 2013), matching travellers to their passport photos is a common task. Previous research has shown that performing face comparisons repeatedly makes faces seem more alike over time (Alenezi & Bindemann, 2013; see also Alenezi et al., 2015). The findings of Experiments 1–3 show that comparing two faces (e.g. in the similarity-listing task) makes those faces appear more similar when they are subjected to a subsequent comparison (i.e. when performing the similarity rating task). It follows that if a given person-photo pairing is repeatedly compared, they will appear more similar. One interesting implication is that person-photo pairings that look dissimilar, and are therefore re-examined, are the ones most likely to be subject to a comparison-related increase in similarity.

One aspect of the stimuli used in Experiments 1–3, which limits the extent to which these findings can be generalised to the applied context of border control, is that the faces were shown from different angles. In the context of border control the compared faces are both seen from the same frontal view. It is possible that comparison will not result in an increase in the perceived similarity of unfamiliar face pairs that are shown from the same angle. This issue was addressed in Experiment 4. Another issue which may limit the extent to which these findings can be extrapolated to an applied context is that they were obtained in a laboratory study.
However, there are several reasons to believe that the effects reported here might be amplified in the applied context of border control. First the stimuli used in Experiments 1–3 were derived from image repositories for research purposes. The photos were taken over a short time-period under similar lighting conditions. It has been shown that face matching accuracy decreases for photographs that have been captured a few months apart (Megreya et al., 2013). The photos that are compared at border control are unlikely to be that recent. Second, the participants in experiments 1–3 only performed a handful of face comparisons. They are therefore unlikely to be subject to the effects of repeated comparison, which have been shown to decrease the likelihood that mismatches are correctly identified (Alenezi & Bindemann, 2013; Alenezi et al., 2015). This general process might interact with the face-pair specific increases in perceived similarity reported here. Another important difference is that whereas task completion in Experiments 1–3 was self-paced and unhurried, Border Force agents are under pressure to meet targets (Toynbee, 2016). The effect of limiting the time available to make a comparison on perceived similarity was explored in Experiment 5.

A final important difference is that border agents are experienced and trained in matching photos to faces. There is some evidence that forensic facial identification examiners are better at matching faces than members of the public (Norell et al., 2015; White, Dunn, Schmid, & Kemp, 2015; White, Phillips, Hahn, Hill, & O'Toole, 2015). However, studies that compared face matching performance between members of the public, and either police, or passport officers (Burton, Wilson, Cowan, & Bruce, 1999; White, Dunn, et al., 2015; White, Kemp, Jenkins, Matheson, et al., 2014), found no expertise-related advantage. Furthermore, there is evidence to suggest that individual differences are a more important contributing factor to face matching performance than training, or length of experience (Bindemann, Avetisyan, & Rakow, 2012; Bobak, Dowsett, et al., 2016; Bobak, Hancock, et al., 2016; Davis et al., 2016; Robertson et al., 2016; White, Kemp, Jenkins, Matheson, et al., 2014). One remaining advantage, is that passport officers can draw on additional information and resources to inform their decisions. In fact, involving another passport officer to lend a fresh pair of eyes could potentially help them overcome biases that are introduced by examining a face more closely. Supporting evidence is provided by the findings that pair working improves face matching accuracy (Dowsett & Burton, 2015), and that response aggregation from expert examiners results in near perfect accuracy (White, Phillips, et al., 2015).

3.4.3 Potential confound

The generalisability of the findings reported in this chapter is potentially limited by the 3:1 ratio of female to male participants. Females have been shown to outperform males on both face recognition (de Frias, Nilsson, & Herlitz, 2006; Herlitz, Nilsson, & Bäckman, 1997; Herlitz, Reuterskiöld, Lovén, Thilers, & Rehnman, 2013), and on face matching tasks (McBain, Norton, & Chen, 2009; Megreya, Bindemann, & Havard, 2011). However, there are three lines of evidence which suggest that the findings reported in this chapter can be generalised to males. First, a recent meta-analysis found that the female advantage in face recognition is mainly attributable to an own-gender bias (Herlitz & Lovén, 2013). Only a small female advantage was found for stimulus sets entirely comprised of male faces, such as the stimulus sets used in the experiments reported here (Herlitz & Lovén, 2013). Second, the female advantage has been found in tasks contexts that rely on memory, or impose time constraints. Neither of these factors were an issue in the studies reported here. Third, an examination of the data for male participants in experiments 1–3, shows the same pattern of results than that observed for the entire sample.

3.4.4 Conclusion

The experiments reported in this chapter examined the effects of comparison on the perceived similarity of unfamiliar face pairs that are relatively easy to tell apart. The current research suggests that comparing the same unfamiliar faces more than once will increase the perceived similarity of the compared faces. This finding has important implications for the security of border control. Higher similarity ratings were obtained for similar face pairs, than for dissimilar pairs (Experiment 1). However, these differences in phenotypic similarity were not found to mediate the effect of comparison on perceived similarity.

Chapter 4: The roles of phenotypic similarity, and of time-to-compare in mediating the effect of comparison on perceived similarity

The three experiments presented in Chapter 3 have shown that comparing two faces by listing commonalities or differences between them is sufficient to increase their perceived similarity. These findings were discussed in relation to the applied context of matching travellers to their photo-identification for border control. This discussion has highlighted two gaps in the research conducted in Experiments 1–3. These gaps are explored in this chapter. The first gap is related to the stimuli used. Whereas the faces that were compared in Experiments 1–3 were shown from different angles, Border Force agents compare a person facing them, to a portrait photo. It is possible that comparing faces from the same view could reduce, or even reverse, the comparison-related increases in perceived similarity that were found in Experiments 1–3. Evidence in support of this argument is provided by the finding that the opportunity to compare two face stimuli that are pictorially identical, and phenotypically difficult to discriminate, decreases the perceived similarity between them (Mundy et al., 2014, 2007). To assess the role of pictorial similarity in mediating the effect of comparison on the perceived similarity of face stimuli, participants in Experiment 4 were asked to compare unfamiliar faces that were shown from the same frontal view.

Another practical implication that has been raised in Chapter 3, is that if a Border Force agent compares a given person-photo pairing repeatedly, the likelihood of a match decision should increase. Participants in Experiments 1–3 performed two discrete comparison tasks (a listing task, and a rating task). The findings reported in Chapter 3 therefore only provide direct evidence that comparison will increase perceived similarity following a previous comparison event. Yet, it is unclear whether Border Force agents subject person-photo pairings to multiple comparisons, or whether they are more likely to compare them once, prolonging the comparison if needed. Research on the time-course of comparison shows that the time available to complete a comparison affects face matching accuracy in the absence of any additional constraints (Özbek & Bindemann, 2011; White, Phillips, et al., 2015). These finding suggests that the perceived similarity of a given stimulus pair can change over the course of a single comparison. The relationship between comparison time and perceived similarity was explored in Experiment 5.

4.1 Experiment 4: The Effect of Comparison on the Perceived Similarity of Morphed faces shown at the same angle

The three experiments reported in Chapter 3 have shown that comparison results in an increase in the perceived similarity of face pairs that are relatively easy to discriminate. This set of results conforms to one of the predictions of structural alignment theory. According to the theory comparison should increase the perceived similarity of unfamiliar faces, because the similarities that are highlighted during the alignment process are rated more highly than the differences (Markman & Gentner, 1993b). However, to increase ecological validity, the faces that were compared in the first three experiments differed both phenotypically (there were differences between the facial features), and pictorially (the faces were presented at different angles). It is possible that the comparison-related increases in perceived similarity found in Chapter 3 are limited to face stimuli that require transformation to be aligned. For face pairs that do not require transformation, habituation to the common features might be the dominant driver of comparison-related changes in perceived similarity. If this is the case, then comparison should result in a decrease in perceived similarity. The first aim of this experiment was to assess whether comparison changes the perceived similarity of face pairs that only vary in their degree of phenotypic resemblance. To isolate the role of phenotypic similarity, both faces in a pair were shown from the same, frontal view. It was assumed that if comparison affects the perceptual similarity of face stimuli that differ phenotypically but not pictorially, then face-pair similarity ratings made after performing a comparison task would differ significantly from similarity ratings that were not preceded by a comparison task.

The second aim of this experiment was to test whether the finding of Experiment 1—that comparison results in an increase in perceived similarity for both similar and dissimilar face pairs—extends to homogenised face pairs shown from the same angle. In particular, this experiment explored whether comparison of similar face pairs would result in an increase in perceived similarity, as predicted by structural alignment theory (Boroditsky, 2007), or in a decrease in perceived similarity, as suggested by habituation theory (Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009). To explore the role of stimulus similarity half of the face pairs were matched to be relatively similar, and half to be relatively dissimilar. The face pairs in the dissimilar condition consisted of unrelated face morphs. The similarity manipulation for similar faces was chosen to emulate the method used in studies of discrimination learning. Comparison of Figures 3 and 16 shows that the similar stimuli used in the current experiment, whilst being very similar, are nevertheless easier to discriminate than those used in discrimination learning studies. The face pairs in the similar condition are however phenotypically more similar than those used in Experiment 1, because the second face in each pair consists of a 50% morph of the first, whereas in Experiment 1, unfamiliar faces were matched based on similarity ratings. It was assumed that if phenotypic similarity mediates the effect of comparison on perceived similarity then there would be a significant interaction between stimulus similarity and perceived similarity.

In the experiments reported in Chapter 3, participants had to explicitly list commonalities and differences. In the discrimination learning literature, on the other hand, participants were not required to make explicit comparisons (e.g. Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009). Therefore, the third aim of this experiment was to assess whether the explicit listing of commonalties (or differences) is a necessary pre-requisite for comparison-induced increases in perceived similarity. To test the role of explicit comparison, two additional conditions were introduced: a pre-exposure condition, and an implicit-comparison condition. It was assumed that if the comparison-induced increases in perceived similarity observed in the first three experiments are limited to explicit comparisons (i.e. listing commonalities or differences), then the implicit-comparison condition would not result in a change in perceived similarity, relative to the no-comparison control. The pre-exposure condition acted as a control for the effect of implicit comparison on perceived similarity. It was assumed that if changes in perceived similarity in the implicit-comparison condition are attributable to pre-exposure, then both the pre-exposure condition and the implicit-comparison condition would be associated with a change in perceived similarity, relative to the no-comparison control. If, on the other hand, they are attributable to the fact that the two stimuli were compared, then there would be no comparison-related change for the pre-exposure condition.

4.1.1 Method

Participants

One hundred and fifty-three Southampton University undergraduates (118 females, 35 males), ranging in age from 18 to 54 years (*Mdn* = 19 years), volunteered to participate. They were randomly allocated to the two between-participant treatment conditions: 76 students participated in the similar-faces condition; the remaining 77 participated in the different-faces condition.

Materials

The stimuli used in the different-faces condition consisted of ten male face averages. Each face average was generated using four colour photos obtained from the Productive Aging Laboratory Face Database (Minear & Park, 2004). The stimuli were presented in pairs. Both faces in a pair were presented in a frontal view (see Figure 16 top row).

For the similar-faces condition 5 additional stimuli were generated by morphing the face pairs in the different-faces condition (see Figure 16 bottom row). For example, if face A and face B were a pair in the different-faces condition, the equivalent pair in the same-faces condition would be Face A and a morphed face that was 50% face A and 50% face B. The morphing was done using Psychomorph version 5.

Figure 16: *An example of the dissimilar (top row) and similar (bottom row) face pairs used in Experiment 4.*

Procedure

The experimental booklets were given to students midway through a lecture. Depending on the between-participant treatment condition the questionnaire contained either five similar or five different face pairs. Each participant completed five within-participant trials: a control, a similarity comparison, a difference comparison, a pre-exposure, and an implicit comparison. The allocation of face pairs to the different treatment conditions was counterbalanced across participants, so that each of the five pairs was used in each condition on an equal number of occasions. The order in which the conditions were presented was randomised across participants.

In the control condition participants were asked to indicate how similar the two faces are on a scale of 1 (*not similar*) to 9 (*very similar*). This is different from the 1 to 7 scale that was used in the previous three experiments. The reason for introducing this scale was a concern that the high similarity of faces in the similar condition (i.e. morphs that had an overall similarity of 50%) may lead to ceiling effects. It was hoped that a scale with a greater range might reveal subtle differences in perceived similarity.

In the similarity-comparison and difference-comparison conditions this similarity rating was preceded by the question “Please describe 3 similarities (or differences) between the two faces”. In the pre-exposure condition the participants were asked to rate whether each face was older or younger than 40 years. The rating was requested for each individual face but the faces were still presented next to each other as in the other conditions. Comparing two men in their early 20s does not aid the judgment on whether each of them is older or younger than 40. It is therefore unlikely that participants will compare the face pairs in the pre-exposure condition. Finally, in the implicit-comparison condition the participants were asked to rate how healthy each of the two faces was. It is assumed that in the absence of a clear standard of what a healthy face looks like, participants are likely to rate the health of the individual faces in a pair in relation to one another.

4.1.2 Results

The mean similarity ratings for the different processing conditions are shown in Figure 17. A 5 (processing: control, similarity comparison, difference comparison, pre-processing, implicit comparison) x 2 (stimulus: similar faces, dissimilar faces) mixed ANOVA revealed a significant main effect of the between-participants stimulus variable, $F(1, 151) = 83.07$, $p < .001$, $\eta_p^2 = .355$. This manipulation check confirms that the similar faces ($M = 5.9$, $SD = 1.6$) were perceived to be more similar to each other than the dissimilar faces ($M = 4.3$, $SD = 1.7$).

There was a significant main effect of the within-participants processing variable, $F(4, 604) = 3.23$, $p < .05$, $\eta_p^2 = .021$. The interaction between stimulus and processing was also significant, $F(4, 604) = 2.40$, $p < .05$, $\eta_p^2 = .016$. To investigate this interaction the data for the similar and dissimilar faces were analysed separately.

Repeated measures ANOVAs revealed a significant main effect of processing for dissimilar faces ($F(4, 304) = 4.19$, $p < .01$, $\eta_p^2 = .052$) but not for similar faces ($F(4, 300) = 0.91$, $p = .459$, $\eta_p^2 = .012$). The data for the similar face pairs was not analysed further. A contrast showed that for the dissimilar face pairs, similarity ratings in the similarity-focussed comparison condition were significantly higher than in the no-comparison control, $F(1, 76) = 15.57$, $p < .001$, $\eta_p^2 = .17$. Similarity ratings in the implicit-comparison condition were also significantly higher than in the no-comparison control, $F(1, 76) = 4.03$, $p < .05$, $\eta_p^2 = .05$. The similarity ratings for the difference comparison condition ($\eta_p^2 = .029$) and pre-exposure condition ($\eta_p^2 = .008$) were not significantly different from the control, $p > .1$.

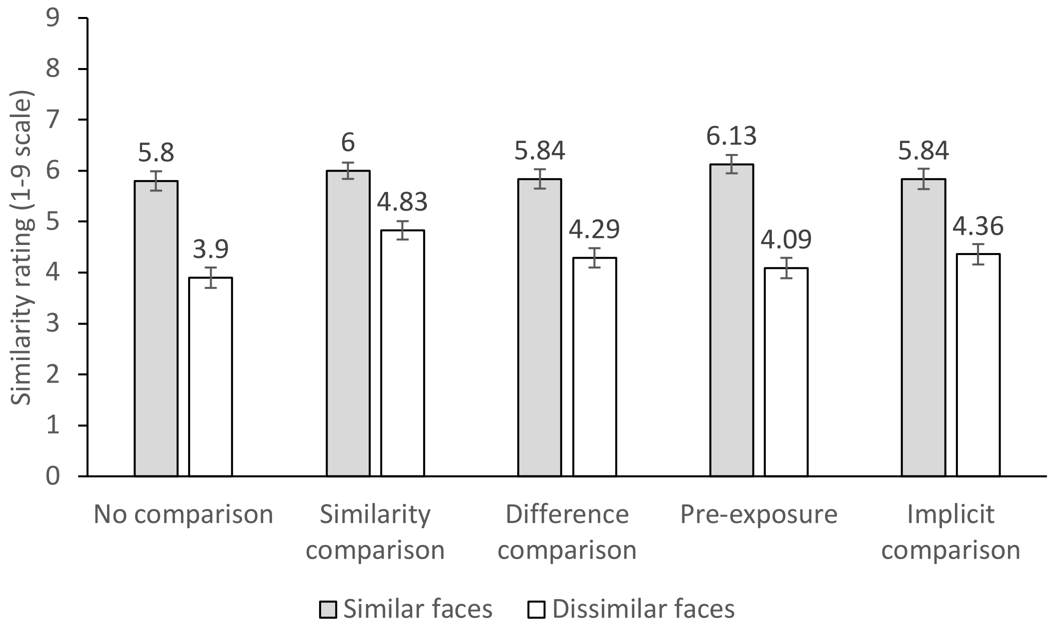


Figure 17: *The mean similarity ratings (and associated standard errors) for the different face-pair pre-processing conditions in Experiment 4.*

4.1.3 Discussion

The first aim of this experiment was to assess whether comparison would result in an increase in the perceived similarity of face stimuli that differ phenotypically but are otherwise identical. For the similar face pairs, none of the experimental manipulations led to an increase in perceived similarity compared to the control. For the dissimilar face pairs, on the other hand, both similarity-focussed comparisons and implicit comparisons resulted in an increase in perceived similarity relative to the no-comparison control. Visual inspection of Figure 17 suggests that difference-focussed comparisons also increased perceived similarity. However, this increase was not statistically significant. The finding that comparison increases the perceived similarity of dissimilar face pairs shown from the same angle demonstrates that pictorial differences are not a necessary pre-requisite to induce comparison related increases in perceived similarity. The comparison-related increases in perceived similarity observed in the first three experiments extend to face pairs that only vary phenotypically, and are easy to align.

The second aim of this experiment was to test whether the finding obtained in Experiment 1—that comparison results in an increase in perceived similarity for both similar and dissimilar face pairs—extends to homogenised face pairs shown from the same angle. One possible interpretation of the null result for the similar face pairs is that for similar face stimuli that are easy to align, any influence of comparison on perceived similarity is either small or non-existent. However, this interpretation is difficult to reconcile with the finding obtained in discrimination learning studies that comparison results in a decrease in perceived similarity of similar faces that are difficult to tell apart. The face pairs used in discrimination learning studies were more similar than the similar face pairs deployed in this experiment. A second possibility, which is compatible with both data sets, is that there are opposing processes at play during comparison, some of which lead to increases in perceived similarity and some to decreases. These opposing processes may have been perfectly balanced for the similar face pairs. A final possibility, which is also compatible with the findings from the discrimination learning literature, is that the rating task may not have been sensitive enough to reveal changes in perceived similarity. Comparison related changes in the perceived similarity of highly similar stimuli have only been observed in studies that deployed indirect measures of similarity. In Experiments 6 and 7 the effect of comparison on perceived similarity was therefore measured using a same–different judgment task.

The third aim of this experiment was to assess whether explicit comparisons are a necessary pre-requisite for comparison-induced increases in perceived similarity. The comparison-related increase in perceived similarity obtained for the implicit comparison condition demonstrates that performing a task that invites comparison between unfamiliar faces—such as assessing the health of the individual faces in a pair—is sufficient to induce comparison-related changes in perceived similarity. An age-categorisation task was used to test whether any comparison-related increases in perceived similarity might instead be attributable to increased familiarity with the compared faces. The age categorisation task did not lead to a significant increase in similarity, compared to the ‘no-comparison’ control. This finding, in conjunction with the finding of Experiment 2 that face descriptions do not increase perceived similarity, suggests that mere familiarity with the faces in a pair is unlikely to be the cause of the comparison induced increase in perceived similarity reported here.

4.2 Experiment 5: The effect of comparison duration on the perceived similarity of face pairs

The first aim of this experiment was to understand the effect of varying the time available to perform a comparison on the perceived similarity of unfamiliar faces. Research on the differential effects of exposure duration on match versus mismatch accuracy has generated conflicting findings. A study in which participants were exposed to face pairs for 200, 500, 1000 and 2000 ms, found that face matching accuracy increased with exposure duration (Özbek & Bindemann, 2011). The increase in performance was driven by an increase in accuracy on match trials. This finding suggests that the perceived similarity of unfamiliar faces increases over the time course of a comparison. The reverse pattern of results was observed in two studies that tested the effects of administering time-pressure flexibly, by setting an overall completion-time target. Both studies found that whilst match accuracy remained relatively constant, performance accuracy on mismatch trials improved (Bindemann et al., 2016; Fysh & Bindemann, 2017). One possible explanation for this discrepancy is that it is attributable to differences in the face stimuli used. However, two of the studies that produced opposing results used the same stimulus set (Bindemann et al., 2016; Özbek & Bindemann, 2011). Another possibility is that the presence of the time pressure task introduced additional response strategies. Participants in the current study were not exposed to any additional time pressure manipulations. It was therefore predicted that longer exposure durations would be associated with higher similarity ratings.

In this experiment, three exposure durations were used to measure time-dependent comparison-related changes in perceived similarity. The shortest exposure duration used was 1 s, the intermediate duration was 3 s, and the longest was 9 s. The choice of exposure duration was informed by research on the time course of comparison. This research has shown that exposure durations of 1 s are sufficient to make simple same–different decisions for a range of visual stimuli (Goldstone & Medin, 1994; Sagi, Gentner, & Lovett, 2012). More nuanced similarity assessments of complex visual stimuli, on the other hand, require exposure durations of 2.7 s seconds or longer (Goldstone & Medin, 1994; Sagi et al., 2012). A study that examined face matching performance in expert forensic examiners and non-experts, found that both groups performed more accurately at long (30 s) exposure durations, than at short (2 s) exposure durations (White, Phillips, et al., 2015). This finding suggests that changes in performance as a function of time available for comparison, can unfold over prolonged exposure durations. The long exposure duration used in the current study was shorter than 30 s. However, there are findings which show that participants do not make full use of the comparison time available, when very long exposure durations are intermixed with shorter exposure durations (Özbek & Bindemann, 2011). The 9 s exposure duration was a trade-off between capturing the effects of longer exposure durations, and ensuring that participants remained engaged.

The second aim of this experiment was to explore whether the effect of comparison duration on the perceived similarity of faces is mediated by the phenotypic similarity of the compared faces. Experiments 1–4 have found comparison-related increases in perceived similarity for a range of face stimuli. Comparison-related decrease in the perceived similarity of unfamiliar faces, on the other hand, have only been observed in discrimination learning studies that have used face stimuli that were identical, except for subtle differences in facial features. Face pairs in the similar faces condition were designed to resemble the face stimuli that were used in studies on discrimination learning. The two faces in a pair consisted of highly similar morphed faces with only 10% variation between them, shown from the same frontal view (see Figure 19). To further increase the homogeneity of the similar faces, the external features between pairs were either identical (e.g. ears, and face shape) or absent (e.g. hair). The stimulus pairs in the dissimilar faces conditions consisted of images of unfamiliar faces that were matched to be relatively dissimilar. Like the similar faces, both faces in a dissimilar pair were shown from the same frontal view.

If phenotypic similarity between faces plays a role in mediating the effects of comparison on perceived similarity, then increasing exposure durations should have a different effect on the perceived similarity of similar and dissimilar face pairs. Based on the findings of Experiments 1–4 (see also Özbek & Bindemann, 2011), it was predicted that the perceived similarity of the dissimilar face pairs would increase with increasing exposure durations. Based on the findings from the discrimination learning literature (Mundy et al., 2014, 2007), it was predicted that the perceived similarity of the similar face pairs would decrease with increasing exposure durations.

4.2.1 Method

Participants

One-hundred-and-forty-six Goldsmiths, University of London undergraduate psychology students participated in exchange for course credit. They were randomly allocated to the two between-participant treatment conditions: 72 students participated in the similar faces condition; the remaining 74 participated in the dissimilar faces condition.

Materials

The stimuli used in the dissimilar faces condition consisted of 24 colour images of male faces taken from the Computer Vision Laboratory, University of Ljubljana, Face Database (http://www.lrv.fri.uni-lj.si/facedb.html). In a pilot study eight volunteers were asked to sort 80 male faces into groups. Each of the twelve dissimilar face pairs consisted of two faces that were clustered together by at most one of the eight volunteers. Examples of the dissimilar face pairs are shown in Figure 18. As can be seen both faces in a pair were shown from a frontal view.

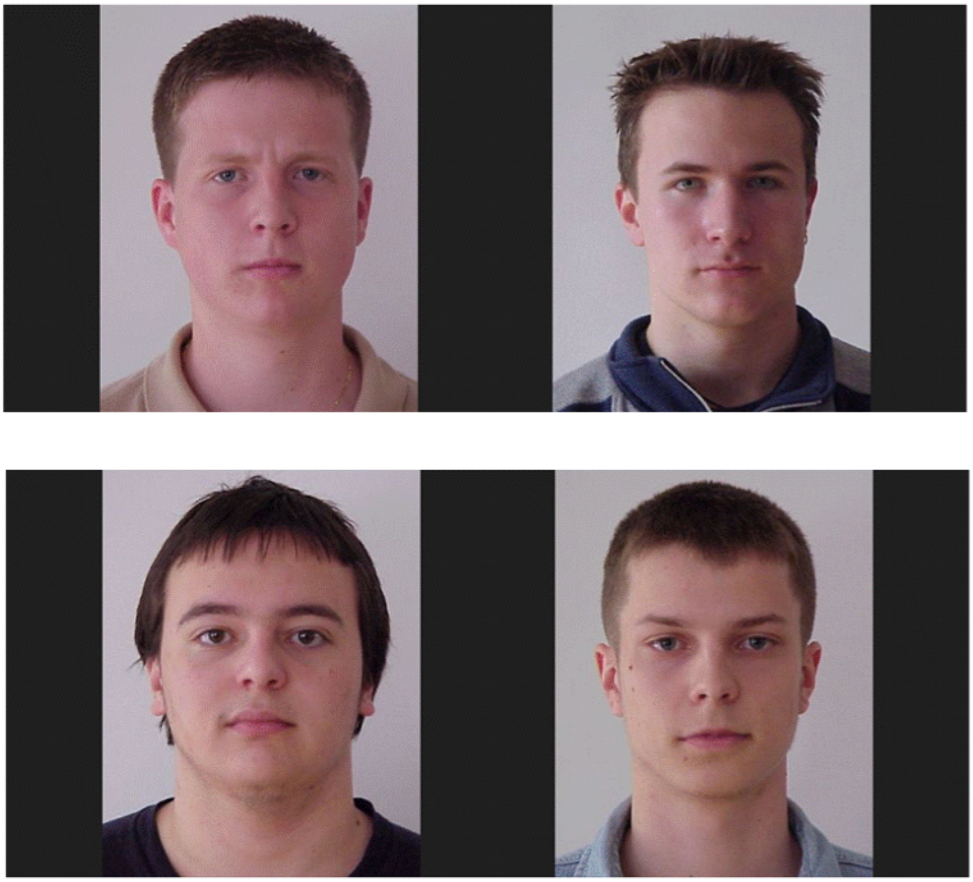


Figure 18: *Two examples of the dissimilar face pairs used in Experiment 5.*

The similar faces are 3D face models that were generated using FaceGen Modeller (version 1). The twelve pairs were generated in a two-stage process. First, twelve 3D face models were generated based on the faces used in the dissimilar condition. Second, similarity matched pairs were created by adding 10% random variation to each of the 12 faces. Examples of the similar face pairs are shown in Figure 19.

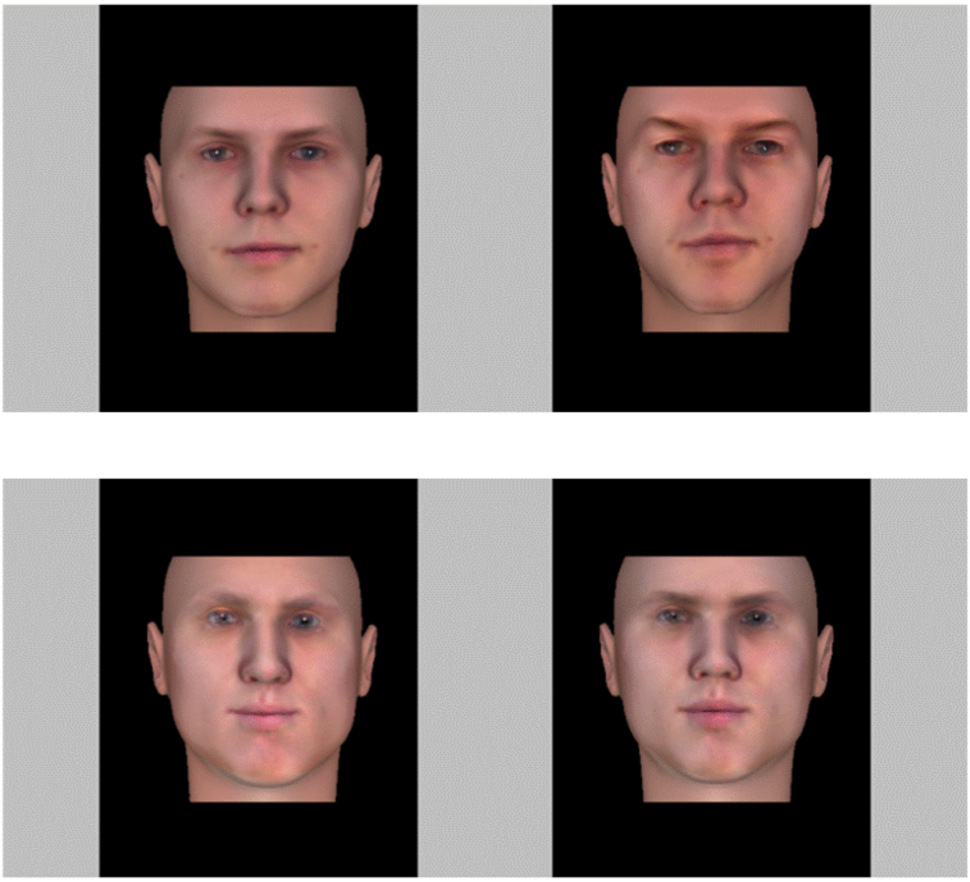


Figure 19: *Two examples of the similar face pairs used in Experiment 5.*

Procedure

Participants were randomly allocated to either the similar or the dissimilar face conditions. Approximately half of the participants were asked to compare similar faces (see Figure 19 for examples) the other half were asked to compare dissimilar faces (see Figure 18 for examples). There were three within-participants conditions: short duration (1 s), medium duration (3 s) and long duration (9 s). Participants received four trials of each condition, presented in a randomised order.

At the beginning of the experiment participants were provided with a brief outline of the experiment and an opportunity to ask questions. After giving informed consent participants read the following instructions on a computer screen:

In this experiment you will be asked to compare pairs of male faces. The face pairs will be presented on the screen for a set time period. The amount of time for which the pairs will be visible will vary from trial to trial. Please compare the faces for the entire duration they are shown. After the faces have disappeared from the screen you will be asked to rate their similarity on a scale of 1 (not at all similar) to 7 (very *similar) using the number keys on the keyboard. Please press the space bar to continue.*

The experiment began with two practice trials to enable participants to gain familiarity with the experimental procedure and the variations in presentation duration. The practice trials were followed by the twelve experimental trials. Each new trial was initiated by a space bar press. In response to a space bar press a face pair would appear on the screen for a fixed duration of 1, 3 or 9 s. After that the pair would disappear from view and the participant would be prompted to rate the similarity of the face pair on a scale of 1 (*not at all similar*) to seven (*very similar*). Participants were not able to rate the similarity of a face pair before it had disappeared from the screen. Presentation order was randomised for each participant. After the final trial participants were thanked for taking part and debriefed.

4.2.2 Results

Figure 20 shows the mean similarity ratings for the high-similarity and the low-similarity face pairs as a function of presentation duration. A 3 (presentation duration: short, medium, long) x 2 (stimulus similarity: similar faces, dissimilar faces) mixed ANOVA revealed a significant main effect of the between-participants similarity variable, $F(1, 144) = 131.87$, $p < .001$, $\eta_p^2 = .478$. This manipulation check confirms that the similar faces ($M = 4.2$, $SD = 0.8$) were perceived to be more similar to each other than the dissimilar faces ($M = 2.5$, $SD = 1.0$). There was a significant main effect of the within-participants duration variable, $F(2, 288) = 7.49$, $p < .01$, $\eta_p^2 = .049$. The interaction between similarity and presentation duration was not significant, $F(2, 288) = 2.52$, $p = .082$, $\eta_p^2 = .017$.

A repeated measures ANOVA of the similar faces data showed a significant main effect of duration, $F(2,142) = 4.01$, $p < .05$, $\eta_p^2 = .02$. Repeated contrasts revealed that face pairs in the medium duration condition were perceived to be more similar than face pairs in the short duration condition, $F(1, 71) = 8.3$, $p < .01$, $\eta_p^2 = .105$. Repeated contrasts further showed that face pairs in the long duration condition were perceived to be less similar than face pairs in the medium duration condition, $F(1, 71) = 4.27$, $p < .05$, $\eta_p^2 = .057$.

A repeated measures ANOVA of the dissimilar faces data showed a significant main effect of duration, $F(2,146) = 6.57$, $p < .01$, $\eta_p^2 = .083$. Repeated contrasts revealed that face pairs in the medium duration condition were perceived to be more similar than face pairs in the short duration condition, $F(1, 73) = 5.14$, $p < .05$, $\eta_p^2 = .066$. There was no significant difference between the long duration condition and the medium duration condition, $F(1, 73) = 1.52$, $p = .221$, $\eta_p^2 = .02$.

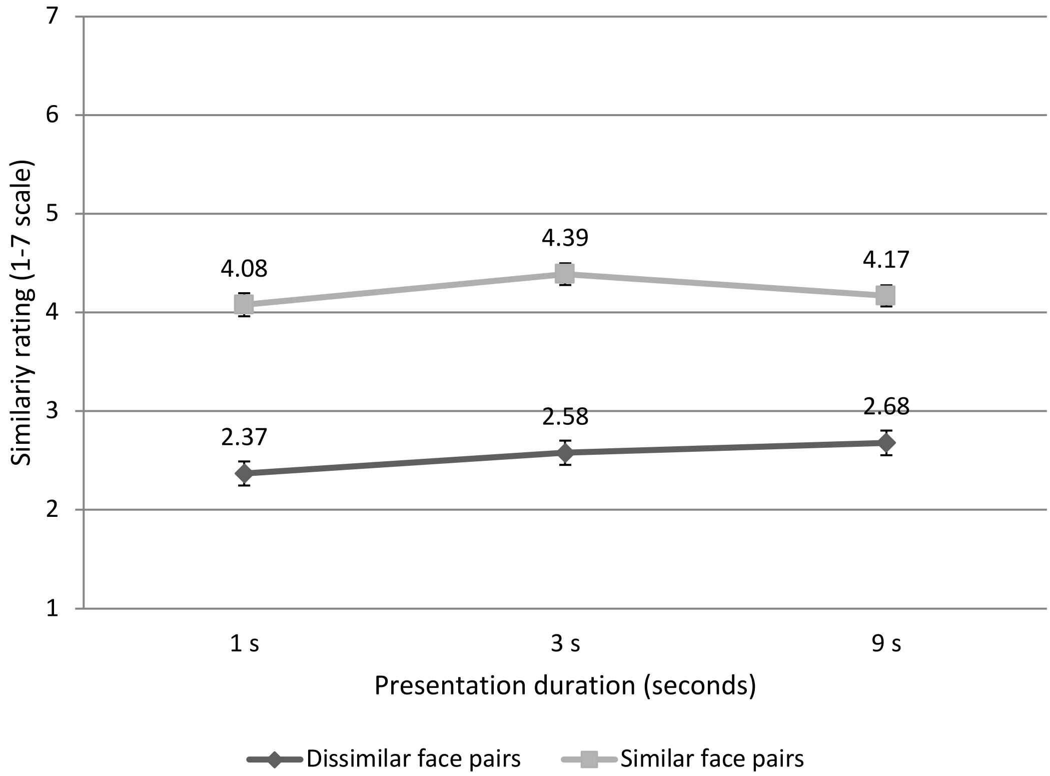


Figure 20: *The mean similarity ratings obtained in Experiment 5 for the similar and dissimilar face pairs, and for all face pairs. The mean similarity ratings (and associated standard errors) are plotted as a function of presentation duration.*

4.2.3 Discussion

The first aim of this experiment was to explore whether the amount of time available for a comparison affects the perceived similarity of the compared faces. As can be seen in Figure 20, similarity ratings for both the similar and the dissimilar face pairs were higher in the medium exposure duration condition, than in the short exposure condition. This finding suggests that increasing the time available for a comparison increases the perceived similarity of the compared faces.

The second aim of this experiment was to assess whether the effect of comparison duration on perceived similarity is mediated by the phenotypic similarity of the compared faces. Visual inspection of Figure 20 suggests that longer presentation durations have different effects on the perceived similarity of phenotypically similar faces, than they do on the perceived similarity of phenotypically dissimilar faces. For the similar face pairs, perceived similarity was significantly lower in the long exposure condition than in the medium exposure condition. For the dissimilar face pairs, on the other hand, similarity ratings in the long exposure condition were slightly higher than in the medium exposure condition. However, this difference was not significant.

This pattern of results suggests that there might be two processes at play in determining the changes in perceived similarity over the time course of comparison. Research on the time course of structural alignment suggests that whereas same–different judgments for complex visual stimuli can be achieved at 1 s exposure durations, a full alignment takes approximately three seconds to complete (Goldstone & Medin, 1994; Sagi et al., 2012). The choice of the short and medium exposure durations was informed by these findings. According to structural alignment theory, comparison increases the perceived similarity of similar stimuli because it highlights the commonalties between them, which are weighted more highly than the differences (Markman & Gentner, 1993b). It can therefore be argued that the higher similarity ratings in the medium exposure duration condition, relative to the short exposure duration condition, are attributable to effects of structural alignment.

The second process is one of habituation. According to the habituation account perceived similarity for similar stimuli decreases as a function of comparison, because the observer habituates to the commonalities between the stimuli. The choice of stimuli in the similar faces condition was based on the stimuli for which comparison-related decreases in perceived similarity have been found. The similar face pairs were designed to be particularly conducive to habituation because they are phenotypically very similar (only 10% variation between the faces). Habituation to shared stimulus attributes may have caused the lower similarity ratings in the long exposure duration condition, relative to the medium exposure duration condition.

The interaction between stimulus similarity and exposure duration was not significant. One possible explanation for this finding is that the choice of exposure durations did not allow for the full unfolding of the effect of habituation. Whereas the impact of the structural alignment process appears at exposure durations of 3 s, the effects of habituation appear to unfold more slowly. To capture the effects of habituation more fully, this experiment needs to be repeated with two additional exposure durations, one at 6 s (to capture the unfolding of the effect of habituation) and one at 12 or 15 s (to capture the full effect of habituation). Conversely, the homogeneity and high similarity of the similar face pairs, made the similar stimuli particularly conducive the effects of habituation. It is possible that, had the two faces in the similar faces condition been shown from two different angles (or had other pictorial differences been introduced), the drop in perceived similarity between the medium and long exposure duration may have been reduced or even reversed. To fully understand how exposure duration interacts with stimulus similarity a wider range of unfamiliar face stimuli needs to be tested that vary systematically in their degree of phenotypic and pictorial similarity.

4.3 Chapter 4 discussion

This chapter addressed two questions that emerged when discussing the implications of the findings of Experiments 1–3 to the applied context of border control. The first question relates to the unfamiliar face stimuli that were used. Would comparison still result in an increase in perceived similarity if the compared faces are shown from the same angle, as they are when comparing a person to a passport photo? Evidence from the discrimination learning literature suggests that this might not be the case (Mundy et al., 2014, 2007). Participants in Experiment 4 were asked to compare pairs of similar and dissimilar face morphs shown from the same angle. The study found comparison-related increases in perceived similarity for the dissimilar but not for the similar face morphs. The comparison-related increase in perceived similarity was found for a similarity-focussed, and an implicit comparison task. The difference-focussed comparison condition failed to reach significance. Together the findings of Experiment 4 suggest that the effects of comparison on perceived similarity are attenuated for stimuli that are both phenotypically similar, and pictorially homogenous.

The second question relates to the nature of the comparison task. Would comparison-related changes still be seen over one continuous comparison, or are the effects observed in Chapter 3 limited to instances in which two comparison tasks (e.g. the similarity-listing task and the similarity rating task) are performed in succession? To address this question participants in Experiment 5 were asked to compare similar and dissimilar face morphs that were presented for three different presentation durations. For both the similar and the dissimilar face morphs perceived similarity was higher at the medium duration, relative to the short duration (see Figure 20). Furthermore, perceived similarity of the similar faces was lower at the long presentation duration, than at the medium presentation duration. For the dissimilar morphs, on the other hand, there was no significant difference between the similarity ratings at the medium presentation duration, and the long presentation duration condition. The findings of Experiments 4 and 5 have implications both for theories of comparisons, and in applied contexts.

4.3.1 Theoretical implications

The comparison-related increases in perceived similarity observed in Chapter 3 constitute a first demonstration that the prediction of structural alignment theory—that comparison will increase the perceived similarity of similar stimuli that share many commonalities and differences—extends to unfamiliar face pairs. However, the findings of Experiments 4 and 5 suggest that two opposing processes might interact in determining the effect of comparison on perceived similarity.

Experiment 4 found a comparison-related increase in perceived similarity for dissimilar face morphs (which were created by morphing two different faces) but not for similar face morphs (which where 50% morphs of each other). One possible interpretation of the null result is that it reflects the influence of a habituation process, which opposes the influence of structural alignment. Alternatively, it may simply reflect the fact that similarity ratings are not sensitive enough to capture small changes in the perceived similarity of faces that are phenotypically very similar. Evidence against the latter interpretation is provided by the findings of Experiment 5. In this experiment, comparison-related decreases in similarity ratings were found for face morphs that resemble each other more closely (there was only 10% variation between them) than those used in Experiment 4.

Habituation has been used to explain comparison-related improvements in performance on same–different judgement tasks for unfamiliar faces (Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009). The findings of Experiments 4 and 5 are consistent with the notion that habituation plays a role in mediating the effects of comparison on perceived similarity. However, the face stimuli were designed to resemble those used in discrimination learning studies. Habituation is a form of learning that helps an organism ignore irrelevant information (Kohn, 2007). It follows that the effects of habituation on perceived similarity may not be limited to stimuli that are highly similar (both phenotypically and pictorially), but also impact on the perceived similarity of unfamiliar face stimuli that have many pictorial differences that make face-specific information difficult to discriminate. In this context, habituation to the 'visual noise' would be beneficial to the comparison process. The two experiments presented in Chapter 4 assessed the role of habituation in mediating the effects of comparison on perceived similarity with face pairs that were not pictorially homogenous.

4.3.2 Applied considerations

The findings reported in Chapter 3 suggest that the act of subjecting a given person-photo pairing to additional scrutiny might increase the likelihood that the pairing is considered a match. However, whereas the person-photo pairings that are compared in the context of border control are seen from the same angle, the stimuli used in Experiments 1–3 were shown from different angles. Experiment 4 assessed whether the findings reported in Chapter 3 extend to faces that are shown from the same frontal view. The experiment found comparison-related increases in perceived similarity for the dissimilar face pairs. This finding suggests that comparison-related increases in perceived similarity might affect face matching performance in the context of border control.

For the similar face pairs, no comparison-related increase was found. This finding could be interpreted as evidence that the findings reported in Chapter 3 are only of limited relevance to the context of border control, because intuitively, the faces that are the most similar are also the faces that are most likely to result in false matches. However, systematic biases towards responding that two faces are the same person have been demonstrated for face pairs that were less similar than the similar faces in Experiment 4 (Alenezi & Bindemann, 2013; see also Alenezi et al., 2015).

Another implication of the findings obtained in Chapter 3 for the border control context is that subjecting a given mismatched pair to intense scrutiny might increase the likelihood that they are erroneously considered a match. It is unclear whether a border force agent is more likely to subject face-photo pairings that are difficult to match (because a lot of time has passed since the photo was taken) to multiple comparisons, or whether they will be examined for longer. The findings reported in Chapter 3 provide evidence that comparison-related increases might affect decisions in the former case. To explore the latter case, Experiment 5 examined changes in perceived similarity as a function of exposure duration. The study found a comparison-duration related increase in perceived similarity for dissimilar face pairs shown from the same angle. This finding supports the notion that comparison-related increases in perceived similarity may affect performance at border control.

4.3.4 Conclusion

The experiments reported in this chapter examined whether the findings obtained in Chapter 3 extend to face pairs that are shown from the same angle (Experiment 4) and to task contexts that involve a single prolonged comparison rather than two, or more, discrete comparisons (Experiment 5). Experiment 4 found a comparison-related increase in perceived similarity for pairs of dissimilar face morphs. Experiment 5 found that the perceived similarity of dissimilar faces increases with time to compare. Together these results support the argument made in Chapter 3, that the process of subjecting a face pair to more scrutiny might influence the outcome of the same–different decision.

A different pattern of results emerged for pairs of highly similar face morphs. Experiment 4 found no comparison-related changes for pairs of face morphs that overlapped by 50% in their similarity. Experiment 5 found that the perceived similarity for similar face morphs is lower at a long exposure duration than at a medium exposure duration. Together these findings suggest that habituation plays a role in determining the effect of comparison on perceived similarity. The next chapter will explore the role of habituation in similarity by using face stimuli that are difficult to compare because pictorial differences degrade the quality of face-related information available.

Chapter 5: Does comparison exert opposing effects on the perceived similarity of degraded and non-degraded unfamiliar face stimuli?

The two preceding chapters have focussed on the applied context of border control. However, there are many other applied situations that involve face comparison. Consider, for example, the task of matching a suspect in the dock, or a photo of a suspect, to CCTV footage of the perpetrator. Whilst there are many parallels between performing a face comparison in this context, and in the context of border control, there are also critical differences. One difference, of particular relevance to the aims of this thesis, is that whereas faces in passport photos are presented in a standardised format, the unfamiliar face stimuli in CCTV footage, are not. A variety of pictorial variables affect the amount and quality of face-specific information available in CCTV footage. These pictorial differences stem from system-level variables (e.g. the position of the camera), situational variables (e.g. ambient light conditions), person-specific variables (e.g. wearing a hat), and interactions between these variables (Ao, Yi, Lei, & Li, 2009). An example of an interaction between system level and person-specific variables, is that the spatial relationship between the CCTV camera and the person being recorded affects the amount of face specific information available for analysis.

According to a British Security Industry Association (BSIA) estimate, there are currently 4 to 6 million CCTV cameras in the UK ("CCTV: Too many cameras useless, warns surveillance watchdog Tony Porter," 2015). Matching a face to CCTV footage is therefore a relatively common task in both forensic and judicial contexts. For example, 95% of murder case investigations undertaken by Scotland Yard in 2009, used CCTV footage as evidence ("The end of the CCTV era?," 2015). Furthermore, it is acceptable practice in UK law for a jury to compare a defendant in the dock to CCTV footage of the culprit, provided the footage is sufficiently clear (Davis & Valentine, 2009; Davis et al., 2012). The pictorial manipulations used in Experiments 1–3 were limited to changes in pose, and rotations in the picture plane.

They therefore only reflect a small subset of the pictorial differences that are encountered when comparing CCTV footage to a photo, or to a person present.

Experiments 1–3 have shown that comparison increases the perceived similarity of unfamiliar faces that are relatively easy to tell apart. For unfamiliar face stimuli that are relatively difficult to tell apart, a different pattern of results emerges. No comparison-related increase in perceived similarity was found in Experiment 4, for phenotypically similar face morphs shown from the same angle. Furthermore, comparison has been shown to decrease the perceived similarity (operationalised as accuracy on discrimination and categorisation learning tasks) of pictorially homogenous face stimuli that were designed to look like identical twins (Mundy et al., 2014, 2007; Mundy, Honey, Downing, et al., 2009). One potential causal mechanism for the comparison-related decrease in perceived similarity is habituation to shared stimulus features (see Mundy et al., 2007; Mundy, Honey, & Dwyer, 2009). Habituation is a form of learning that enables individuals to adapt to their environment by tuning out irrelevant information. It follows that comparison-related decreases in perceived similarity may not be limited to unfamiliar face stimuli that are pictorially homogenous. Habituation might also play a role in comparisons that involve a stimulus in which the signal (i.e. the face) is surrounded by noise (i.e. pictorial attributes that degrade the amount, or quality of visual information available). A prediction that follows from this is that comparison might also decrease the perceived similarity of unfamiliar face stimuli, if at least one of the stimuli is of low image quality. The hypothesis that degrading the quality of one of the compared stimuli will result in a comparison-related decrease in perceived similarity, was tested in Experiments 6 and 7.

In Experiment 6 visual noise was added to one of the faces in a pair by pixelating it. This manipulation was chosen to mimic the reduction in image resolution that arises when low resolution CCTV footage is used, or when footage captured with a wide-angle lens is enlarged. A study on the effect of image pixelation, a pictorial manipulation similar to the one deployed in Experiment 6, found a gradual decrease in face matching accuracy with increasing levels of pixelation (Bindemann et al., 2013). In Experiment 7 participants were asked to compare high quality stills with CCTV footage of a person walking. These stimuli were chosen because they reflect the types of stimuli that might be compared in criminal investigations. They have been shown to result in mistaken identifications in a previous study (Davis & Valentine, 2009).

To measure perceived similarity participants in Experiments 6 and 7 were asked to decide whether two unfamiliar faces were the same person. The dependent variable was chosen to increase the ecological validity of the experiments reported in this chapter. Same–different decisions are common in applied settings, such as criminal investigations and passport control. Similarly, jury members can also be asked to decide whether a photo of the culprit matches the defendant in the dock (Davis et al., 2012). The use of both direct (e.g. similarity ratings) and indirect (e.g. accuracy) measures of similarity in this thesis, rests on the assumption that both reflect the same underlying similarity construct. This issue is discussed in Chapter 1, which concludes that although the two measures capture slightly different aspects of the similarity construct, overall, they reflect the same construct.

Some of the variance between the two measures is caused by methodological differences. For example, whereas same–different judgement tasks are generally performed under timed conditions, both during the initial pre-exposure phase and at test, studies that deploy similarity ratings are generally untimed. The fact that same–different judgments are often performed under time-constraints, rather than the use of a same–different judgment per se, may affect the similarity construct that is captured. To control for the effects of time-pressure, participants were free to perform the similarity- and difference-listing tasks without time constraints. The same–different judgement tasks were also untimed.

5.1 Experiment 6: The effect of comparison on same–different judgments for difficult to discriminate faces

The first aim of this experiment was to assess whether, and how, comparison affects the perceived similarity of unfamiliar face stimuli that differ in image quality. The pictorial manipulation consisted of adding visual noise to one of the images in a pair, to mimic poor quality CCTV footage. The stimuli used in this experiment were the similar face pairs used in the Experiment 1. For stimuli that are relatively easy to discriminate, same–different ratings are not a sensitive measure of changes in perceived similarity. The assumption that the dissimilar face pairs—even with added pixelation (see Figure 21)— would be too easy to discriminate under unlimited viewing conditions, to reveal changes in perceived similarity, was tested in a pilot study with 47 participants. The study found that performance accuracy was high in all conditions, the proportion of correct, 'different' responses was 89% in the no-comparison control condition, 91% in the similarity comparison condition, and 96% in the difference comparison condition. A Cochran's Q test, confirmed that there were no significant differences in the proportions of 'same' responses between the three treatment conditions, χ^2 (2) = 1.27, p = 0.53. The dissimilar face pairs were therefore excluded to avoid ceiling effects.

The second aim of the current study was to assess the merits of an alternative interpretation for comparison-related changes in the proportion of 'same' responses. Namely, that they reflect changes in observers' ability to accurately tell the unfamiliar faces apart. An additional same face condition was therefore introduced, in which both images in a pair depicted the same person. It was assumed that if a comparison-related increase in 'same' responses for the match pairs, is associated with an equivalent decrease in the number of 'same' responses for the mismatch pairs, this would suggest that comparison improves accuracy on the task. If, on the other hand, the number of 'same' responses for both the match and the mismatch pairs either increases or decreases (i.e. if the direction of change is the same for both), this would suggest that comparison alters the perceived similarity of the stimuli.

Figure 21: *An example of the pixelated dissimilar face pairs used in the pilot study of Experiment 6.*

5.1.1 Method

Participants

One hundred and fifty-seven University of Southampton undergraduate students (108 females, 49 males), ranging in age from 18 to 30 years (*Mdn* = 19 years), volunteered to take part. They were randomly allocated to the two treatment conditions: 80 students participated in the similar faces condition; the remaining 77 participated in the same faces condition.

Materials

Examples of the stimuli used are shown in Figure 22. The stimuli consisted of pairs of black-and-white male faces shown from different angles. The stimuli were modified versions of the face stimuli used in Experiment 1. To simulate low quality CCTV imagery, and to make the stimulus pairs less easy to discriminate, one face in each pair was manipulated using Corel Photo-Paint version X3. The manipulations consisted of Gaussian noise, and a canvas texture effect. For face pairs in the same face condition the face that was manipulated was an identity match to the non-manipulated face; for similar face pairs the manipulated face was matched to be similar (see Experiment 1 for a description of the matching procedure). In each pair the non-manipulated face was shown from a frontal view and the manipulated face from a three-quarter view.

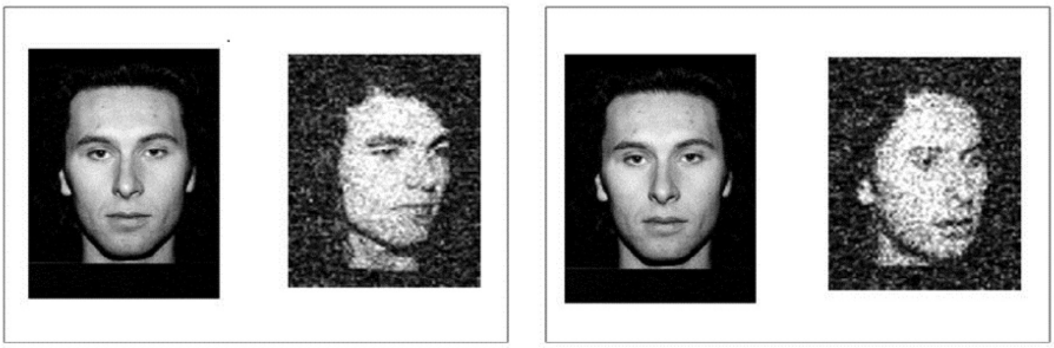


***Figure 22:** An example of the similar (left) and same (right) face pairs that were used in Experiment 6.*

Procedure

The participants were asked to fill in a short five-page booklet at the beginning of a lecture. Depending on the treatment condition the questionnaire contained either three same face pairs, or three similar face pairs. Each participant completed a control condition, a similarity-focused comparison condition, and a difference-focused comparison condition. In the no-comparison control condition participants were asked to decide whether the two images are of the same person or not. In the similarity-focused and difference-focused comparison conditions this yes-no rating was preceded by the question "Please describe 3 similarities (differences) between the two faces". The allocation of face pairs to the treatment conditions, and the order of presentation were counterbalanced across participants.

5.1.2 Results

The dichotomous same–different rating data was subjected to Cochran's Q tests to determine whether there are significant differences between the related treatment conditions. There is no equivalent test to the mixed ANOVA for dichotomous data. The interaction between face pair identity (same or similar) and the comparison conditions (similarity-focussed, difference-focussed, no-comparison control) was therefore not analysed.

The percentage of 'same' responses for each of the treatment groups is shown in Figure 23. A Cochran's Q test of the same–different ratings for the similar face pairs revealed that 'same' responses were significantly less likely in the similarity-focused comparison condition, than in the no-comparison control condition, $Q(1)= 3.93$, $p < 0.05$, $\eta^2 = .049$. 'Same' responses for similar faces were also significantly less likely following difference-focused comparisons relative to the no-comparison control, $Q(1)= 6.4$, $p < 0.05$, $\eta^2 = .08$.

As can be seen in Figure 23, for the face pairs in the same face pair condition there were fewer 'same' responses after similarity-focused comparisons than for the no-comparison control. This difference in the proportion of same responses did not reach statistical significance, $Q(1)= 3.67$, $p = 0.056$, $\eta^2 = .046$. Visual inspection of Figure 23 further reveals a small difference in the number of 'same' responses between the difference-focused comparison condition and the no-comparison control condition for the same face pairs. This difference was not statistically significant, $Q(1) = 0.73$, $p = 0.39$, $\eta^2 = .009$.

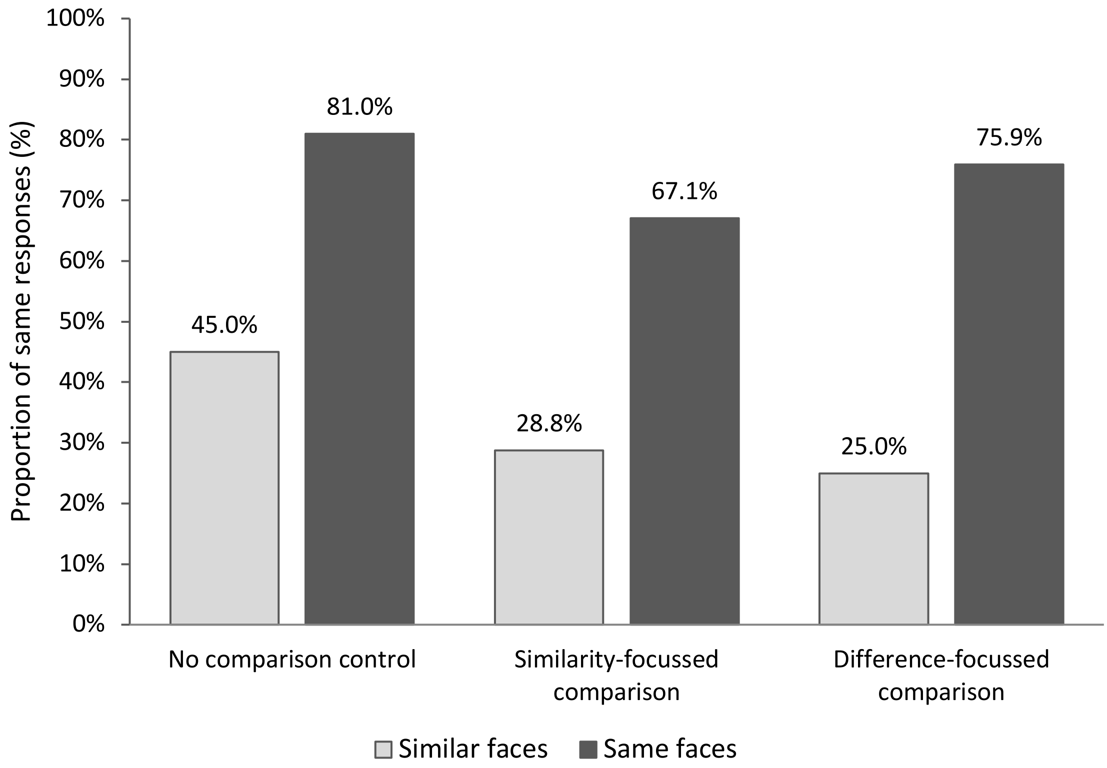


Figure 23: *The proportion of 'same' responses for similar and same face pairs found in Experiment 6.*

5.1.3 Discussion

For the similar face pairs both similarity- and difference-focussed comparisons resulted in a significant decrease in the proportion of 'same' responses, relative to the control. As can be seen in Figure 23, a similar pattern of results was observed for the same face pairs. However, for the same face pairs the comparison-related decreases in perceived similarity were not statistically significant. Nevertheless, taken together the findings for the similar and same face pairs suggest that comparison reduces the perceived similarity of unfamiliar face stimuli that are difficult to discriminate due to differences in image quality. This pattern of results does not conform to the alternative argument that comparison simply improves face matching accuracy. If this were the case, there should have been a comparison-related increase in the number of 'same' responses for the same face pairs.

It is important to note that discrimination learning studies that have analysed performance on match and mismatch trials separately have shown that the comparison-related improvements in accuracy are driven by improvements on different trials (Lavis et al., 2011; Lavis & Mitchell, 2006; T. Wang et al., 2012). Performance accuracy on same trials is generally high, regardless of treatment condition because matching identical stimuli is easier than discriminating between highly similar stimuli. The findings from discrimination learning studies are therefore compatible with the notion that comparison decreases the perceived similarity of unfamiliar face pairs. It is also important to note that even though comparison does appear to reduce the perceived similarity of both similar and same face pairs, this effect is attenuated for the same face pairs, suggesting that some discrimination learning also occurred. These issues will be further discussed in the general discussion of this chapter.

Unlike the findings of Experiments 1–4, the results of this experiment do not conform to the predictions of structural alignment theory. Instead these findings support a hypothesis that was derived from the habituation account of discrimination learning. According to this hypothesis, comparison should increase the perceived similarity of unfamiliar face stimuli that are difficult to compare because the image quality of one of the face stimuli is degraded. In this context, habituation to visual noise was predicted to direct attention towards the faces. The faces, in turn, are likely to look rather dissimilar due to the many pictorial differences between the stimuli. The findings of the current study lend support to the notion that structural alignment theory and the habituation account of comparison need to be combined, to fully account for the effects of comparison on the perceived similarity of unfamiliar face stimuli. This topic will be further discussed in the general discussion of this chapter.

It could be argued that the decrease in similarity ratings following difference-focused comparisons supports the notion that difference-focussed comparisons enhance perceptual differences (Hassin, 2001). However, the finding that similarity-focused comparisons also decrease perceived similarity stands in direct contrast to the complementary argument that similarity judgments enhance perceptual similarities. Furthermore, the finding that both types of comparison resulted in a decrease in perceived similarity, indicates that participants were not biased to respond in accordance with perceived task demands.

In sum, the current study has found that both similarity- and difference-focused comparisons decrease the perceived similarity of unfamiliar face stimuli that are difficult to discriminate, due to pixelation of one of the stimuli. The fact that the proportion of 'same' ratings was lower following comparison of both similar and same face pairs suggests that this effect is not attributable to an improvement in the ability to tell the faces apart. Experiment 7 tested the robustness of the finding that comparison decreases the perceived similarity of unfamiliar face stimuli if the quality of one of the stimuli is degraded, with a different stimulus set.

5.2 Experiment 7: The effect of comparison on same–different judgments for pairs of static and non-static stimuli

In this experiment participants were asked to decide whether a face shown in a photo is the same person than a face shown in a video clip. The reasons for using video footage were twofold. First the video footage was used to test the robustness of the finding obtained in Experiment 6—that comparison reduces the perceived similarity of unfamiliar face stimuli, if the quality of face specific information is reduced. The second reason for deploying video clips was that they are used in forensic investigations, for example, when matching a photo to CCTV footage. Knowing whether the findings of Experiment 6 extend to these stimuli therefore has important practical implications. This experiment also tested whether the decreases in perceived similarity observed in Experiment 6 might be attributable to participants' greater familiarity with the stimuli. As in Experiment 2, a face description task was used to test this alternative hypothesis.

The stimulus pairs were comprised of a black-and-white high resolution image, paired with a medium range video clip showing a person performing choreographed actions. The amount of face-specific information available in these clips was limited, both by the fact that the recording showed the whole person (and the face therefore took up a relatively small proportion of the scene), and by the fact that the image resolution was lower in the video footage (see Figure 24 for an example). The two faces in a pair were matched to phenotypically resemble each other (see Davis & Valentine, 2009). As in Experiment 6, the proportion of 'same' ratings was used as a measure of perceived similarity.

5.2.1 Method

Participants

Participants were recruited on two open days, one held at Totton College and one held at Goldsmiths, University of London. One hundred and seventy-nine open day visitors (136 females, 43 males), ranging in age from 17 to 63 years (*Mdn* = 18 years), volunteered to take part. They were randomly allocated to the four treatment conditions: 44 participated in the no-comparison control condition; 45 in the similarity-focused comparison condition; 45 in the difference-focused comparison condition; and the remaining 45 participated in the face-description condition.

Materials

The stimuli used were black-and-white photos and short video clips of four male rugby players (see Figure 24 for an example). The black-and-white photos show a headshot of the player in a three-quarter view. The colour video clips show each player walking, briefly facing the camera and ascending stairs. Each matched pair consisted of two different individuals. The players were matched into pairs based on the likelihood that they would be mistaken for one another. A detailed description of how the stimuli were created and paired can be found in Davis and Valentine (2009).

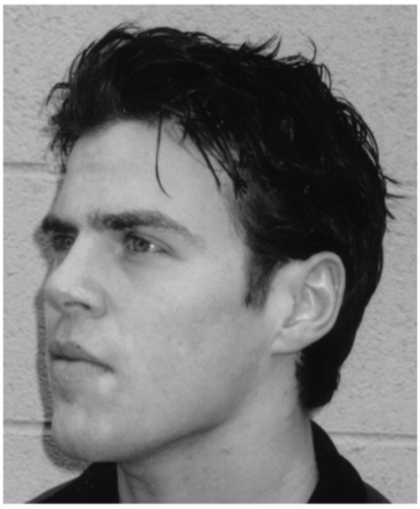


Figure 24: *One of the two player pairs used in Experiment 7. The black and white photo that was included in the experimental booklet is shown on the left. A still from the video clip that was looped is shown on the right. Both player pairs were individuals that have been shown to be easily mistaken for one another (Davis & Valentine, 2009). In particular the pair shown above was mistaken for each other on 67% of different trials when they were the first or only pair to be compared, and on 44% of different trails when they were the second pair to be compared (Davis & Valentine, 2009)*

Procedure

Participants were tested in small groups of five to ten. Each test session started with a short introduction to the experiment during which informed consent was obtained. Following the introduction, a short video clip of one of the rugby players was projected on a screen in the front of the room. The video clip was continuously looped until all participants had completed the experimental task and had handed back their booklets. Each participant took part in one of four conditions: a no-comparison control condition, a similarity-focused comparison condition, a difference-focused comparison condition, and a face-description condition. Participants were asked to remain silent and to keep responses private. They were given performance feedback at the end of the session.

Participants in the no-comparison control condition had to decide whether the face in the photo and the face in the video footage were the same person or not (i.e. "*yes, same*" or "*no, different*"). Participants in the similarity-focused condition had to describe three similarities between the two faces before making a judgment about whether they are the same person. Similarly, participants in the difference-focused condition had to describe three differences before judging whether the two faces were the same. Finally, participants in the face-description condition had to first give a brief description of the face in the photo, then a brief description of the face in the video before judging whether the two faces were the same.

5.2.2 Results

The dichotomous same–different rating data was subjected to Chi-square tests to determine whether there are significant differences between the unrelated treatment conditions. The proportion of 'same' responses in the three pre-processing conditions (similarity-focussed comparison, difference-focussed comparison, describe) was compared to the proportion of same responses in a no pre-processing control.

The proportion of same responses is shown in Figure 25. As can be seen in Figure 25, participants in the similarity-focused comparison condition were less likely to respond that the two faces were the same than participants in the no-comparison control condition. A Chi-square test revealed that this difference in performance on the same–different judgment task was significant $\chi^2(1, N = 99) = 4.54$, $p < 0.05$, $\Phi_c = 0.211$. Visual inspection of Figure 25 further shows that there were fewer same responses in the difference-focused comparison condition than in the no-comparison control. However, a Chi-square test showed that this difference was not statistically significant, $\chi^2(1, N = 99) = 0.75$, $p > 0.1$, $\Phi_c = 0.028$. Finally, participants in the face-description condition also made fewer same responses than participants in the control condition. This difference was not statistically significant, $\chi^2(1, N = 99) = 1.17$, $p > 0.1$, $\Phi_c = 0.138$.

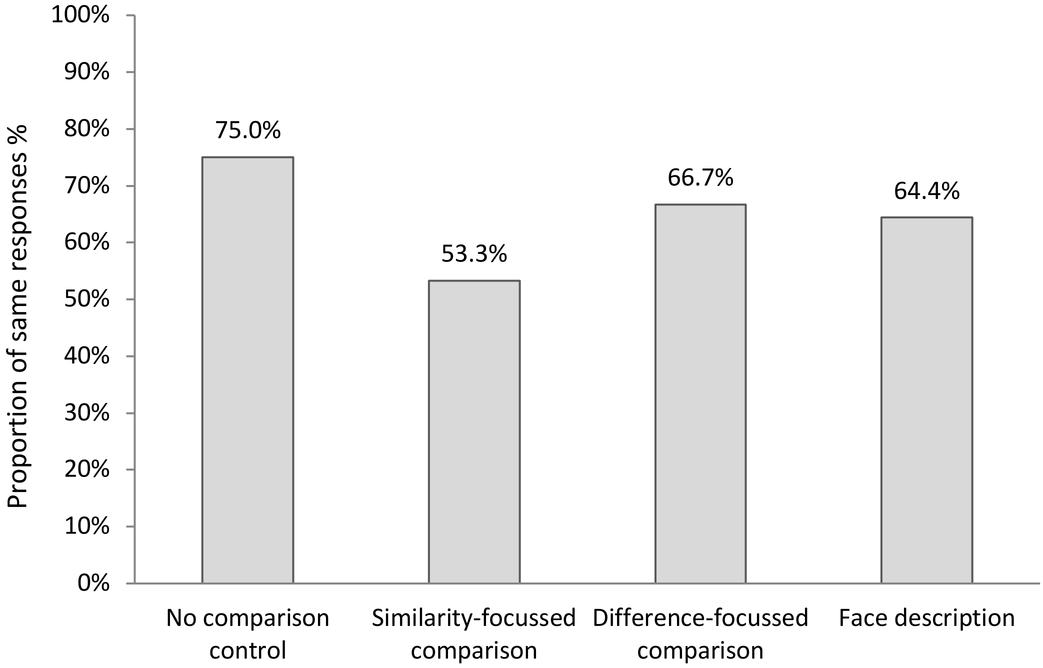


Figure 25: *The proportion of 'same' responses obtained in Experiment 7 for the different treatment conditions.*

5.2.3 Discussion

As can be seen in Figure 25, both similarity- and difference-focused comparisons resulted in a decrease in the proportion of 'same' responses for the unfamiliar face pairs. However, only the decrease in the number of 'same' responses following similarity-focused comparisons differed significantly from the control. The results for the similarity-focused comparisons replicate the findings of Experiment 7, and extend them to comparisons involving video footage. Together the two experiments demonstrate that, whilst comparison increases the perceived similarity of unfamiliar face pairs that are easy to discriminate, it results in a decrease in perceived similarity if the amount of face-specific information available for one of the stimuli is degraded. The implications of this finding for theories of similarity will be further discussed in the general discussion of this chapter.

As can be seen in Figure 25, there was a reduction in the number of 'same' responses following face descriptions compared to the no-comparison control. However, this finding was not significant. The absence of a significant finding mirrors the results of Experiment 2. Thus, whilst there is evidence to suggest that face descriptions can improve recognition accuracy (e.g Brown & Lloyd-Jones, 2006;

Bruce et al., 2001), it seems unlikely that verbal descriptions are driving the effect of comparison on perceived similarity reported here.

Overall performance accuracy was quite poor. Only 25% of participants in the control condition correctly judged that the two faces were not the same person. However, poor matching performance was also found in the study which first deployed these face pairs, under more optimal viewing conditions (Davis & Valentine, 2009). In particular, the study found a false alarm rate of 67% for one of the two target pairs used here, when it was either the first or the only pair that was matched (Davis & Valentine, 2009). Considering that the experimental set-up in Davis and Valentine's (2009) study was somewhat more favourable to correct identifications—participants compared video footage to a person present in the room and could therefore examine the faces from multiple angles—these findings can be treated as compatible.

The generalisability of the findings reported in this experiment is potentially limited by the age range of the participants. Some of the participants were 17-year-old college students. There is evidence to suggest that face recognition abilities are not fully mature at this age (Susilo, Germine, & Duchaine, 2013). However, a study that assessed the developmental trajectory of face identification ability found no significant difference in performance accuracy on a same–different identity perception task between older adolescents (15.90–18.00 years) and adults (18.01–33.15 years) (Fuhrmann et al., 2016). This finding suggests that age is unlikely to have been a limiting factor on performance for the youngest participants. A related issue is that people are generally better at recognising individuals who are of a similar age to themselves (Rhodes & Anastasi, 2012). However, this effect appears to be most pronounced in younger adults (Wiese, Komes, & Schweinberger, 2013). Considering that the male faces used as stimuli were in their 20s, and therefore of a similar age to the younger participants, and that the own-age bias has been demonstrated to affect memory, not perception, the own-age bias is also unlikely to have impacted on the results.

5.3 Chapter 5 discussion

In Experiments 6 and 7, participants were asked to decide whether two unfamiliar face stimuli were the same, under unlimited viewing conditions. In both experiments the quality of one of the stimuli was degraded. In Experiment 6, one of the two unfamiliar faces was pixelated; in Experiment 7, video footage was used that was both of lower resolution than the photographs, and captured the face from a greater distance (the video showed the entire person). Both studies found a comparison-related decrease in perceived similarity.

Experiment 7 assessed whether this decrease in perceived similarity might be attributable to greater familiarity with the stimuli by asking participants to describe the individual faces in a pair. There was no significant comparison-related change in perceived similarity for the face description condition. This finding suggests that familiarity is unlikely to be the driver of the comparison-related decrease in perceived similarity found in Experiments 6 and 7. The difference-focussed comparison condition in Experiment 7 also failed to reach significance. A similar null result for the difference-focussed comparisons condition was observed in Experiments 3 and 4. A possible reasons why the comparison-related changes in perceived similarity are less stable following difference-focussed comparisons will be discussed in Chapter 6.

Experiment 6 tested whether the decrease in 'same' responses following comparison might be attributable to better discrimination learning, by comparing performance in a same faces condition, with performance in a similar faces condition. A comparison-related decrease in the number of same responses was found both for the similar and for the same faces condition. However, the comparison-related decreases in perceived similarity were only significant in the similar faces condition. Nevertheless, these findings suggest that improvements in accuracy are not the driver of the comparison-related changes in the number of 'same' responses. If this were the case then comparison would have resulted in a reduction in the number of 'same' responses in the same faces condition. This finding differs from the general finding of discrimination learning studies which suggest that the opportunity to compare stimuli that are difficult to discriminate, improves the ability to later discriminate between them. There are several reasons for the existence of this discrepancy.

First, changes in performance accuracy in discrimination learning studies is generally driven by an improvement in participants' ability to tell two different stimuli apart. Performance on same trials is generally high in all conditions because they involve stating that two identical images are the same. In other words, "because the stimuli are very difficult to discriminate, participants are much more likely to miss a difference that is present than they are to mistakenly perceive a difference that is not present (resulting in an overall bias towards responding 'same')"(Lavis & Mitchell, 2006). A comparison-related improvement in the number of correct 'different' responses discrimination learning study is compatible with the comparison-related decrease in the number of 'same' responses observed in Experiments 6 and 7.

A second important difference is that the stimuli used in discrimination learning studies have clear diagnostic features. If one looks at the face pairs or checkerboards for long enough, one will eventually find the important difference by which the stimuli can be discriminated. Be it a slight difference in a checkerboard pattern, or a slight difference in the shape of a nose. The face stimuli that are captured in CCTV footage do not necessarily have such clear cues to be discovered. The poverty of the stimulus is well illustrated by the finding that two images of the same person can look so different that a range of face matching algorithms incorrectly match them on 100% of trials (Phillips, Beveridge, et al., 2011). This begs the question whether a correct 'same' response by a human reflects any learning, or whether it simply results from a misinterpretation of other information. The face stimuli that are encountered in forensic settings have many diagnostic and non-diagnostic perceptual differences between them. In the absence of feedback (which is often provided on test trials in discrimination learning studies) the human observer needs to rely on his or her subjective interpretation of this ambiguous information.

It is important to note that this pattern does not exclude the existence of discrimination learning. If there are differences that allow for the discrimination of stimuli then they are likely to be uncovered during comparison. However, for naturally occurring stimuli such as pictures of faces, there may simply not be a clear discriminating feature between stimuli. A same–different decision, will be strongly affected by the relative weighing up of pictorial differences between stimuli. It is in these cases that the effect of comparison-related decreases in perceived similarity are the most likely to have an impact. This hypothesis could be tested by comparing the effect of comparison on same-different judgements between distinctive and average faces. The number of 'same' responses following comparison should be lower for distinctive mismatch pairs (because they have more distinguishing features) than for average mismatch pairs.

5.3.1 Theoretical implications

The findings of Experiments 6 and 7, in combination with those of Experiments 1–5 suggest that two opposing processes are at play when comparing faces. A structural alignment process which increases the perceived similarity of the compared faces, and a habituation process which decreases it. The effect of the alignment process is dominant for a wide range of stimuli, provided the differences between them are perceptually accessible. The habituation process is dominant when habituation to extraneous pictorial information (i.e. visual information that is extraneous to, and detracts from, the task at hand) is beneficial because the compared stimuli are difficult to discriminate.

5.3.2 Applied implications

These findings have clear implications for the forensic task of identifying a suspect in CCTV footage. They suggest that the comparison process might introduce a bias to say that the two stimuli are not the same person. It is important to note that this bias has been demonstrated in the absence of any other contextual information. There is evidence to suggest that the provision of false contextual information can lead fingerprint experts to reverse their decision (Dror & Charlton, 2006; Dror et al., 2006, 2005). These findings highlight the importance of providing individuals, who are required to perform a face matching task in a professional or legal context, with feedback training, where they are asked to compare faces for which the true identity is known. At best, they will learn how to interpret these stimuli more accurately. At a minimum they will learn about the variance that exists in these types of stimuli, that are introduced by pictorial differences.

5.3.3 Outstanding questions

The stimulus pairs used in Experiments 6 and 7 consisted of one degraded (e.g. pixelated face in Experiment 6), and one non-degraded (e.g. the high-quality photo in Experiment 7) stimulus. It is therefore not clear whether the changes in perceived similarity are attributable to the fact that it was difficult to get face-specific information from one of the face stimuli, or to the fact that the two stimuli were pictorially very different. Based on the habituation account it can be predicted that when both stimuli are degraded, comparison should have the same effect on the perceived similarity of the compared stimuli, then when only one stimulus is degraded. This question was not addressed in this chapter because a comparison that involves at least one relatively high quality stimulus seemed to have higher ecological validity. After all, a suspect in the dock, or an image of a suspect held on file, is likely to be of high quality. However, this question could be addressed in a future study in which participants are asked to compare two screenshots taken from CCTV footage. One interesting additional question that could be asked in this context is whether face-extraneous cues might influence participants' decisions. For example, if both CCTV clips are recorded in the same location, would it increase the likelihood that two faces are judged to be the same?

A second issue is that the effect of stimulus degradation was correlated with other pictorial differences. Specifically, the two faces in Experiment 6 were shown from different angles. The stimuli in Experiment 7 could be compared from a variety of different views, because one of the stimuli was presented in a video clip. It remains to be established whether comparison will still result in a decrease in perceived similarity if all other pictorial variables are held constant. This could be tested by using the same face view (i.e. both frontal) used in Experiments 6.

Another outstanding issue is related to the dependent variable used. Experiments 1–5 have used similarity ratings to measure perceived similarity, and have found comparison-related increases in perceived similarity. Experiments 6 and 7 have used same-different judgments, and have found comparison-related decreases in perceived similarity. The same–different judgement task was used to increase the ecological validity of the experiments reported in this chapter. Together the findings from Experiments 6 and 7 show that comparing two face stimuli that are difficult to tell apart (because at least one of the stimuli is of low quality) will increase the likelihood that they will be judged to be the same. The use of similarity ratings does not allow for such a strong conclusion. Nevertheless, the assumption that the changes in the proportion of 'same' responses will be reflection in changes in similarity ratings needs to be empirically tested. This issue will be further discussed in the general discussion.

5.3.4 Conclusion

Experiments 6 and 7 have shown that comparing unfamiliar face stimuli will reduce the perceived similarity between them, if at least one of the stimuli is of low quality which makes it difficult to abstract face-specific information. Together with the findings from Experiments 1–5, these findings suggest that habituation and structural alignment theory need to be considered together to account for comparison-related changes in perceived similarity. The findings obtained in Experiments 6 and 7 have implications for the context of forensic and judicial comparisons that involve CCTV footage. They suggest that the act of comparison might bias the decision towards deciding that two faces are not a match. This bias is particularly likely to play a role for ambiguous stimuli that have multiple diagnostic and non-diagnostic similarities and differences.

Chapter 6: General discussion

6.1 Summary of findings

This thesis had two aims. The first aim was to assess the effect of comparison on the perceived similarity of range of unfamiliar face stimuli. The second aim was to understand whether, and in what manner, the nature of the comparison task influences perceived similarity. To address these aims seven experiments were conducted, which are presented in the empirical chapters of this thesis.

6.1.1 Chapter 3

The first three experiments, presented in Chapter 3, explored the effects of comparison on the perceived similarity of faces that are relatively easy to tell apart. In Experiments 1 and 2, participants were asked to compare pairs of unfamiliar faces shown from two different angles: a frontal, and a ¾ view. In Experiment 3 participants compared pairs of faces that were rotated in the picture plane: one face in a pair was presented upright, the other upside down. A comparison-related increase in perceived similarity was found in all three experiments. Furthermore, the findings of Experiment 1—which asked participants to compare phenotypically similar and dissimilar face pairs—suggest that the increase in perceived similarity is not sensitive to variations in the phenotypic similarity of unfamiliar faces. Together these findings show that comparison increases the perceived similarity of unfamiliar face stimuli that are relatively easy to tell apart, both because there are clear phenotypic differences between them, and because the pictorial differences still allow for the abstraction of facial information.

The comparison-related increases in perceived similarity found in Experiments 1–3 were largely unaffected by variations in the comparison task. Face pairs that were subjected to either a similarity- or a difference-focussed comparison task, were given higher similarity ratings than face pairs that were not compared before the similarity rating task. There was one exception to this pattern: difference-focussed comparisons did not result in a significant increase in the perceived similarity for the pairs of upright and inverted faces. This finding suggests that difference-focussed comparisons might be more vulnerable to variations in task difficulty than similarity-focussed comparisons. Evidence that the higher similarity ratings in the comparison conditions are attributable to comparison—rather than increased familiarity with, or better encoding of, the compared stimuli—is provided by the finding that performance of a face-description task prior to a similarity rating, did not result in higher similarity ratings, relative to a control.

6.1.2 Chapter 4

Experiments 4 and 5, which are presented in Chapter 4, were designed to test the boundary conditions of the results obtained in the first three experiments. The first boundary condition was related to the stimuli. The unfamiliar face stimuli used in Experiments 1–3 were both phenotypically and pictorially dissimilar, it is therefore not possible to pinpoint the locus of the comparison-related increase in perceived similarity. The relative contribution of phenotypic resemblance was isolated in Experiment 4, by asking participants to compare face morphs shown from the same angle. Half the participants compared face morphs that resembled each other quite closely phenotypically, the other half compared face morphs that were less similar. The study found comparison-related increases in perceived similarity for the less similar face morphs, but not for the face morphs that resembled each other quite closely. Experiment 4 also assessed whether explicit comparisons are necessary for comparison-related increases in perceived similarity to emerge. It was found that whereas performing an explicit similarity- or difference-listing task is not a necessary requirement for comparison-related increases in perceived similarity to emerge, performing a comparison is.

The second boundary was related to the comparison task. It is unclear from the findings of Experiments 1–4 whether the comparison-related increases in perceived similarity are limited to situations where there are two discrete instances of comparison (i.e. the comparison task, and the similarity assessment task), or whether they would also arise within a continuous comparison sequence. The time-course of comparison-related changes in perceived similarity over a single comparison sequence were assessed in Experiment 5, by measuring the effect of varying the presentation duration of face pairs on perceived similarity. A different pattern of results emerged for the similar and the dissimilar face pairs. For the dissimilar face pairs—which consisted of frontal photos of unfamiliar faces that were matched to be phenotypically dissimilar—comparison resulted in a significant increase in perceived similarity between the short (1 s) and the intermediate (3 s) exposure duration. A slight increase was also seen between the intermediate (3 s) and the long (9 s) exposure duration, however, this increase was not found to be significant. For the similar face pairs—which consisted of frontal images of highly similar face morphs with only 10% variance between them—similarity ratings were higher at the intermediate exposure duration, relative to the short exposure duration. This pattern of results is the same than that obtained for the dissimilar pairs. However, similarity ratings at the long-exposure duration condition were found to be significantly lower than at the intermediate exposure duration.

6.1.3 Chapter 5

The findings of Experiments 4 and 5, together with the findings from discrimination learning studies that have used unfamiliar face stimuli (Dwyer et al., 2011, 2009, Mundy et al., 2014, 2007; Mundy, Honey, Downing, et al., 2009), suggest that comparison-related decreases in perceived similarity might be limited to face stimuli that are highly similar phenotypically, and pictorially homogenous. The hypothesis that comparison-related decreases in perceived similarity might also be found for stimuli that are pictorially dissimilar, was tested in Experiments 6 and 7. In Experiment 6, participants were asked to compare two unfamiliar face stimuli that differed in image quality. This manipulation was designed to assess whether pictorial differences that make it more difficult to abstract information from one of the images, would affect perceived similarity. There were two stimulus similarity conditions. Face pairs in the 'same' condition consisted of a degraded and a non-degraded image of the same individual. Face pairs in the 'similar' condition were different individuals that were matched to resemble each other phenotypically. The study found a significant decrease in the number of 'same' responses following both similarity- and difference-focussed comparisons for the similar face pairs. A comparison-related decrease in perceived similarity was also found for the same pairs, but it was not statistically significant.

Experiment 7 was a partial replication of Experiment 6, with a different stimulus set. Participants were asked to decide whether an unfamiliar face shown in a photograph matched a person presented in video footage. The quality of the video recording resulted in a reduction in the amount of detailed, face-specific information that could be abstracted. The experiment found a comparison-related decrease in the number of 'same' responses, following similarity-focussed comparisons. Together the findings of Experiments 6 and 7 suggest that comparison-related decreases in perceived similarity are not restricted to stimuli that are pictorially homogenous. Instead they appear to be limited to unfamiliar face stimuli for which differences are difficult to assess, be it because these differences are subtle and hidden in a sea of sameness, or because they are ambiguous and surrounded by visual noise.

6.1.4 Summary and overview of discussion

This thesis set out to explore two related aims. The first aim was to understand the effect of comparison on the perceived similarity of a range of unfamiliar face stimuli. The second aim was to understand the role that the comparison task plays in mediating the effect of comparison on the perceived similarity of unfamiliar faces. To meet these aims seven experiments were conducted that assessed the effects of varying both stimulus attributes, and the comparison task, on perceived similarity. The findings were discussed in relation to applied contexts in which face matching performance is critical, as well as in relation to theoretical accounts of comparison. The three main findings from this research are:

1. Comparing unfamiliar face stimuli can result in both increases, and decreases in the perceived similarity of the compared stimuli.
2. Whereas comparative processes give rise to changes in the perceived similarity of unfamiliar face stimuli, they do not drive the direction of the effect.
3. The direction of the comparison-related changes in perceived similarity found in this research is driven by the phenotypic and pictorial similarity of the unfamiliar face stimuli.

The implications that emerge from the research presented in this thesis are threefold:

1. The research can be used as a framework for a systematic approach to the study of naturalistic face stimuli that considers both pictorial and phenotypic differences.
2. The research can be used to constrain, and refine theoretical accounts about the mechanisms of comparison.
3. The research informs our understanding about face matching performance in applied settings.

Before turning to the discussion of these implications, it is important to first discuss the two main assumptions that have been made in this thesis, the potential limitation they impose on the generalisability of this research, and how these limitations could be addressed.

6.2 Caveats and next steps

This thesis assumed that the likelihood that two unfamiliar faces will be judged to be the same person is largely determined by the perceived similarity of the stimulus pair. This assumption has informed the decision to treat both similarity ratings and same–different judgments as measures of perceived similarity. Similarity ratings constitute a direct measure of similarity, whereas same–different judgments constitute an indirect measure. It has been argued in section 1.2.3 that direct and indirect measures of perceived similarity reflect the same underlying similarity construct. Face space constitutes a model of this similarity construct. The face space framework posits that faces are represented as points in a multidimensional similarity space (Valentine, 1991; Valentine et al., 2016). Faces with a high degree of phenotypic similarity are represented close together. Faces that are dissimilar are represented far apart. However, whilst the existence of a strong correlation between direct (e.g. similarity ratings) and indirect (e.g. responses on a same–different judgement task) measures of similarity, indicates that they measure a common underlying similarity construct, the strength of the correlation varies across studies (Medin et al., 1993; Palmer, 1978; Tversky & Gati, 1982). Some of this variation is likely to be attributable to experimental manipulations that tend to co-occur with a given dependent variable. For example, studies that assess response accuracy are often timed. The contribution of these factors was controlled for in this thesis by using the same design in all experiments (except for Experiment 5). For example, both the same–different judgements in Experiments 6 and 7, and the similarity ratings in Experiments 1–4, were untimed.

Nevertheless, it is an important next step in this body of work, to test this underlying assumption empirically, by replicating Experiment 6 and–or 7, using a similarity rating task to measure perceived similarity. If a comparison-related decrease in perceived similarity is found, then it would provide converging evidence for the argument made in this thesis. Alternatively, finding that under unconstrained viewing conditions, similarity ratings and same–different judgments nevertheless diverge, would provide an interesting avenue for the improvement of forensic best practices. It would suggest that performing both same–different judgements and similarity ratings might result in less bias. To test the underlying assumption fully, it is equally important to assess whether comparison of the unfamiliar face stimuli used in Experiments 1–5 will increase the number of 'same' responses on a same–different judgement task. For some of these stimuli (e.g. the dissimilar faces used in Experiment 5) this could lead to ceiling effects under unconstrained conditions. However, this issue could be addressed by limiting the amount of time available to perform the comparison, and for the same–different decision. The prioritisation of which subset of the stimuli used in Experiments 1–5 to test, could be based on the stimuli that are of most relevance to the applied context of border control.

Another assumption that has been made in this thesis is that both similarity ratings and same–different responses reflect perceived similarity. However, it is not possible to determine whether the changes in perceived similarity reported in Experiments 1–7, are attributable to a change in how the compared stimuli are perceived, or to a change in response criterion. Signal detection theory breaks down responses on same–different tasks into two components: a decision criterion and a sensitivity score. The sensitivity score reflects changes in the perceiver's ability to abstract information from the visual stimulus. In the case of comparing highly similar face stimuli a change in sensitivity would allow for the detection of subtle differences. The decision criterion reflects a response bias. A liberal response criterion requires relatively little evidence (signal) to make a response. A conservative response criterion requires a lot of evidence (signal) to make a response. In the context of a face matching task, a liberal response bias would lead to erroneous 'same' responses on target absent trials, whereas a conservative response bias would lead to misses on target present trials. It is unclear whether the comparison-related changes in perceived similarity that were found in Experiments 1–7, are attributable to a shift in response criterion, a change in sensitivity, or a combination of both.

Understanding whether the comparison-related changes are driven by shifts in response criterion or changes in sensitivity, has practical implications, because changes in response criterion are thought to be more amenable to behavioural intervention (Stanislaw & Todorov, 1999; but see Aberg & Herzog, 2012). The data that were generated in Experiments 1–7 is not conducive to an analysis based on signal detection theory, because each participant performed only one trial per condition. This issue could be addressed in future studies by requiring participants to complete more trials. However, there is a risk inherent in requiring participants to perform multiple trials, which justifies the approach taken here, at least as an initial step to assess the presence of an effect. The process of task repetition can obscure the effects that are of interest. This has been shown to be the case for a research question closely related to the ones under investigation in this thesis. Specifically, it has been shown that repetition of the comparison task has a general effect of increasing the number of 'same' responses on face matching tasks, which can fully or partially mask other effects, including the effect of time-pressure on performance accuracy (Bindemann et al., 2016; Fysh & Bindemann, 2017).

6.3 A framework for studying the effects of comparison on the perceived similarity of naturalistic face stimuli

6.3.1 The specifics of the comparison task

One of the aims of this thesis was to understand how the nature of the comparison task influences the perceived similarity of unfamiliar face stimuli. It was predicted that a variety of different comparison tasks would result in a change in perceived similarity (Hypothesis 3). This hypothesis was supported by the findings. Comparison-related changes in perceived similarity were found for a similarity-listing task (Experiments 1–4, 6, and 7), a difference-listing task (Experiments 1, 2, and 6), an implicit comparison task (Experiment 4), and when participants were instructed to compare faces presented to them on a screen (Experiment 5). Whilst comparison-related changes were observed in all experiments for similarity-focused comparisons, the findings for difference-focused comparisons were less robust. No clear causal factor could be identified based on an examination of potential differences between the stimulus sets, or any other variables. It is possible that the focus of the comparison task has a subtle effect on the relative weighting given to the common and distinctive features of the compared stimuli (Markman, 1996; Medin, Goldstone, & Gentner, 1990; Tversky & Gati, 1978), which affected the impact of difference-focussed comparisons on perceived similarity. No change in perceived similarity was found when participants were asked to describe each face in a pair individually (Experiments 2 and 7), or when they answered a simple two-alternative choice question for each face (Experiment 4). Together the findings for the comparison and control conditions suggest that the effects of comparison on perceived similarity of unfamiliar faces are both unique to comparison, and robust to variations in task instructions.

Previous research suggests that similarity-focused and difference-focused comparisons might have opposing effects on perceived similarity (Hassin, 2001). This hypothesis was not supported by the findings reported here. In each experiment the effect of all comparison manipulations, including similarity- and difference-focused comparisons, was always found to be in the same direction. For example, in Experiment 4 both similarity-focussed, and implicit comparisons were found to result in an increase in perceived similarity in the dissimilar faces condition. Similarly, in Experiment 7 both similarity-, and difference-focussed comparisons resulted in a decrease in the number of ‘same’ responses in the similar faces condition. This pattern of results suggests that the nature of the comparison task does not drive the direction of the effect of comparison on perceived similarity.

It was predicted that the effect of comparison on the perceived similarity of unfamiliar face stimuli would change as a function of comparison duration (Hypothesis 4). This hypothesis was supported by the findings of Experiment 5. Furthermore, the findings of Experiment 5 suggest that the effect of comparison duration on perceived similarity is influenced by the phenotypic similarity of the face pairs. However, this interpretation is tentative due to various shortcomings of the experimental design, including the fact that phenotypic similarity was conflated with stimulus artificiality, and that only a limited range of comparison durations were tested. The next section examines the role of phenotypic similarity, and pictorial differences in mediating the effect of comparison on perceived similarity.

6.3.2 The role of phenotypic similarity and pictorial differences

Another aim of this thesis was to assess the effect of comparison on the perceived similarity of a range of unfamiliar face stimuli. To address this aim, the experiments presented in the empirical chapters of this thesis have examined the effect of varying both the phenotypic similarity of the compared faces, and the pictorial differences between them. The term phenotypic similarity was introduced to refer to the natural variations that occur in faces due to genes and the impact of the environment. It was predicted that comparison would change the perceived similarity of phenotypically similar and dissimilar face pairs (Hypothesis 1). The term pictorial difference refers to all other differences that exist between face stimuli, including those introduced by system variables (e.g. the camera used), environmental variables (e.g. ambient lighting), person variables (e.g. whether the subject was wearing a hat), and the interactions between them (e.g. distance between face and camera). It was predicted that comparison would alter the perceived similarity of face stimuli across a range of pictorial differences (Hypothesis 2). The unfamiliar face stimuli used in Experiments 1–7 varied both in the degree to which they phenotypically resemble each other, and pictorially. For example, consider the face pair labelled ‘Exp 6 same’ in the top right-hand corner of Figure 8. The pictorial differences between the two faces consist of a difference in pose, and the fact that the face on the right-hand-side is pixelated. The two faces are phenotypically identical because they are the same person, and the two photographs were taken on the same day.

In Experiments 1–4 comparison resulted in an increase in perceived similarity. As can be seen in Figure 26, the increase in perceived similarity was relatively impervious to variations in phenotypic similarity. For example, a comparison-related increase was found for both the phenotypically similar and dissimilar face pairs in Experiment 1. In Experiments 6 and 7 comparison resulted in a decrease in perceived similarity. As can be seen in Figure 27, the face pairs used in Experiments 6 and 7 were phenotypically similar. The similar face pairs used in Experiment 6 were the same faces than the similar faces for which a comparison-related increase was found in Experiment 1. This raises the question—why can comparison both increase and decrease the perceived similarity of a given face pair? One obvious difference between the two stimulus sets is the pictorial pixelation manipulation deployed in Experiment 6. Similarly, as can be seen in Figure 27, there were multiple pictorial differences between the paired face stimuli used in Experiment 7. In Experiments 1–4 on the other hand, there were either no pictorial differences between the two faces in a pair (Experiment 4), or a single pictorial difference (Experiments 1–3).

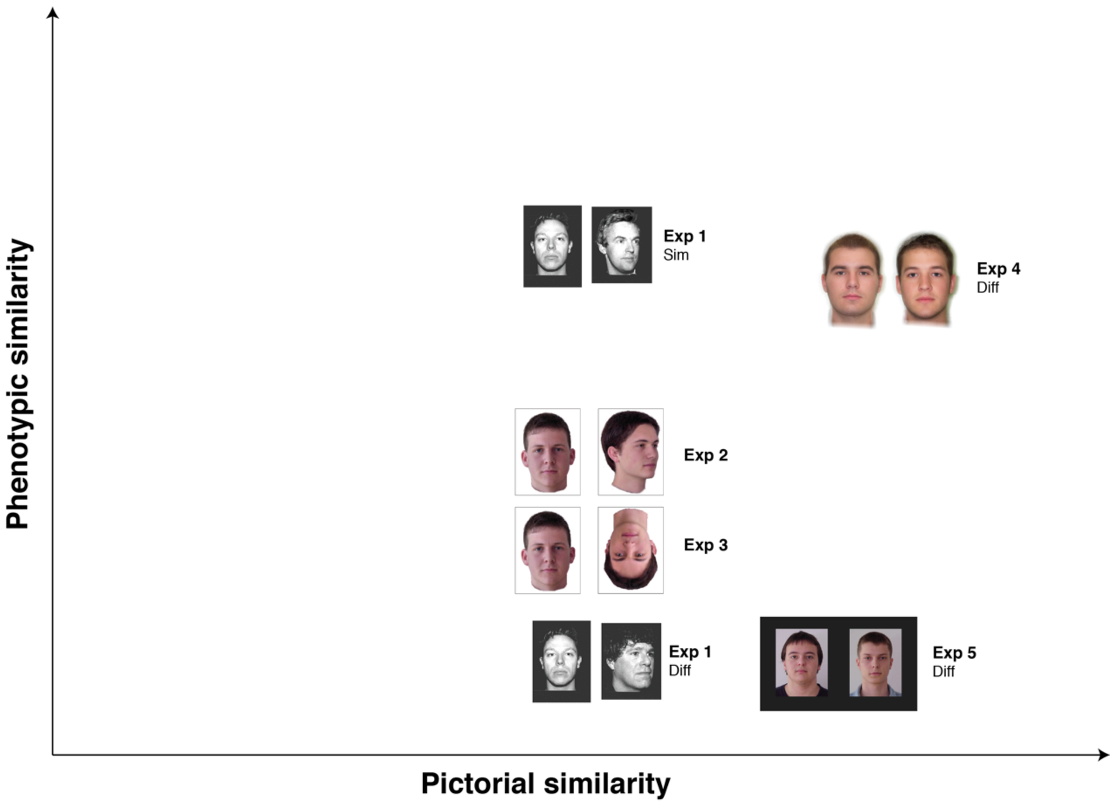


Figure 26: *The stimuli for which comparison-related increases in similarity have been found. The comparison manipulation used in experiment 5, was different from that used in the other experiments. The different face pair from experiment 5 is nevertheless included here because the overall trend was an increase in perceived similarity.*

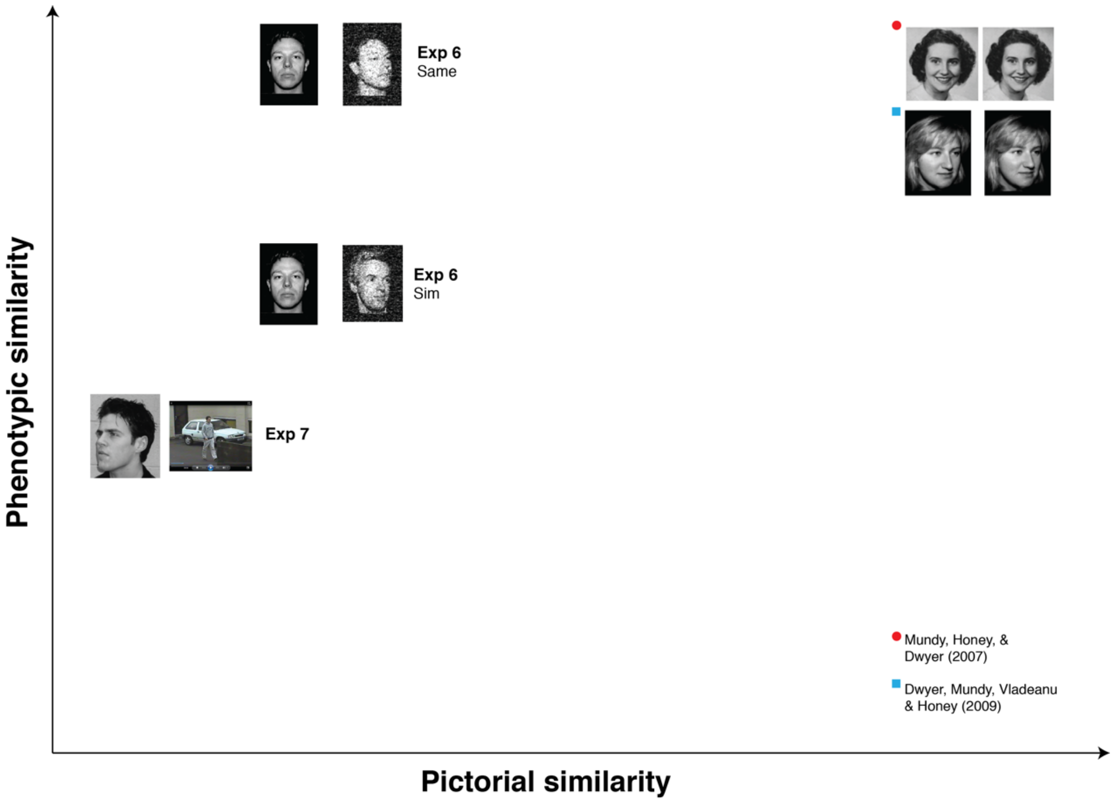


Figure 27: *Examples of unfamiliar face stimuli for which comparison-related decreases in similarity have been found. These face pairs are the stimuli used in Experiments 6 and 7, as well as examples taken from the discrimination learning literature. In Experiments 6 and 7 perceived similarity was operationalised as the total number of 'same' responses. In studies on discrimination learning it is operationalised as better performance accuracy.*

Comparison of the stimuli used in Experiments 1–4, and those used in Experiments 6 and 7 suggests that pictorial differences play a crucial role in driving the direction of the effect of comparison on perceived similarity. However, the role of pictorial differences becomes less clear if one considers the stimuli that were used in studies on discrimination learning. As can be seen in Figure 27 discrimination learning studies have found that comparison decreases the perceived similarity of homogenised face pairs that are pictorially identical, and difficult to tell apart phenotypically (Dwyer et al., 2009; Mundy et al., 2007). The face pairs that were used in discrimination learning studies were designed to be as similar as identical twins. The stimuli used in Experiment 7 were chosen because they resulted in many misidentification errors. In Experiment 6 one of the faces in each pair was pictorially degraded (through pixelation). The pictorial manipulation was used to increase the difficulty of the same–different judgment task. The finding that comparison can decreases the perceived similarity of face pairs that have many pictorial differences, as well as of face pairs that have none, can be explained by the fact that that the faces used in Experiments 6 and 7, and the faces deployed in discrimination learning studies are difficult to individuate.

It was argued in Chapter 2 that it is not possible to predict, based on the existing literature on comparison, at what degree of phenotypic stimulus similarity the effect of comparison on perceived similarity is likely to change direction. Together the findings of Experiments 1–7 suggest that comparison will only result in a decrease in perceived similarity for stimuli that are difficult to individuate. The only unfamiliar face pairs that most people find very difficult to individuate due to their phenotypic resemblance are twins. It follows that in the absence of pictorial differences comparison is likely to result in an increase in perceived similarity for most unfamiliar face stimuli. However, the prediction about whether the comparison of two faces will result in an increase or a decrease in perceived similarity can only ever be probabilistic, due to individual differences in the face processing abilities of the observer. Two faces might appear as doppelgängers to someone who suffers from prosopagnosia. For that person, comparison is likely to result in a decrease in the perceived similarity of the face pair. A super-recogniser, on the other hand, would find it trivially easy to tell the two faces apart. It follows that for the super-recogniser the perceived similarity of the face pairs is likely to increase following comparison.

The above argument assumes that the effects of comparison on perceived similarity operate in the same way across the entire spectrum of face-processing abilities. However, the relationship between individual differences in face-processing abilities and the effects of comparison on the perceived similarity of unfamiliar faces was not examined in this thesis. It is possible that comparison has a different effect on perceived similarity for individuals who are at the extreme ends of the continuum of face processing skills. There is evidence to suggest that the superior face-processing skills of super-recognisers are attributable, at least in part, to their holistic processing abilities. Yet the comparison-related increase in perceived similarity for pairs of upright and inverted faces found in Experiment 3 show that the effects of comparison on perceived similarity observed in Experiments 1–7 are unlikely to be dependent on holistic processing. It is plausible that comparison may exert a different effect on perceived similarity for super-recognisers, because they are more likely than typical perceivers to deploy holistic processing when performing a face comparison.

A full understanding of the effects of comparison on the perceived similarity of unfamiliar faces could lead to the development of new diagnostic tests to identify super-recognisers. Some of the tests that are used to measure face-matching accuracy are not difficult enough to make fine distinctions at the higher levels of face-matching abilities. Similarity ratings are more nuanced, and could therefore potentially be used to identify super-recognisers, provided the pattern of responses that is likely to distinguish super-recognisers from the general population is known. A useful starting point would be to conduct a study in which typical perceivers, known super-recognisers, and people with prosopagnosia are asked to compare a set of faces that showcase a range of phenotypic variation, and to assess if there are any between-group differences in the direction, and the magnitude of the similarity ratings.

In sum, the findings obtained in Experiments 1–7, together with the findings obtained in discrimination learning studies suggest that the pictorial differences between face stimuli, and the phenotypic similarity of the depicted faces both play crucial roles in driving the direction of the effect of comparison on perceived similarity. One limitation of Experiments 1–7 is that the stimuli used only reflect a small subset of the variation that is found in images of faces. The stimuli depict Caucasian male faces, with an approximate age range of 18 to 30 years. It therefore remains to be established whether comparison increases the perceived similarity of face pairs that straddle category boundaries, for example, a pair in which one face is female, and the other male. Similarly, the pictorial differences that were assessed where largely limited to variations in pose. Experiment 7 displayed a wider range of pictorial differences that are found between ambient stimuli. However, because there were so many differences it is impossible to determine the relative contribution of each difference.

Arguably the role of pictorial differences is more important in applied settings. Firstly, whereas the faces used in studies on discrimination learning were designed to match each other so closely that they resemble identical twins, the majority of unfamiliar faces that are confused in forensic settings are not identical twins. Secondly, whereas the stimuli used in studies on discrimination learning are pictorially identical, this is unlikely to be the case for stimuli that are encountered in applied settings that require face matching. The research reported in the empirical chapters of this thesis has shown that pictorial differences play an important role in determining the perceived similarity of faces. Yet, whilst phenotypic similarity was manipulated as a between-subjects variable, the effect of varying pictorial differences between stimuli was not systematically examined. The next section outlines s a framework that can be used to model pictorial differences.

6.3.4 Augmenting the face-space model

Faces range widely in their degree of phenotypic similarity (Sheehan & Nachman, 2014). The face-space model has been developed to represent human perception of this phenotypic variation within a multi-dimensional similarity space (Valentine et al., 2016). Faces that are located close to each other on various dimensions in this space are perceived to be similar; faces whose attributes are separated by a larger distance are perceived to be less similar. The face-space model has been informed by, and contributed towards insights in a variety of research areas that involve face processing, including the own-race bias, facial caricature recognition, and facial adaptation (for a review see Valentine et al., 2016).

Whilst the face-space model has proven very useful to the study of cognitive processes that involve faces, it does not currently capture pictorial differences. This thesis proposes a complementary model to simulate the effects of pictorial variables. A possible instantiation of this model is shown in Figure 28. The model captures three variables that jointly give rise to pictorial differences, namely person variables, system variables, and environment variables. As can be seen, any of these variables in isolation can affect the pictorial qualities of an image. For example, the face of the person in the 'occlusion' example is very difficult to identify, because it is covered both with clothing, and by the subject's hands. In most ambient images, multiple pictorial variables affect the overall appearance of a face. For example, the 'backlit face' in Figure 28 is difficult to see clearly due to an interaction between where the person is standing in relation to the sun, and the ambient light conditions.

The pictorial model could be used to systematically test the effect of varying different pictorial variables on perceived similarity, and face matching accuracy. For example, given an equal amount of distortion it is possible that people are more likely to believe their own assessment when a face is backlit than when the image quality is reduced. Simply because people have experience of identifying others in different light conditions but less experience in using CCTV footage to do so. Finally, it would also be possible to combine this model with faces that are artificially generated from a face-space model. One hypothesis that could be tested by combining the two is that faces which are more distinctive, are easier to identify at a given level of degradation, than those that are not. The combinatorial possibilities are endless. Priority should be given to those combinations that are of either applied importance (e.g. distortion introduced when images are enlarged), or theoretical interest (e.g. to test the own-race bias on face perception). A systematic understanding of how pictorial variables interact would make it possible to make probabilistic predictions about human face-matching accuracy.

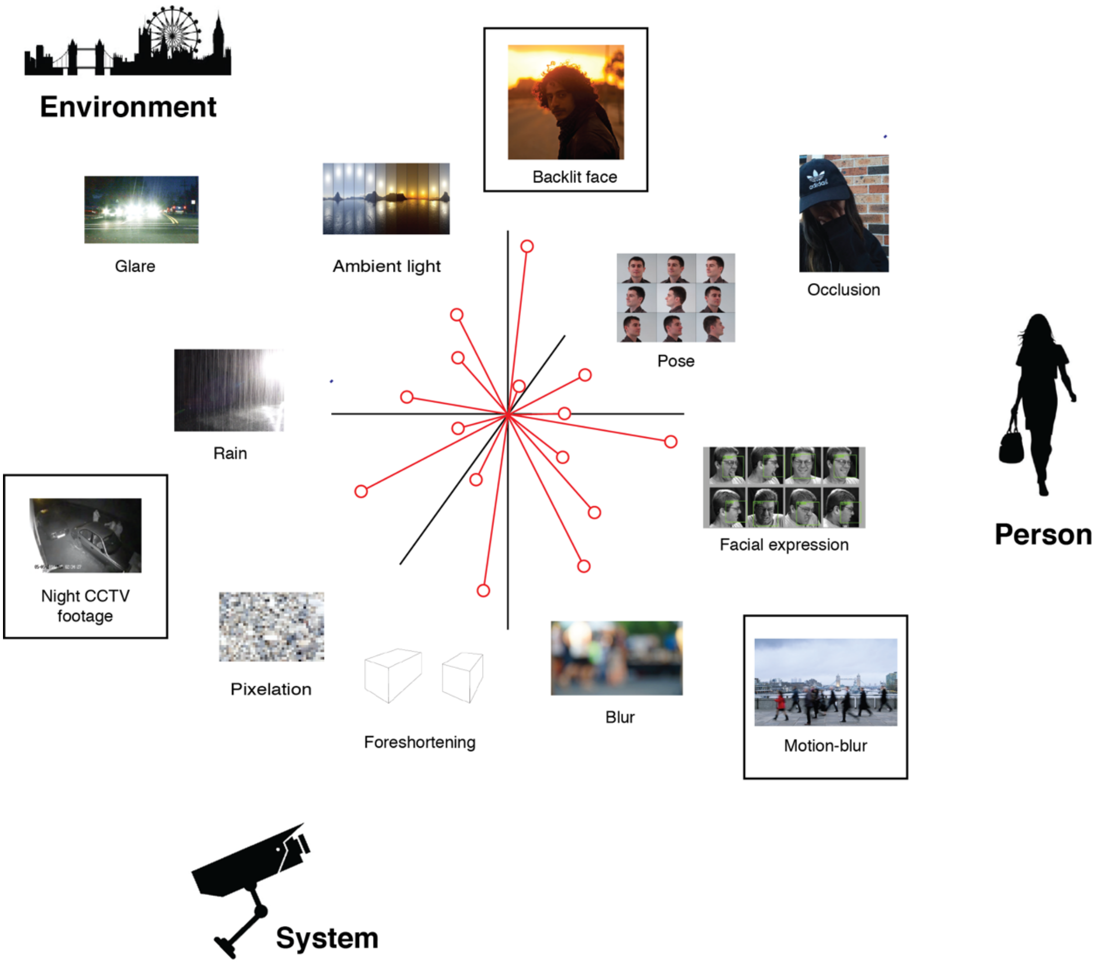


Figure 28: *Example of pictorial similarity space. The variables in black boxes reflect interactions between two variables. For example, motion blur arises through the interaction between person and system.*

6.3.5 Summary

This thesis has shown that the direction of the effect of comparison on the perceived similarity of unfamiliar face stimuli is driven by pictorial and phenotypic variations between the compared stimuli. Two predictions about the effect of comparison on perceived similarity have emerged from the research presented in the empirical chapters of this thesis. The first prediction is that for most naturally occurring face stimuli, comparison will result in an increase in perceived similarity. The second prediction is that comparison will decrease the perceived similarity of unfamiliar face pairs that are difficult to discriminate, either because they are phenotypically highly similar, or because the pictorial differences between them reduce the quality or the quantity of face-specific information that is available for comparison. The next section discusses the causal mechanisms that give rise to these differences.

6.4 Why does comparison have opposing effects on the perceived similarity of unfamiliar face stimuli, depending on their discriminability?

6.4.1 Theoretical accounts of comparison

Two theoretical accounts of comparison—which make opposing predictions about the effect of comparison on the perceived similarity of unfamiliar face stimuli—were used to understand the interactions between cognitive processes and stimulus attributes that give rise to changes in the perceived similarity of unfamiliar face stimuli. The first theoretical account is structural alignment theory. According to structural alignment theory, comparison increases the perceived similarity of alignable stimuli, such as faces, because the structural alignment process facilitates the discovery of commonalities, and enhances their salience (Markman & Gentner, 1993b, 1993c). The theory predicts that comparison should increase the perceived similarity of unfamiliar faces, because the commonalities that are discovered during the alignment process are weighted more highly than the differences (Markman & Gentner, 1993b).

The second theory is the habituation account of comparison, which predicts that comparison should decrease the perceived similarity of unfamiliar face stimuli that are pictorially homogenous, and phenotypically difficult to tell apart. According to this account, observers habituate to commonalities between stimuli over time, which enables them to direct their attention towards the differences by which the stimuli can be discriminated. In other words, “the operation of short-term adaptation/habituation effects has enduring repercussions for the attentional weighting given to the unique and common features” (Mundy et al., 2007, p. 136). In the introductory chapter to this thesis it was argued that neither theory can fully explain the range of comparison-related changes in perceived similarity that have been observed for visual stimuli. This argument held for the findings reported in the empirical chapters of this thesis.

Structural alignment theory is compatible with the findings of Experiments 1–3: that comparison results in an increase in perceived similarity of unfamiliar face stimuli that are relatively easy to tell apart. Similarly, the comparison-related increase seen for the dissimilar faces in Experiment 4 is also compatible with the theory. With a modification to Mundy’s definition, the habituation account of comparison was used to predict the findings of Experiments 6 and 7. According to the modified view, participants habituate, not only to shared features between homogenous stimuli—but also to the visual noise inherent in a low-quality image—which directs their attention towards the differences between the faces.

Overall, the findings of experiments 1–7 are best interpreted if one considers that there are two processes involved in generating the comparison-related changes in perceived similarity, which operate at the same time, and in opposition to each other. According to this view, structural alignment is dominant, because it is an automatic process that guides attention (Markman & Gentner, 2005). Habituation, on the other hand, can only operate on visual stimuli once they are attended to.

A combined account can explain the findings of Experiment 5. The short and the medium presentation durations were selected based on research on the time-course of structural alignment. The higher similarity ratings seen in the 3 s duration condition, relative to the 1 s duration condition conform to the predictions of structural alignment theory. However, the finding that there is a subsequent decline for the similar but not the dissimilar face pairs, suggests that the two processes operate simultaneously. A combined account could also explain the null result for the similar face pairs in Experiment 4. It could be explained by the suggestion that for phenotypically similar stimuli shown from the same angle, the two processes sometimes cancel each other out. This assumption could be tested by assessing the time-course of comparison for the similar face pairs used in Experiment 4.

6.4.2 Can theoretical accounts of comparison further our understanding of face matching in applied settings?

As has been discussed in detail in Chapter 2 professionals who perform face-matching tasks in applied settings are subject to a variety of influences, including contextual information, time pressure, monotony, distractions, and the need to perform multiple tasks at once. Contextual information about face pairs has been shown to exert a strong influence on the perceived similarity of faces. Participants rated adult-child pairs as more similar when they believed that the pairs depicted a parent and child, than when they believed that the two individuals were unrelated (Bressan & Dal Martello, 2002). Similarly, compared to participants who were not given any additional information, participants who were told that an eyewitness had made a positive identification assigned higher similarity ratings to face pairs depicting the suspect and the culprit (Charman et al., 2009). It is unlikely that structural alignment theory and habituation—which are both passive processes that happen automatically—can fully capture the cognitive processes that are involved in generating these biases. Yet, both theoretical accounts can contribute towards an explanation of how the effects of biases and other task constrains might exert their effect on face-matching performance.

The increases in perceived similarity that have been observed in response to the provision of false information about face pairs has been ascribed to the motivation of the observer to "seek confirming evidence, by looking for common traits rather than for differences" (Bressan & Dal Martello, 2002, p. 217). However, the findings reported in the empirical chapters of this thesis suggest that task focus (i.e. looking for common traits or differences) alone is unlikely to change the outcome of a comparison. For example, in Experiments 1 and 2, comparison was found to result in an increase in perceived similarity regardless of whether participants were instructed to look for commonalities or differences. This pattern of results was consistent with structural alignment theory, which posits that comparison should increase the perceived similarity of unfamiliar face pairs because the similarities that are revealed during alignment, are weighted more highly than the differences (Boroditsky, 2007;

Markman & Gentner, 1993b). According to structural alignment theory both similarity- and difference-focused comparisons should result in an increase in perceived similarity, because the alignment process is not dependent on the focus of the comparison. Based on this analysis it seems more plausible that the effect of contextual information on perceived similarity is attributable to a change in the response criterion. According to this argument face pairs that are believed to be unrelated need to reach a higher level of similarity before they are rated as similar than face pairs that are believed to be related.

Another factor that is likely to affect face matching performance in applied settings is the need to perform multiple tasks concurrently. A study that assessed whether embedding a passport style photograph in an ID document affects face matching accuracy found that participants were more likely to make ‘same’ responses in the passport frame condition, relative to a no-frame control condition (McCaffery & Burton, 2016). The increase in the number of same responses might be caused by a disruption of habituation. As outlined above, when comparing homogenous stimuli observers habituate to the commonalities between them, which enables them to direct their attention towards the differences. This habituation process may be disrupted if observers have to look away frequently to focus on other pieces of information. This interpretation, as well as the explanation offered for the effect of contextual information on the perceived similarity of unfamiliar face pairs is tentative and requires further study. However, the interpretations illustrate that theoretical accounts of comparison, combined with empirical data can be used to interpret the findings of applied research.

6.5 Future work in applied settings

The research presented in this thesis was partially motivated by, and grounded in, two applied contexts that involve the comparison of unfamiliar faces. The first context is matching travellers to passport photos for border control. The second context is comparing a photo of a suspect, to CCTV footage of the perpetrator.

6.5.1 Implications for passport identifications

Research findings, which show that unfamiliar face matching is error prone even when only two faces need to be matched, and high quality face stimuli are used

(Henderson et al., 2001; Megreya & Bindemann, 2009; Megreya & Burton, 2008; Megreya et al., 2013), suggest that the task of matching travellers to passport photos might be more error prone than commonly believed. This argument is strengthened by the finding that face matching accuracy is significantly worse for image pairs that were captured a few months apart, than for images that were captured on the same day (Megreya et al., 2013; see also Bindemann & Sandford, 2011).

The findings of Experiments 1–4 suggest that comparison will result in an increase in the perceived similarity between a given traveller and their passport photo. This comparison-related increase in perceived similarity is likely to be amplified in the context of border control for several reasons. First, comparisons at passport control are unlikely to engage habituation, because Border Force agents need to perform multiple tasks concurrently to both verify the identities of travellers, and assess their right to enter the country. Second, the likelihood that images of different individuals are erroneously judged to be the same increases with task repetition (Alenezi & Bindemann, 2013; Alenezi et al., 2015). Third, the comparison-related increases in perceived similarity might be further amplified in this context by category-level processing. The notion that Border Force agents would process faces at the categorical level seems somewhat counterintuitive, because their role is to perform identity matches. However, whilst they may not rely on simple gender or age-based categories, they might rely on other, more functional categories, when performing their duties. For example, they may deploy a simple 'suspicious – not suspicious' heuristic to determine who to scrutinise more closely. To illustrate this reasoning, here is an extract from an interview with a Border Force agent: "First impressions are very important … For example, if you stopped someone who is innocent but just acting nervously, they'd be looking at you going, 'What are you doing?' They want to interact with you" (Snowdon, 2009 ,Paragraph 11). In this scenario, the face pairs that are not suspicious are more likely to be subjected to a less effortful, categorical form of processing.

According to Carvalho and Goldstone (2014) when two objects are presumed to belong to the same category, attention is directed towards the similarities between them. This attentional bias could increase the perceived similarity of the compared faces. This hypothesis could be tested in an experiment that provides category information about the faces in a pair. For two given faces, the comparison related increase should be attenuated if they are presented as belonging to different groups, as opposed to the same group. Another way in which the category of a face might influence performance at border control is through the own-ethnicity bias. As discussed in Chapter 2 one of the two dominant causal explanations attributes the own-ethnicity advantage in face recognition and face matching performance to the category-based encoding of other-ethnicity faces (for a review see Wilson, Hugenberg, & Bernstein, 2013). It would be interesting to explore whether comparison affects the perceived similarity of own-ethnicity and other-ethnicity faces differently. If the effect of comparison on perceived similarity is affected by the own-ethnicity bias then it would be expected that comparison-related decreases in perceived similarity would appear sooner for other-ethnicity faces (say when two faces have a 90% phenotypic similarity) than for own ethnicity faces (which might need to have a 98% phenotypic similarity).

6.5.2 Implications for CCTV identifications

Unfamiliar face matching from CCTV footage is a challenging task because the amount and quality of face-specific information available is often poor. It is therefore unsurprising that matching unfamiliar faces from CCTV footage has been found to be error prone (Bruce et al., 2001; Davies & Thasen, 2000; Lee et al., 2009), even when the person who is matched is physically present (Davis & Valentine, 2009). The findings of Experiments 6 and 7 suggest that one source of errors on this task are comparison-related decreases in perceived similarity. One possible solution to improve unfamiliar face-matching accuracy that has shown promise is the provision of performance feedback (White, Kemp, Jenkins, & Burton, 2014). However, feedback may not be effective in counteracting the comparison-related bias towards ‘different’ responses observed in Experiments 6 and 7. It has been argued in this thesis that the cognitive process that gives rise to comparison-related decreases in perceived similarity is habituation, which is an automatic, passive process that may not be affected by performance feedback. This assumption could be tested in an experiment that provides performance feedback on a series of trials that resemble the experimental tasks deployed in Experiment 7.

Another possible solution to improve performance on the task of matching unfamiliar faces from CCTV footage is to provide specialist training. Feature-by-feature comparisons have been shown to improve unfamiliar face matching accuracy in untrained students (Towler et al., 2017). This finding suggests that feature-comparison training might help to counteract the comparison-related bias towards ‘different’ responses identified in Experiments 6 and 7. However, the study also compared the performance of untrained students to those of experienced forensic facial examiners. The superior performance of the examiners, who were presumably more effective at performing the feature comparison task than the untrained students, was found to be driven by mismatch trial accuracy (Towler et al., 2017). Considering that the similarity-, and difference-focussed listing tasks that were used in Experiments 6 and 7 are likely to have encouraged feature-based comparisons, this finding suggests that—rather than counteracting the comparison-related decrease in perceived similarity—detailed feature-based comparisons increase it. The latter interpretation is supported by the findings of Experiment 3, which have found a comparison-related change in the perceived similarity of stimulus pairs that are particularly amenable to feature-based processing. More research is needed to establish the relationship between feature-based processing and the effect of comparison on perceived similarity. A useful starting point would be to contrast performance on three comparison tasks: an open-ended comparison task (similar to the one deployed in the experiments reported here), a feature-by-feature comparison task, and a holistic comparison task (in which participants are asked to compare the two faces in a pair on holistic traits).

6.6 Conclusion

A series of experiments conducted by Loftus and her colleagues has highlighted the fallibility and malleability of eye-witness memory (Loftus, 2005; Loftus & Hoffman, 1989; Loftus et al., 1987). These findings have informed best practices in the context of criminal justice. More recent evidence suggest that unfamiliar face matching is also surprisingly error prone, even when high quality images are used. The experiments reported in this thesis offer one possible explanation for these misidentifications. They suggest that changes in perceived similarity arise through the interaction between comparative processes and stimulus attributes. Performing a comparison once, before performing a rating task, is sufficient to alter the perceived similarity of unfamiliar face stimuli. Furthermore, perceived similarity also appears to change over the course of a single comparison. Alternative accounts based on increases in familiarity are unlikely to explain these findings.

Whilst comparison is necessary, it does not drive the direction of the comparison-related changes in perceived similarity. These are determined predominantly by the stimulus. In particular, comparison appears to increase the perceived similarity of unfamiliar face stimuli that are relatively easy to tell apart, and decreases the perceived similarity of unfamiliar face stimuli that are difficult to discriminate. These comparison-related changes in perceived similarity are driven both by the phenotypic resemblance between the two faces (e.g. the interaction between genes and environment) and by the pictorial similarities and differences between the stimuli (i.e. all the other differences between the compared stimuli that arise from environmental, system, and person variables). This thesis has proposed a model within which to systematically study the effects of these variables. Doing so could be used to further develop theories of comparison, and more importantly could be used to predict performance in applied settings.

References

Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, *12*(3), 3–3. https://doi.org/10.1167/12.3.3

Abudarham, N., & Yovel, G. (2016). Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision*, *16*(3), 40. https://doi.org/10.1167/16.3.40

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*(6), 735–753. https://doi.org/10.1002/acp.2968

Alenezi, H. M., Bindemann, M., Fysh, M. C., & Johnston, R. A. (2015). Face matching in a long task: enforced rest and desk-switching cannot maintain identification accuracy. *PeerJ*, *3*, e1184. https://doi.org/10.7717/peerj.1184

Amishav, R., & Kimchi, R. (2010). Perceptual integrality of componential and configural information in faces. *Psychonomic Bulletin and Review*, *17*(5), 743–748. https://doi.org/10.3758/PBR.17.5.743

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *The Quarterly Journal of Experimental Psychology*, *68*(10), 2041–2050. https://doi.org/10.1080/17470218.2014.1003949

Angulo, R., & Alonso, G. (2013). Attentional changes in human perceptual learning. *Behavioural Processes*, *98*, 61–68. https://doi.org/10.1016/j.beproc.2013.05.009

Ao, M., Yi, D., Lei, Z., & Li, S. Z. (2009). Face Recognition at a Distance: System Issues. In M. Tistarelli, S. Z. Li, & R. Chellappa (Eds.), *Handbook of Remote Biometrics* (pp. 155–167). https://doi.org/10.1007/978-1-84882-385-3_6

Ashworth, A. R. S., Vuong, Q. C., Rossion, B., & Tarr, M. J. (2008). Recognizing rotated faces and Greebles: What properties drive the face inversion effect? *Visual Cognition*, *16*(6), 754–784. https://doi.org/10.1080/13506280701381741

Attwood, A. S., Penton-Voak, I. S., Burton, A. M., & Munafò, M. R. (2013). Acute anxiety impairs accuracy in identifying photographed faces. *Psychological Science*, *24*(8), 1591–4. https://doi.org/10.1177/0956797612474021

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia*, *49*(9), 2541–2552. https://doi.org/10.1016/j.neuropsychologia.2011.05.002

Bailey, T. M., & Hahn, U. (2001). Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory and Language*, *44*(4), 568–591. https://doi.org/10.1006/jmla.2000.2756

Baker, K., Laurence, S., & Mondloch, C. (2016). How does a newly encountered face become familiar? *Journal of Vision*, *16*(12), 913. Retrieved from http://dx.doi.org/10.1167/16.12.913

Bakhshi, S., Shamma, D. a., & Gilbert, E. (2014). Faces engage us. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, 965–974. https://doi.org/10.1145/2556288.2557403

Barton, J. J. S. (2009). What is meant by impaired configural processing in acquired prosopagnosia? *Perception*, *38*(2), 242–260. https://doi.org/10.1068/p6099

Barton, J. J. S., & Cherkasova, M. V. (2005). Impaired spatial coding within objects but not between objects in prosopagnosia. *Neurology*, *65*(2), 270–274. https://doi.org/10.1212/01.wnl.0000168901.98616.1c

Bassok, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 522–533. https://doi.org/10.1037/0278-7393.16.3.522

Beattie, L., Walsh, D., McLaren, J., Biello, S. M., & White, D. (2016). Perceptual impairment in face identification with poor sleep. *Royal Society Open Science*, *3*(10), 160321–160328. https://doi.org/10.1098/rsos.160321

Bedford, F. L. (1999). Keeping perception accurate. *Trends in Cognitive Sciences*, *3*(1), 4–11. https://doi.org/10.1016/S1364-6613(98)01266-2

Behrmann, M., Avidan, G., Marotta, J. J., & Kimchi, R. (2005). Detailed Exploration of Face-related Processing in Congenital Prosopagnosia: 1. Behavioral Findings. *Journal of Cognitive Neuroscience*, *17*(7), 1130–1149. https://doi.org/10.1162/0898929054475154

Behrmann, M., Richler, J. J., Avidan, G., & Kimchi, R. (2015). Holistic face perception. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organisation* (pp. 758–774). Oxford, United Kingdom: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199686858.013.010

Behrmann, M., Scherf, K. S., & Avidan, G. (2016). Neural mechanisms of face perception, their emergence over development, and their breakdown. *Wiley Interdisciplinary Reviews. Cognitive Science*. https://doi.org/10.1002/wcs.1388

Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar-face matching. *Applied Cognitive Psychology*, *27*(6), 707–717. https://doi.org/10.1002/acp.2970

Bindemann, M., Avetisyan, M., & Blackwell, K.-A. (2010). Finding needles in haystacks: Identity mismatch frequency and facial identity verification. *Journal of Experimental Psychology: Applied*, *16*(4), 378–386. https://doi.org/10.1037/a0021893

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology. Applied*, *18*(3), 277–91. https://doi.org/10.1037/a0029635

Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. R. (2008). Brain potential correlates of face recognition: Geometric distortions and the N250r brain response to stimulus repetitions. *Psychophysiology*, *45*(4), 535–544. https://doi.org/10.1111/j.1469-8986.2008.00663.x

Bindemann, M., Fysh, M., Cross, K., & Watts, R. (2016). Matching faces against the clock. *I-Perception*, *7*(5), 1–18. https://doi.org/10.1177/2041669516672219

Bindemann, M., & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception*, *40*(5), 625–627. https://doi.org/10.1068/p7008

Biotti, F., Wu, E., Yang, H., Jiahui, G., Duchaine, B., & Cook, R. (2017). Normal composite face effects in developmental prosopagnosia. *Cortex*, *95*, 63–76. https://doi.org/10.1016/j.cortex.2017.07.018

Blanton, A., Allen, K. C., Miller, T., Kalka, N. D., & Jain, A. K. (2016). A Comparison of Human and Automated Face Verification Accuracy on Unconstrained Image Sets. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 229–236). IEEE. https://doi.org/10.1109/CVPRW.2016.35

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016). An in-depth cognitive examination of individuals with superior face recognition skills. *Cortex*, *82*, 48–62. https://doi.org/10.1016/j.cortex.2016.05.003

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem:

Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0148148

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). Super-recognisers in Action: Evidence from Face-matching and Face Memory Tasks. *Applied Cognitive Psychology*, *30*(1), 81–91. https://doi.org/10.1002/acp.3170

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology*, *7*(SEP). https://doi.org/10.3389/fpsyg.2016.01378

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*(5), 527–536. Retrieved from http://dx.doi.org/10.1080/13506280244000168%5Cnpapers3://publication/doi/10.1080/13506280244000168

Boroditsky, L. (2007). Comparison and the development of knowledge. *Cognition*, *102*(1), 118–128. https://doi.org/10.1016/j.cognition.2002.08.001

Bressan, P., & Dal Martello, M. F. (2002). Talis pater, talis filius: Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, *13*(3), 213–218. https://doi.org/10.1111/1467-9280.00440

Brignall, M. (2016, November 26). Handcuffed in Waitrose: the innocent man in search of justice. *The Guardian*. Retrieved from https://www.theguardian.com/business/2016/nov/26/innocent-man-handcuffed-waitrose-mistaken-identity-nightmare-christopher-seddon

Brown, C., & Lloyd-Jones, T. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory & Cognition*, *34*(2), 277–286. https://doi.org/10.3758/BF03193406

Brown, C., & Lloyd-Jones, T. J. (2005). Verbal facilitation of face recognition. *Memory & Cognition*, *33*(8), 1442–1456. https://doi.org/10.3758/BF03193377

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*(1), 105–116. https://doi.org/10.1111/j.2044-8295.1982.tb01795.x

Bruce, V. (2012). Familiar face recognition. In *Craniofacial Identification* (pp. 1–10). https://doi.org/10.1017/CBO9781139049566.001

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. https://doi.org/10.1037/1076-898X.5.4.339

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207–218. https://doi.org/10.1037/1076-898X.7.3.207

Bruce, V., Valentine, T., & Baddeley, A. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, *1*(2), 109–120. https://doi.org/10.1002/acp.2350010204

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Bruyer, R., & Crispeels, G. (1992). Expertise in person recognition. *Bulletin of the Psychonomic Society*, *30*(6), 501–504. https://doi.org/10.3758/BF03334112

Bruyer, R., Galvez, C., & Prairial, C. (1993). Effect of disorientation on visual analysis, familiarity decision and semantic decision on faces. *British Journal of*

Psychology (London, England : 1953), *84 (Pt 4)*, 433–41. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8298857

Buolamwini, J. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification *. Proceedings of Machine Learning Research* (Vol. 81). Retrieved from http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces. *Vision Research*, *47*(16), 2164–2169. https://doi.org/10.1016/j.visres.2007.04.018

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1467–1485. https://doi.org/10.1080/17470218.2013.800125

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943–958. https://doi.org/10.1111/j.2044-8295.2011.02039.x

Burton, A. M., White, D., & McNeill, A. (2010). The glasgow face matching test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

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. https://doi.org/10.1111/1467-9280.00144

Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., & Rossion, B. (2010). Holistic perception of the individual face is specific and necessary: Evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia*, *48*(14), 4057–4092. https://doi.org/10.1016/j.neuropsychologia.2010.09.017

Busigny, T., & Rossion, B. (2010). Acquired prosopagnosia abolishes the face inversion effect. *Cortex*, *46*(8), 965–981. https://doi.org/10.1016/j.cortex.2009.07.004

Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research*, *38*(15–16), 2455–2468. https://doi.org/10.1016/S0042-6989(97)00469-0

Cabeza, R., & Kato, T. (2000). Features are Also Important: Contributions of Featural and Configural Processing to Face Recognition. *Psychological Science*, *11*(5), 429–433. https://doi.org/10.1111/1467-9280.00283

Campbell, A., & Tanaka, J. W. (2018). Inversion Impairs Expert Budgerigar Identity Recognition: A Face-Like Effect for a Nonface Object of Expertise. *Perception*, *47*(6), 647–659. https://doi.org/10.1177/0301006618771806

CCTV: Too many cameras useless, warns surveillance watchdog Tony Porter. (2015). Retrieved August 28, 2017, from http://www.bbc.co.uk/news/uk-30978995

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the Diagnostic Utility of Facial Composites: Beliefs of Guilt Can Bias Perceived Similarity Between Composite and Suspect. *Journal of Experimental Psychology: Applied*, *15*(1), 76–90. https://doi.org/10.1037/a0014682

Charman, S. D., Kavetski, M., & Mueller, D. H. (2017). Cognitive bias in the legal system: Police Officers evaluate ambiguous evidence in a belief-consistent Manner. *Journal of Applied Research in Memory and Cognition*, *6*(2), 193–202. https://doi.org/https://doi.org/10.1016/j.jarmac.2017.02.001

Chin, M., Evans, K., Wolfe, J., & Tanaka, J. W. (2017). Inversion Effects in the Expert Classification of Mammograms and Faces. *Cognitive Research:*

Principles and Implications, *17*(10). https://doi.org/10.1167/17.10.1226
Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*(8), 985–994. https://doi.org/10.1068/p3335
Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, *11*(7), 857–869. https://doi.org/10.1080/13506280444000021
Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, *17*(1), 97–116. https://doi.org/10.1080/09541440340000439
Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, *17*(1), 97–116. https://doi.org/10.1080/09541440340000439
Cohen, M. E., & Carr, W. J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, *6*(4), 383–384. https://doi.org/10.3758/BF03333209
Cole, S. (2005). *More Than Zero: Accounting for Error in Latent Fingerprint Identification. Journal of Criminal Law & Criminology* (Vol. 95). https://doi.org/10.3366/ajicl.2011.0005
Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, *29*(8), 893–909. https://doi.org/10.1068/p2949
Collishaw, S. M., & Hole, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception*, *31*(3), 287–296. https://doi.org/10.1068/p3195
Corcoran, K. (2015, September 17). Father-of-four faces being sent back to Zimbabwe because he gained nine stone and no longer looks like his passport photo. *MailOnline*. Retrieved from http://www.dailymail.co.uk/news/article-3238172/Father-four-faces-sent-Zimbabwe-gained-nine-stone-no-longer-looks-like-passport-photo.html
Davies, G. M., & Young, A. W. (2017). Research on face recognition: The Aberdeen influence. *British Journal of Psychology*, *108*(4), 812–830. https://doi.org/10.1111/bjop.12243
Davies, G., & Thasen, S. (2000). Closed-circuit television: How effective an identification aid? *British Journal of Psychology*. https://doi.org/10.1348/000712600161907
Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating Predictors of Superior Face Recognition Ability in Police Super-recognisers. *Applied Cognitive Psychology*, *30*(6), 827–840. https://doi.org/10.1002/acp.3260
Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23*(4), 482–505. https://doi.org/10.1002/acp.1490
Davis, J. P., & Valentine, T. (2015). Human Verification of Identity from Photographic Images. In *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV* (pp. 211–238). Chichester, England: Wiley-Blackwell. https://doi.org/10.1002/9781118469538.ch9
Davis, J. P., Valentine, T., & Wilkinson, C. M. (2012). Facial image comparison. In C. M. Wilkinson & C. Ryan (Eds.), *Craniofacial identification* (pp. 136–153). Cambridge: Cambridge University Press. https://doi.org/DOI: https://doi.org/10.1017/CBO9781139049566.012

de Frias, C., Nilsson, L.-G., & Herlitz, A. (2006). Sex Differences in Cognition are Stable Over a 10-Year Period in Adulthood and Old Age. *Aging, Neuropsychology, and Cognition*, *13*(3–4), 574–587. https://doi.org/10.1080/13825580600678418

De Gelder, B., Bachoud-Lévi, A. C., & Degos, J. D. (1998). Inversion superiority in visual agnosia may be common to a variety of orientation polarised objects besides faces. *Vision Research*, *38*(18), 2855–2861. https://doi.org/10.1016/S0042-6989(97)00458-6

De Gelder, B., & Rouw, R. (2000). Paradoxical configuration effects for faces and objects in prosopagnosia. *Neuropsychologia*, *38*(9), 1271–1279. https://doi.org/10.1016/S0028-3932(00)00039-7

DeGutis, J., Mercado, R. J., Wilmer, J., & Rosenblatt, A. (2013). Individual Differences in Holistic Processing Predict the Own-Race Advantage in Recognition Memory. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0058253

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Delamater, a R., & Joseph, P. (2000). Common coding in symbolic matching tasks with humans: training with a common consequence or antecedent. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *53*(3), 255–73. https://doi.org/10.1080/713932729

DeNicola, C. A., Holt, N. A., Lambert, A. J., & Cashon, C. H. (2013). Attention-orienting and attention-holding effects of faces on 4- to 8-month-old infants. *International Journal of Behavioral Development*, *37*(2), 143–147. https://doi.org/10.1177/0165025412474751

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology. General*, *115*(2), 107–17. https://doi.org/10.1037/0096-3445.115.2.107

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, *106*(3), 433–445. https://doi.org/10.1111/bjop.12103

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Quarterly Journal of Experimental Psychology*, *69*(1), 1–10. https://doi.org/10.1080/17470218.2015.1017513

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification*, *56*(4), 600–616. https://doi.org/Doi: 10.1016/s1355-0306(98)72118-5

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74–78. https://doi.org/10.1016/j.forsciint.2005.10.017

Dror, I. E., Péron, A. E., Hind, S. L., & Charlton, D. (2005). When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, *19*(6), 799–809. https://doi.org/10.1002/acp.1130

Dror, I. E., Wertheim, K., Fraser-Mackenzie, P., & Walajtys, J. (2012). The Impact of Human–Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts. *Journal of*

Forensic Sciences, *57*(2), 343–352. https://doi.org/10.1111/j.1556-4029.2011.02013.x
Dwyer, D. M., Mundy, M. E., & Honey, R. C. (2011). The role of stimulus comparison in human perceptual learning: effects of distractor placement. *Journal of Experimental Psychology. Animal Behavior Processes*, *37*, 300–307. https://doi.org/10.1037/a0023078
Dwyer, D. M., Mundy, M. E., Vladeanu, M., & Honey, R. C. (2009). Perceptual learning and acquired face familiarity: Evidence from inversion, use of internal features, and generalization between viewpoints. *Visual Cognition*, *17*(3), 334–355. https://doi.org/10.1080/13506280701757577
Edmond, G., Davis, J. P., & Valentine, T. (2015). Expert analysis: Facial image comparison. In T. Valentine & J. P. Davis (Eds.), *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV* (pp. 239–262). Chichester, England: Wiley-Blackwell.
Edmond, G., & Wortley, N. (2016). Interpreting image evidence: Facial mapping, police familiars and super-recognisers in England and Australia. *Journal of International and Comparative Law*, *3*(2), 473–522.
Ellis, H. C., Feuge, R. L., Long, K. K., & Pegram, V. G. (1964). Evidence for acquired equivalence o cues in a perceptual task. *Perceptual and Motor Skills*, *19*, 159–162.
Ellis, H. D. (1991). The Effects of Distinctiveness, Presentation Time and Delay on Face Recognition. *European Journal of Cognitive Psychology*, *3*(1), 137–145. https://doi.org/10.1080/09541449108406223
Engfors, L. M., Jeffery, L., Gignac, G. E., & Palermo, R. (2017). Individual differences in adaptive norm-based coding and holistic coding are associated yet each contributes uniquely to unfamiliar face recognition ability. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 281–293. https://doi.org/10.1037/xhp0000265
Etchells, D. B., Brooks, J. L., & Johnston, R. A. (2017). Evidence for view-invariant face recognition units in unfamiliar face learning. *Quarterly Journal of Experimental Psychology*, *70*(5), 874–889. https://doi.org/10.1080/17470218.2016.1248453
Facial Identification Scientific Working Group FISWG. (2012a). *Guidelines for Facial Comparison Methods*. Retrieved from https://fiswg.org/FISWG_GuidelinesforFacialComparisonMethods_v1.0_2012_02_02.pdf
Facial Identification Scientific Working Group FISWG. (2012b). *Recommendations for a Training Program in Facial Comparison*. Retrieved from https://fiswg.org/FISWG_RecommendationsForTrainingProgram_v1.0_2012_02_02.pdf
Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 628–634. https://doi.org/10.1037/0096-1523.21.3.628
Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*(3), 482–498. https://doi.org/10.1037/0033-295X.105.3.482
Favelle, S., Hill, H., & Claes, P. (2017). About Face: Matching Unfamiliar Faces Across Rotations of View and Lighting. *I-Perception*, *8*(6). https://doi.org/10.1177/2041669517744221
Favelle, S. K., Palmisano, S., & Avery, G. (2011). Face viewpoint effects about three

axes: The role of configural and featural processing. *Perception*, *40*(7), 761–784. https://doi.org/10.1068/p6878

Findley, K. A., & Scott, M. S. (2006). The Multiple Dimensions of Tunnel Vision in Criminal Cases. *Wisconsin Law Revieve*, *2*(2005), 292–397. https://doi.org/10.3366/ajicl.2011.0005

Finzi, R. D., Susilo, T., Barton, J. J. S., & Duchaine, B. C. (2016). The role of holistic face processing in acquired prosopagnosia: evidence from the composite face effect. *Visual Cognition*, *24*(4), 304–320. https://doi.org/10.1080/13506285.2016.1261976

Förster, B., Gebhardt, R. P., Lindlar, K., Siemann, M., & Delius, J. D. (1996). Mental-rotation effect: A function of elementary stimulus discriminability? *Perception*, *25*(11), 1301–1316. https://doi.org/10.1068/p251301

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configurai information: Direct evidence. *Perception*, *29*(2), 159–170. https://doi.org/10.1068/p3012

Frowd, C. D. (2015). Facial composites and techniques to improve image recognisability. In T. Valentine & J. P. Davis (Eds.), *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV* (pp. 43–70). Chichester, England: Wiley-Blackwell.

Frowd, C. D. (2017). Facial composite systems: Production of an identifiable face. In M. Bindemann & A. M. Megreya (Eds.), *Face Processing: Systems, Disorders and Cultural Differences* (pp. 55–86). Nova Science.

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., Mclanaghan, S., & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime and Law*, *11*(1), 33–52. https://doi.org/10.1080/10683160310001634313

Frowd, C. D., Hancock, P. J. B., Bruce, V., Skelton, F. C., Atherton, C., Nelson, L., … Sendrea, G. (2011). Catching More Offenders with EvoFIT Facial Composites: Lab Research and Police Field Trials. *Global Journal of Human Social Science*, *11*(3), 35–46. Retrieved from http://globaljournals.org/GJHSS_Volume11/5-Catching-More-Offenders-with-Evofit.pdf

Frowd, C. D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., … Hancock, P. J. B. (2011). The psychology of face construction: Giving evolution a helping hand. *Applied Cognitive Psychology*, *25*(2), 195–203. https://doi.org/10.1002/acp.1662

Frowd, C. D., Pitchford, M., Skelton, F., Petkovic, A., Prosser, C., & Coates, B. (2012). Catching even more offenders with EvoFIT facial composites. In *Proceedings - 3rd International Conference on Emerging Security Technologies, EST 2012* (pp. 20–26). https://doi.org/10.1109/EST.2012.26

Frowd, C. D., Skelton, F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., … Hancock, P. J. B. (2012). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied*, *18*(2), 224–238. https://doi.org/10.1037/a0027393

Fuhrmann, D., Knoll, L. J., Sakhardande, A. L., Speekenbrink, M., Kadosh, K. C., & Blakemore, S.-J. (2016). Perception and recognition of faces in adolescence. *Scientific Reports*, *6*(1), 33497. https://doi.org/10.1038/srep33497

Fysh, M. C., & Bindemann, M. (2017). Effects of time pressure and time passage on face-matching accuracy. *Royal Society Open Science*, *4*(6), 1–13. https://doi.org/10.1098/rsos.170249

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. https://doi.org/10.1038/72140

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of Experimental Psychology. Human Perception and Performance*, *28*(2), 431–446. https://doi.org/10.1037/0096-1523.28.2.431

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training “greeble” experts: A framework for studying expert object recognition processes. *Vision Research*, *38*(15–16), 2401–2428. https://doi.org/10.1016/S0042-6989(97)00442-2

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy*. *Cognitive Science*, *7*(2), 155–170. Retrieved from http://dx.doi.org/10.1016/S0364-0213(83)80009-3%5Cnhttp://doi.wiley.com/10.1016/S0364-0213(83)80009-3%5Cnhttp://onlinelibrary.wiley.com/doi/10.1207/s15516709cog0702_3/abstract

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition*, *29*(4), 565–577. https://doi.org/10.3758/BF03200458

Gentner, D., & Markman, A. B. (1997). Structure Mapping in Analogy and Similarity. *American Psychologist*, *52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Gentner, D., & Namy, L. L. (1999). Comparison in the Development of Categories. *Cognitive Development*, *14*(4), 487–513. https://doi.org/10.1016/S0885-2014(99)00016-7

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277–300. https://doi.org/10.1016/S0364-0213(86)80019-2

Gibson, E. J. (1969). Principles of perceptual learning and development. *Principles of Perceptual Learning and Development.*, 353. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1969-35014-000&lang=fr&site=ehost-live

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces Attract Infants’ Attention in Complex Displays. *Infancy*, *14*(5), 550–562. https://doi.org/10.1080/15250000903144199

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. https://doi.org/10.1167/6.6.272

Going, M., & Read, J. D. (1974). Effects of Uniqueness, Sex of Subject, and Sex of Photograph on Facial Recognition. *Perceptual and Motor Skills*, *39*, 109–110. https://doi.org/10.2466/pms.1974.39.1.109

Goldstone, R. L. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 381–386. https://doi.org/10.3758/BF03204653

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 29–50. https://doi.org/10.1037/0278-7393.20.1.29

Grill-Spector, K., Sayres, R., & Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neuroscience*, *9*(9), 1177–1185. https://doi.org/10.1038/nn1745

Grimm, L. R., Rein, J. R., & Markman, A. B. (2012). Determining transformation distance in similarity: Considerations for assessing representational changes a priori. *Thinking & Reasoning*, *18*(1), 59–80. https://doi.org/10.1080/13546783.2011.649094

Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition*, *97*(3), 227–267. https://doi.org/10.1016/j.cognition.2004.09.006

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*, 1–32. https://doi.org/doi:10.1016/S0010-0277(02)00184-1

Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: associative and nonassociative mechanisms. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *56*(1), 43–55. https://doi.org/10.1080/02724990244000151

Hall, G., Mitchell, C., Graham, S., & Lavis, Y. (2003). Acquired equivalence and distinctiveness in human discrimination learning: Evidence for associative mediation. *Journal of Experimental Psychology: General*, *132*(2), 266–276. https://doi.org/10.1037/0096-3445.132.2.266

Hampton, J. A. (1995). Testing the Prototype Theory of Concepts. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1995.1031

Hancock, P. J. B., Bruce, V., & Mike Burton, A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(00)01519-9

Harley, E. M., Pope, W. B., Villablanca, J. P., Mumford, J., Suh, R., Mazziotta, J. C., … Engel, S. A. (2009). Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. *Cerebral Cortex*, *19*(11), 2746–2754. https://doi.org/10.1093/cercor/bhp051

Hassin, R. R. (2001). Making features similar: comparison processes affect perception. *Psychonomic Bulletin & Review*, *8*(4), 728–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11848592

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, *106*(2), 1017–1027. https://doi.org/10.1016/j.cognition.2007.04.002

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the Faces of Robbers Captured on Video. *Applied Cognitive Psychology*, *15*(4), 445–464. https://doi.org/10.1002/acp.718

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, *21*(9–10), 1306–1336. https://doi.org/http://dx.doi.org/10.1080/13506285.2013.823140

Herlitz, A., Nilsson, L. G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition*, *25*(6), 801–811. https://doi.org/10.3758/BF03211324

Herlitz, A., Reuterskiöld, L., Lovén, J., Thilers, P. P., & Rehnman, J. (2013). Cognitive Sex Differences Are Not Magnified as a Function of Age, Sex Hormones, or Puberty Development During Early Adolescence. *Developmental Neuropsychology*, *38*(3), 167–179.

https://doi.org/10.1080/87565641.2012.759580

Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology*, *13*(2), 357–371. https://doi.org/10.1348/135532507X238682

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(4), 986. https://doi.org/10.1037/0096-1523.22.4.986

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, *62*(2), 201–222. https://doi.org/10.1016/S0010-0277(96)00785-8

Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hockley, W., Hemsworth, D., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition*, *27*(I), 128–138. https://doi.org/10.3758/BF03201219

Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. *Cognition*, *113*(1), 62–79. https://doi.org/10.1016/j.cognition.2009.07.010

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

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*(4), 332–340. https://doi.org/10.3758/BF03197035

Honey, R. C., & Bateson, P. (1996). Stimulus comparison and perceptual learning: Further evidence and evaluation from an imprinting procedure. *The Quarterly Journal of Experimental Psychology Section B : Comparative and Physiological Psychology*, *49*(3), 259–269. https://doi.org/10.1080/713932631

Honey, R. C., Bateson, P., & Horn, G. (1994). The role of stimulus comparison in perceptual learning: an investigation with the domestic chick. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *47*(1), 83–103. https://doi.org/10.1080/14640749408401349

Honey, R. C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology. Animal Behavior Processes*, *15*(4), 338–346. https://doi.org/10.1037/0097-7403.15.4.338

Honey, R. C., & Hall, G. (1991). The Quarterly Journal of Experimental Psychology Section B Acquired equivalence and distinctiveness of cues using a sensory-preconditioning procedure Acquired Equivalence and Distinctiveness of Cues Using a Sensory-Preconditioning Procedure. *The Quarterly Journal Of Experimental Psychology*, (907217966). https://doi.org/10.1080/14640749108401263

Hong Liu, C., Collin, C. A., Burton, A. M., & Chaudhuri, A. (1999). Lighting direction affects recognition of untextured faces in photographic positive and negative. *Vision Research*, *39*(24), 4003–4009. https://doi.org/10.1016/S0042-6989(99)00109-1

Huang, L., Song, Y., Li, J., Zhen, Z., Yang, Z., & Liu, J. (2014). Individual differences in cortical face selectivity predict behavioral performance in face recognition. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00483

Itz, M. L., Schweinberger, S. R., & Kaufmann, J. M. (2017). Caricature generalization benefits for faces learned with enhanced idiosyncratic shape or texture. *Cognitive, Affective and Behavioral Neuroscience*, *17*(1), 185–197. https://doi.org/10.3758/s13415-016-0471-y

Jeckeln, G., Hahn, C. A., Noyes, E., Cavazos, J. G., & O'Toole, A. J. (2018). Wisdom of the social versus non-social crowd in face identification. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12291

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Phil. Trans. R. Soc. B*, 1671–1683. https://doi.org/10.1098/rstb.2010.0379

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323. https://doi.org/10.1016/j.cognition.2011.08.001

Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(98)01152-8

Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: Effects of lighting direction, inversion, and brightness reversal. *Perception*, *21*, 365–375. https://doi.org/10.1068/p210365n

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: a review. *Memory*, *17*(5), 577–596. https://doi.org/10.1080/09658210902976969

Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition*, *4*(1), 59–67. https://doi.org/10.1080/713756748

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, *13*(4), 289–303. https://doi.org/10.3758/BF03202498

Jones, K. A., Crozier, W. E., & Strange, D. (2017). Believing is seeing: Biased viewing of body-worn camera footage. *Journal of Applied Research in Memory and Cognition*, *6*(4), 460–474. https://doi.org/https://doi.org/10.1016/j.jarmac.2017.07.007

Jones, S. P., & Dwyer, D. M. (2013). Perceptual learning with complex visual stimuli is based on location, rather than content, of discriminating features. *Journal of Experimental Psychology: Animal Behavior Processes*, *39*(2), 152–165. https://doi.org/10.1037/a0031509

Kaiser, D. H., Sherburne, L. M., Steirn, J. N., & Zentall, T. R. (1997). Perceptual learning in pigeons: Decreased ability to discriminate samples mapped onto the same comparison in many-to-one matching. *Psychonomic Bulletin & Review*, *4*(3), 378–381.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Kassin, S. M., Bogart, D., & Kerner, J. (2012). Confessions that corrupt: Evidence from the dna exoneration case files. *Psychological Science*, *23*(1), 41–45. https://doi.org/10.1177/0956797611422918

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, *27*(2), 187–203. https://doi.org/10.1023/A:1022599230598

Kaufmann, J. M., Schweinberger, S. R., & Burton, A. M. (2009). N250 ERP

Correlates of the Acquisition of Face Representations across Different Images. *Journal of Cognitive Neuroscience*, *21*(4), 625–641. https://doi.org/10.1162/jocn.2009.21080

Kemp, R., Towell, N., & Pike, G. (1997). When Seeing should not be Believing: Photographs, Credit Cards and Fraud. *Appl. Cognit. Psychol*, *11*(October 1995), 211–222. https://doi.org/10.1002/(SICI)1099-0720

Kerstholt, J., Eikelboom, A., Dijkman, T., Stoel, R., Hermsen, R., & van Leuven, B. (2010). Does suggestive information cause a confirmation bias in bullet comparisons? *Forensic Science International*, *198*(1–3), 138–142. https://doi.org/10.1016/j.forsciint.2010.02.007

Kimchi, R., & Amishav, R. (2010). Faces as perceptual wholes: The interplay between component and configural properties in face processing. *Visual Cognition*, *18*(7), 1034–1062. https://doi.org/10.1080/13506281003619986

Kimchi, R., Behrmann, M., Avidan, G., & Amishav, R. (2012). Perceptual separability of featural and configural information in congenital prosopagnosia. *Cognitive Neuropsychology*, *29*(5–6), 447–463. https://doi.org/10.1080/02643294.2012.752723

Kohn, A. (2007). Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. https://doi.org/10.1152/jn.00086.2007

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, *21*(1), 38–43. https://doi.org/10.1177/0956797609356508

Kramer, R. S. S., & Ritchie, K. L. (2016). Disguising Superman: How Glasses Affect Unfamiliar Face Matching. *Applied Cognitive Psychology*, *30*(6), 841–845. https://doi.org/10.1002/acp.3261

Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision*, *15*(4), 1. https://doi.org/10.1167/15.4.1

Kramer, T. H., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal, and eyewitness memory. *Law and Human Behavior*, *14*(2), 167–184. https://doi.org/10.1007/BF01062971

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, *66*(5), 651–654. https://doi.org/10.1037/0021-9010.66.5.651

Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behavior*, *38*(3), 256–270. https://doi.org/10.1037/lhb0000066

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260. https://doi.org/10.1037//0033-295X.107.2.227

Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*(1), 330–342. https://doi.org/10.1016/j.cognition.2007.07.012

Lavis, Y., Kadib, R., Mitchell, C., & Hall, G. (2011). Memory for, and salience of, the unique features of similar stimuli in perceptual learning. *Journal of Experimental Psychology. Animal Behavior Processes*, *37*(2), 211–9. https://doi.org/10.1037/a0021888

Lavis, Y., & Mitchell, C. (2006). Effects of preexposure on stimulus discrimination: an investigation of the mechanisms responsible for human perceptual learning. *Q*

J Exp Psychol (Hove), *59*(12), 2083–2101. https://doi.org/10.1080/17470210600705198
Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition*, *61*(2), 139–158. https://doi.org/10.1016/j.bandc.2005.11.005
Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(2), 513–536. https://doi.org/10.1080/713755889
Leder, H., & Carbon, C. C. (2006). Face-specific configural processing of relational information. *British Journal of Psychology*, *97*(Pt 1), 19–29. https://doi.org/10.1348/000712605X54794
Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature Effects, Distinctiveness, and Identification: Testing the Face-Space Framework. *Psychological Science*, *11*(5), 379–385. https://doi.org/10.1111/1467-9280.00274
Lee, W. L., Wilkinson, C., Memon, A., & Houston, K. (2009). Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage: An evaluation of the effect of training on facial identification ability. *AXIS Online Jounal of Centre for Anatomy and Human Identification*, *1*(1), 19–28.
Liesefeld, H. R., Fu, X., & Zimmer, H. D. (2015). Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed-accuracy trade-offs. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(4), 1140–1151. https://doi.org/10.1037/xlm0000081
Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(3), 212–228. https://doi.org/10.1037/0278-7393.5.3.212
Lin, H. Y., & Hsieh, H. Y. (2012). The process of mental rotation in design: Holistic or piecemeal. In *2012 5th International Congress on Image and Signal Processing, CISP 2012* (pp. 766–769). https://doi.org/10.1109/CISP.2012.6469883
Liu, C. H., & Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. *Cognition*, *83*(1), 31–48. https://doi.org/10.1016/S0010-0277(01)00164-0
Liu, T. T., & Behrmann, M. (2014). Impaired holistic processing of left-right composite faces in congenital prosopagnosia. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00750
Lobmaier, J. S., Bölte, J., Mast, F. W., & Dobel, C. (2010). Configural and featural processing in humans with congenital prosopagnosia. *Advances in Cognitive Psychology*, *6*(6), 23–34. https://doi.org/10.2478/v10053-008-0074-4
Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review*, *79*(5), 410–419. https://doi.org/10.1037/h0033129
Lockhead, G. R. (1979). Holistic versus analytic process models: A reply. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 746–755. https://doi.org/10.1037/0096-1523.5.4.746
Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705
Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: the creation of

new memories. *Journal of Experimental Psychology. General*, *118*(1), 100–104. https://doi.org/10.1037/0096-3445.118.1.100

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about “weapon focus.” *Law and Human Behavior*, *11*(1), 55–62. https://doi.org/10.1007/BF01044839

Longmore, C. a, Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(1), 77–100. https://doi.org/10.1037/0096-1523.34.1.77

Love, B. C., Rouder, J. N., & Wisniewski, E. J. (1999). A structural account of global and local processing. *Cognitive Psychology*, *38*(2), 291–316. https://doi.org/10.1006/cogp.1998.0697

Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmet, N., Melsom, E. I., & Lindsay, R. C. L. (2012). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior*, *36*(6), 513–526. https://doi.org/10.1037/h0093937

Markman, A. B. (1996). structural alignment in similarity and difference judgments. *Psychonomic Bulletin & Review*, *3*(2), 227–230. https://doi.org/10.3758/BF03212423

Markman, A. B., & Gentner, D. (1993a). All differences are not created equal: A structural alignment view of similarity. *Advances in Neural Information Processing Systems*, *5*(4), 682–686. https://doi.org/10.1006/cogp.1993.1011

Markman, A. B., & Gentner, D. (1993b). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, *32*(4), 517–535. https://doi.org/10.1006/jmla.1993.1027

Markman, A. B., & Gentner, D. (1993c). Structural Alignment during Similarity Comparisons. *Cognitive Psychology*, *25*(4), 431–467. https://doi.org/http://dx.doi.org/10.1006/cogp.1993.1011

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition*, *24*(2), 235–249. https://doi.org/10.3758/BF03200884

Markman, A. B., & Gentner, D. (2005). Nonintentional similarity processing. In T. Hassin, J. Bargh, & J. Uleman (Eds.), *The New Unconscious* (pp. 107–137). New York: Oxford University Press.

Martini, M., Bufalari, I., Stazi, M. A., & Aglioti, S. M. (2015). Is that me or my twin? Lack of self-face recognition advantage in identical twins. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0120900

Matthews, C. M., & Mondloch, C. J. (2017). Improving Identity Matching of Newly Encountered Faces: Effects of Multi-image Training. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1016/j.jarmac.2017.10.005

Mauro, R., & Kubovy, M. (1992). Caricature and face recognition. *Memory & Cognition*, *20*(4), 433–440. https://doi.org/10.3758/BF03210927

McBain, R., Norton, D., & Chen, Y. (2009). Females excel at basic face perception. *Acta Psychologica*, *130*(2), 168–173. https://doi.org/10.1016/j.actpsy.2008.12.005

McCaffery, J. M., & Burton, A. M. (2016). Passport Checks: Interactions Between Matching Faces and Biographical Details. *Applied Cognitive Psychology*, *30*(6), 925–933. https://doi.org/10.1002/acp.3281

McGugin, R. W., Gatenby, J. C., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proceedings of the National Academy of*

Sciences, *109*(42), 17063–17068. https://doi.org/10.1073/pnas.1116333109
McGugin, R. W., Newton, A. T., Gore, J. C., & Gauthier, I. (2014). Robust expertise effects in right FFA. *Neuropsychologia*, *63*, 135–144. https://doi.org/10.1016/j.neuropsychologia.2014.08.029
McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research*, *69*, 10–22. https://doi.org/10.1016/j.visres.2012.07.014
McGugin, R. W., Tanaka, J. W., Lebrecht, S., Tarr, M. J., & Gauthier, I. (2011). Race-Specific Perceptual Discrimination Improvement Following Short Individuation Training With Faces. *Cognitive Science*, *35*(2), 330–347. https://doi.org/10.1111/j.1551-6709.2010.01148.x
Mcintyre, A. H., Hancock, P. J. B., Kittler, J., & Langton, S. R. H. (2013). Improving discrimination and face matching with caricature. *Applied Cognitive Psychology*, *27*(6), 725–734. https://doi.org/10.1002/acp.2966
McKone, E., & Robbins, R. (2012). Are Faces Special. In *Oxford Handbook of Face Perception*. https://doi.org/10.1093/oxfordhb/9780199559053.013.0009
Mckone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? toward a new theory of holistic processing. *Psychonomic Bulletin and Review*, *16*(5), 778–797. https://doi.org/10.3758/PBR.16.5.778
Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity Involving Attributes and Relations: Judgments of Similarity and Difference Are Not Inverses. *Psychological Science*, *1*(1), 64–69. https://doi.org/10.1111/j.1467-9280.1990.tb00069.x
Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*(2), 254–278. https://doi.org/10.1037/0033-295X.100.2.254
Medin, D. L., Goldstone, R. L., & Markman, A. B. (1995). Comparison and choice: Relations between similarity processes and decision processes. *Psychonomic Bulletin & Review*, *2*(1), 1–19. https://doi.org/10.3758/BF03214410
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207
Megreya, A. M., & Bindemann, M. (2009). Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception*, *38*(12), 1831–1848. https://doi.org/10.1068/p6385
Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology*, *25*(1), 30–37. https://doi.org/10.1080/20445911.2012.739153
Megreya, A. M., Bindemann, M., & Havard, C. (2011). Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica*, *137*(1), 83–89. https://doi.org/10.1016/j.actpsy.2011.03.003
Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876. https://doi.org/10.3758/BF03193433
Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364–372. https://doi.org/10.1037/a0013464

Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching face images taken on the same day or months apart: The limitations of photo ID. *Applied Cognitive Psychology*, *27*(6), 700–706. https://doi.org/10.1002/acp.2965

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

Meissner, C. A., & Brigham, J. C. (2001). Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review. *Psychology, Public Policy, and Law*. https://doi.org/10.1037//1076-8971.7.1.3

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own- and other-race faces. *Visual Cognition*, *21*(9–10), 1287–1305. https://doi.org/10.1080/13506285.2013.832451

Menon, N., White, D., & Kemp, R. I. (2015). Variation in Photos of the Same Face Drives Improvements in Identity Verification. *Perception*, *44*(11), 1332–1341. https://doi.org/10.1177/0301006615599902

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*(1), 55–73. https://doi.org/10.1080/13506280500158761

Mitchell, C. J., & Hall, G. (2014). Can Theories of Animal Discrimination Explain Perceptual Learning in Humans? *Psychological Bulletin*, *140*(1), 283–307. https://doi.org/10.1037/a0032765

Monahan, J. S., & Lockhead, G. R. (1977). Identification of Integral Stimuli. *J. Exp. Psychol. Gen.*, *106*(1), 94–110. https://doi.org/10.1037/0096-3445.106.1.94

Moore, R. M., & Johnston, R. A. (2013). Motivational incentives improve unfamiliar face matching accuracy. *Applied Cognitive Psychology*, *27*(6), 754–760. https://doi.org/10.1002/acp.2964

Moreno-Fernández, M. M., Salleh, N. M., & Prados, J. (2015). Evidence for feature and location learning in human visual perceptual learning. *Psicológica*, *36*, 185–204. Retrieved from https://www.uv.es/psicologica/articulos2.15/2MORENOFERNANDEZ.pdf

Mundy, M. E., Downing, P. E., Honey, R. C., Singh, K. D., Graham, K. S., & Dwyer, D. M. (2014). Brain correlates of experience-dependent changes in stimulus discrimination based on the amount and schedule of exposure. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0101011

Mundy, M. E., Honey, R. C., Downing, P. E., Wise, R. G., Graham, K. S., & Dwyer, D. M. (2009). Material-independent and material-specific activation in functional MRI after perceptual learning. *Neuroreport*, *20*(16), 1397–401. https://doi.org/10.1097/WNR.0b013e32832f81f4

Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology. Animal Behavior Processes*, *33*(2), 124–38. https://doi.org/10.1037/0097-7403.33.2.124

Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2009). Superior discrimination between similar stimuli after simultaneous exposure. *Quarterly Journal of Experimental Psychology (2006)*, *62*(1), 18–25. https://doi.org/10.1080/17470210802240614

Murphy, J., Gray, K. L. H., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin and Review*, *24*(2), 245–261. https://doi.org/10.3758/s13423-016-1131-5

National Academy of Sciences. (2014). *Identifying the Culprit: Assessing Eyewitness Identification*. https://doi.org/https://doi.org/10.17226/18891

Navarro, A., Arriola, N., & Alonso, G. (2016). Instruction-driven processing in human perceptual learning. *The Quarterly Journal of Experimental Psychology*, *69*(8), 1583–1605. https://doi.org/10.1080/17470218.2015.1088556

Nelson, J. B., & Sanjuan, M. C. (2009). Perceptual learning in a human conditioned suppression task. *International Journal of Comparative Psychology*, *22*(4), 206–220. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-18538-002&site=ehost-live

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9998–10003. https://doi.org/10.1073/pnas.1102433108

Newell, F. N., Chiroro, P., & Valentine, T. (1999). Recognizing unfamiliar faces: The effects of distinctiveness and view. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. https://doi.org/10.1080/713755813

Norell, K., Läthén, K. B., Bergström, P., Rice, A., Natu, V., & O'Toole, A. (2015). The Effect of Image Quality and Forensic Expertise in Facial Image Comparisons. *Journal of Forensic Sciences*, *60*(2), 331–340. https://doi.org/10.1111/1556-4029.12660

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology. General*, *115*(1), 39–61. https://doi.org/10.1037/0096-3445.115.1.39

Nothdurft, H. C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception & Psychophysics*, *52*(4), 355–375. https://doi.org/10.3758/BF03206697

Novick, L. R. (1988). Analogical Transfer, Problem Similarity, and Expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510–520. https://doi.org/10.1037/0278-7393.14.3.510

Novick, L. R. (1990). Representational transfer in problem solving. *Psychological Science*, *1*(2), 128–132.

Noyes, E., & Jenkins, R. (2017). Camera-to-subject distance affects face configuration and perceived identity. *Cognition*, *165*, 97–104. https://doi.org/10.1016/j.cognition.2017.05.012

Noyes, E., Phillips, P. J., & O'Toole, A. J. (2017). What is a super-recogniser? In M. Bindemann & A. M. Megreya (Eds.), *Face processing: Systems, Disorders, and Cultural Differences* (pp. 173–201). New York, NY: Nova.

O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology Public Policy and Law*, *15*(4), 315–334. https://doi.org/10.1037/a0017881

O'Toole, A. J., An, X., Dunlop, J., Nato, V., & Phillips, P. J. (2012). Comparing face recognition algorithms to humans on challenging tasks. *ACM Transactions on Applied Perception*, *9*(4), 1–13. https://doi.org/10.1145/2355598.2355599

O'Toole, A. J., Jonathon Phillips, P., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., & Dunlop, J. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Research*, *51*(1), 74–83. https://doi.org/10.1016/j.visres.2010.09.035

O'Toole, A. J., & Phillips, P. J. (2015). Evaluating Automatic Face Recognition

Systems with Human Benchmarks. In T. Valentine & J. P. Davis (Eds.), *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV.* (pp. 263–283). Chichester, England: Wiley-Blackwell. https://doi.org/10.1002/9781118469538.ch11

O'Toole, A. J., Phillips, P. J., Jiang, F., Ayyad, J., Pénard, N., & Abdi, H. (2007). Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(9), 1642–1646. https://doi.org/10.1109/TPAMI.2007.1107

O'Toole, A. J., Vetter, T., & Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing. *Vision Research*, *39*(18), 3145–3155. https://doi.org/10.1016/S0042-6989(99)00034-6

Osborne, N. K. P., Woods, S., Kieser, J., & Zajac, R. (2014). Does contextual information bias bitemark comparisons? *Science and Justice*, *54*(4), 267–273. https://doi.org/10.1016/j.scijus.2013.12.005

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*(2), 185–200. https://doi.org/10.1037/0033-295X.97.2.185

Özbek, M., & Bindemann, M. (2011). Exploring the time course of face matching: Temporal constraints impair unfamiliar face identification under temporally unconstrained viewing. *Vision Research*, *51*(19), 2145–2155. https://doi.org/10.1016/j.visres.2011.08.009

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia*, *49*(5), 1226–1235. https://doi.org/10.1016/j.neuropsychologia.2011.02.021

Palmer, S. E. (1978). Structural aspects of visual similarity. *Memory & Cognition*, *6*(2), 91–97. https://doi.org/10.3758/BF03197433

Phillips, P. J., Beveridge, J. R., Draper, B. A., Givens, G., O'Toole, A. J., Bolme, D. S., … Weimer, S. (2011). An introduction to the good, the bad, & the ugly face recognition challenge problem. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011* (pp. 346–353). https://doi.org/10.1109/FG.2011.5771424

Phillips, P. J., Jiang, F., Narvekar, A., Ayyad, J., & O'Toole, A. J. (2011). An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception*, *8*(2), 1–11. https://doi.org/10.1145/1870076.1870082

Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, *32*, 74–85. https://doi.org/https://doi.org/10.1016/j.imavis.2013.12.002

Pickel, K. L., Ross, S. J., & Truelove, R. S. (2006). Do weapons automatically capture attention? *Applied Cognitive Psychology*, *20*(7), 871–893. https://doi.org/10.1002/acp.1235

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*. https://doi.org/10.1007/s00221-011-2579-1

Podgorny, P., & Garner, W. R. (1979). Reaction time as a measure of inter- and intraobj ect visual similarity : Letters of the alphabet. *Perception & Psychophysics*, *26*(1), 37–52. https://doi.org/10.3758/BF03199860

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. https://doi.org/10.1037/0033-

295X.88.2.93
Ramon, M., Busigny, T., & Rossion, B. (2010). Impaired holistic processing of unfamiliar individual faces in acquired prosopagnosia. *Neuropsychologia*, *48*(4), 933–944. https://doi.org/10.1016/j.neuropsychologia.2009.11.014
Recio, S. A., Iliescu, A. F., Mingorance, S. P., Bergés, G. D., Hall, G., & de Brugada, I. (2016). The role of instructions in perceptual learning using complex visual stimuli. *Journal of Experimental Psychology: Animal Learning and Cognition*, *42*(4), 359–365. https://doi.org/10.1037/xan0000113
Redfern, A. S., & Benton, C. P. (2017). Expressive faces confuse identity. *I-Perception*, *8*(5). https://doi.org/10.1177/2041669517731115
Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1961–1973. https://doi.org/10.1037/xhp0000400
Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80*(3), 313–331. https://doi.org/10.1111/j.2044-8295.1989.tb02323.x
Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146–174. https://doi.org/10.1037/a0025750
Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware Person Recognition From The Body When Face Identification Fails. *Psychological Science*, *24*(11), 2235–43. https://doi.org/10.1177/0956797613492986
Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, *22*(4), 464–471. https://doi.org/10.1177/0956797611401753
Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision*, *15*(9), 15. https://doi.org/10.1167/15.9.15
Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, *51*(3), 333–342. https://doi.org/10.1016/j.visres.2010.11.014
Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology*, *3*(DEC). https://doi.org/10.3389/fpsyg.2012.00553
Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition*, *20*(2), 143–169. https://doi.org/10.1080/13506285.2012.654624
Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681. https://doi.org/10.1016/S0022-5371(75)80055-7
Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897–905. https://doi.org/10.1080/17470218.2015.1136656
Rivest, J., Moscovitch, M., & Black, S. (2009). A comparative case study of face recognition: The contribution of configural and part-based recognition systems, and their interaction. *Neuropsychologia*, *47*(13), 2798–2811. https://doi.org/10.1016/j.neuropsychologia.2009.06.004
Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79.

https://doi.org/10.1016/j.cognition.2006.02.008
Robertson, D. J., Kramer, R. S. S., & Burton, A. M. (2017). Fraudulent ID using face morphs: Experiments on human and automatic recognition. *PLoS ONE*, *12*(3). https://doi.org/10.1371/journal.pone.0173319
Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face recognition by metropolitan police super-recognisers. *PLoS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0150036
Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 629–639. https://doi.org/10.1037/0278-7393.13.4.629
Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 456–468. https://doi.org/10.1037/0278-7393.15.3.456
Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *128*(2), 274–289. https://doi.org/10.1016/j.actpsy.2008.02.003
Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica*, *132*(3), 300–312. https://doi.org/10.1016/j.actpsy.2009.08.002
Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*. https://doi.org/10.1080/13506285.2013.772929
Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: evidence from the face composite illusion. *Journal of Vision*, *8*(4), 3.1-13. https://doi.org/10.1167/8.6.157
Rossion, B., & Curran, T. (2010). Visual expertise with pictures of cars correlates with rt magnitude of the car inversion effect. *Perception*, *39*(2), 173–183. https://doi.org/10.1068/p6270
Rossman, I. L., & Goss, A. E. (1951). The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *Journal of Experimental Psychology*, *42*(3), 173–182. https://doi.org/10.1037/h0054210
Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: No special role for surface reflectance processing. *Neuropsychologia*, *50*(2), 334–340. https://doi.org/10.1016/j.neuropsychologia.2011.12.004
Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin and Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252
Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science*, *36*(6), 1019–1050. https://doi.org/10.1111/j.1551-6709.2012.01250.x
Saumier, D., Arguin, M., & Lassonde, M. (2001). Prosopagnosia: A case study involving problems in processing configural information. In *Brain and Cognition* (Vol. 46, pp. 255–259). https://doi.org/10.1006/brcg.2000.1271
Schwaninger, A., Lobmaier, J. S., Wallraven, C., & Collishaw, S. (2009). Two routes to face perception: Evidence from psychophysics and computational modeling. *Cognitive Science*, *33*(8), 1413–1440. https://doi.org/10.1111/j.1551-

6709.2009.01059.x
Schwarzer, G., & Massaro, D. W. (2001). Modeling Face Identification Processing in Children and Adults. *Journal of Experimental Child Psychology*, *79*(2), 139–161. https://doi.org/10.1006/jecp.2000.2574
Schwarzlose, R. F. (2005). Separate Face and Body Selectivity on the Fusiform Gyrus. *Journal of Neuroscience*, *25*(47), 11055–11059. https://doi.org/10.1523/JNEUROSCI.2621-05.2005
Searston, R. A., & Tangen, R. A. (2015). Putting Bias Into Context : The Role of Familiarity in Identification. *Law and Human Behavior*, 1–15. https://doi.org/10.1037/lhb0000154
Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, *75*(2), 221–242. https://doi.org/10.1111/j.2044-8295.1984.tb01895.x
Sergent, J., & Takane, Y. (1987). Structures in two-choice reaction-time data. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037//0096-1523.13.2.300
Sheehan, M. J., & Nachman, M. W. (2014). Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nature Communications*, *5*, 4800. https://doi.org/10.1038/ncomms5800
Smalarz, L., Madon, S., Yang, Y., Guyll, M., & Buck, S. (2016). The perfect match: Do criminal stereotypes bias forensic evidence analysis? *Law and Human Behavior*, *40*(4), 420–429. https://doi.org/10.1037/lhb0000190
Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*(1), 3–21. https://doi.org/10.1037//0033-295X.99.1.3
Smith, L. A., & Gentner, D. (2014). The Role of Difference-Detection in Learning Contrastive Categories. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1473–1478).
Snowdon, G. (2009). Stamp duty. Retrieved September 27, 2017, from https://www.theguardian.com/money/2009/jun/27/immigration-officer-heathrow
Solomon, C. J., Gibson, S. J., & Mist, J. J. (2013). Interactive evolutionary generation of facial composites for locating suspects in criminal investigations. *Applied Soft Computing Journal*, *13*(7), 3298–3306. https://doi.org/10.1016/j.asoc.2013.02.010
Sporer, S. L., Trinkl, B., & Guberova, E. (2007). Matching faces: Differences in processing speed of out-group faces by different ethnic groups. *Journal of Cross-Cultural Psychology*, *38*(4), 398–412. https://doi.org/10.1177/0022022107302310
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704
Stanovich, K. E. (1979). Studies of letter identification using qualitative error analysis: effects of speed stress, tachistoscopic presentation, and word context. *Journal of Experimental Psychology. Human Perception and Performance*, *5*(4), 713–733. https://doi.org/10.1037/0096-1523.5.4.713
Strathie, A., & McNeill, A. (2016). Facial wipes don’t wash: Facial image comparison by video superimposition reduces the accuracy of face matching decisions. *Applied Cognitive Psychology*, *30*(4), 504–513. https://doi.org/10.1002/acp.3218
Strathie, A., Mcneill, A., & White, D. (2012). In the Dock: Chimeric Image Composites Reduce Identification Accuracy. *Applied Cognitive Psychology*, *26*(1), 140–148. https://doi.org/10.1002/acp.1806

Suchow, J. (2018). Haven't we met before? On doppelgängers and perception. *Aeon*. Retrieved from https://aeon.co/ideas/havent-we-met-before-on-doppelgangers-and-perception

Sunday, M. A., Richler, J. J., & Gauthier, I. (2017). Limited evidence of individual differences in holistic processing in different versions of the part-whole paradigm. *Attention, Perception, and Psychophysics*, *79*(5), 1453–1465. https://doi.org/10.3758/s13414-017-1311-z

Susilo, T., Germine, L., & Duchaine, B. (2013). Face recognition ability matures late: Evidence from individual differences in young adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1212–1217. https://doi.org/10.1037/a0033469

Susilo, T., McKone, E., Dennett, H., Darke, H., Palermo, R., Hall, A., … Rhodes, G. (2010). Face recognition impairments despite normal holistic processing and face space coding: Evidence from a case of developmental prosopagnosia. *Cognitive Neuropsychology*, *27*(8), 636–664. https://doi.org/10.1080/02643294.2011.613372

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16*(2), 145–151. https://doi.org/10.1111/j.0956-7976.2005.00795.x

Tanaka, J. W., & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225–245. https://doi.org/10.1080/14640749308401045

Tanaka, J. W., & Gauthier, I. (1997). Expertise in Object and Face Recognition. *Psychology of Learning and Motivation - Advances in Research and Theory*, *36*(C), 83–125. https://doi.org/10.1016/S0079-7421(08)60282-0

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

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective & Behavioral Neuroscience*, *9*(1), 122–131. https://doi.org/10.3758/CABN.9.1.122

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: a review of the literature. *Quarterly Journal of Experimental Psychology (2006)*, *0218*(August 2016), 1–37. https://doi.org/10.1080/17470218.2016.1146780

Tanzer, M., Freud, E., Ganel, T., & Avidan, G. (2013). General holistic impairment in congenital prosopagnosia: Evidence from Garner's speeded-classification task. *Cognitive Neuropsychology*, *30*(6), 429–445. https://doi.org/10.1080/02643294.2013.873715

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition*, *67*(1–2), 73–110. https://doi.org/10.1016/S0010-0277(98)00023-7

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*(2), 233–282. https://doi.org/10.1016/0010-0285(89)90009-1

The Comptroller and Auditor General. (2013). *The Border Force: securing the border*.

The end of the CCTV era? (2015). Retrieved August 28, 2017, from http://www.bbc.co.uk/news/magazine-30793614

The Innocence Project. (n.d.). Eyewitness Misidentification. Retrieved May 5, 2017, from https://www.innocenceproject.org/causes/eyewitness-misidentification/

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*(5), 42–50. https://doi.org/10.1109/38.946630

Towler, A., Kemp, R. I., & White, D. (2017). *Unfamiliar face matching systems in applied settings*. *Face Processing: Systems, Disorders and Cultural Differences*.

Towler, A., White, D., & Kemp, R. I. (2014). Evaluating training methods for facial image comparison: The face shape strategy does not work. *Perception*, *43*(2–3), 214–218. https://doi.org/10.1068/p7676

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47–58. https://doi.org/10.1037/xap0000108

Toynbee, P. (2016, October 11). Take back control? Our Border Force is in no fit state to do its job. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2016/oct/11/take-back-control-border-force-lax-security

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, *36*(12), 1761–1771. https://doi.org/10.1016/0042-6989(95)00230-8

Turano, M. T., Marzi, T., & Viggiano, M. P. (2016). Individual differences in face processing captured by ERPs. *International Journal of Psychophysiology*, *101*, 1–8. https://doi.org/10.1016/j.ijpsycho.2015.12.009

Tversky, A. (1977). Features of Similarity. *American Psychological Association*, *84*(4), 327–352. https://doi.org/10.1037/h0021465

Tversky, A., & Gati, I. (1978). Studies of Similarity. *Cognition and Categorization*, 79–98.

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, *89*(2), 123–154. https://doi.org/10.1037/0033-295X.89.2.123

Urcuioli, P. J., & Lionello-DeNolf, K. M. (2005). The role of common reinforced comparison responses in acquired sample equivalence. *Behavioural Processes*, *69*(2), 207–22. https://doi.org/10.1016/j.beproc.2005.02.005

Urcuioli, P. J., Zentall, T. R., Jackson-Smith, P., & Steirn, J. N. (1989). Evidence for common coding in many-to-one matching: Retention, intertrial interference, and transfer. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*(3), 264–273. https://doi.org/10.1037/0097-7403.15.3.264

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Valentine, T., & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica*, *61*(3), 259–273. https://doi.org/10.1016/0001-6918(86)90085-5

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition*, *16*(6), 556–566. https://doi.org/10.3758/BF03197057

Valentine, T., & Endo, M. (1992). Towards an Exemplar Model of Face Processing: The Effects of Race and Distinctiveness. *The Quarterly Journal of Experimental Psychology Section A*, *44*(4), 671–703. https://doi.org/10.1080/14640749208401305

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology*, *69*(10), 1996–2019. https://doi.org/10.1080/17470218.2014.990392

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefèvre, P. (2010). Face inversion impairs holistic perception: evidence from gaze-contingent stimulation. *Journal of Vision*, *10*(5), 10. https://doi.org/10.1167/10.5.10

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2016). General and specific factors in the processing of faces. *Vision Research*, 1–11. https://doi.org/10.1016/j.visres.2016.12.014

Vokey, J. R., & Hockley, W. E. (2012). Unmasking a shady mirror effect : Recognition of normal versus obscured faces. *Quarterly Journal of Experimental Psychology (2006)*, *65*(April 2014), 37–41. https://doi.org/10.1080/17470218.2011.628399

Wallis, G. (2013). Toward a unified model of face and object recognition in the human visual system. *Frontiers in Psychology*, *4:497*(AUG), 1–25. https://doi.org/10.3389/fpsyg.2013.00497

Wallraven, C., Schwaninger, A., Schuhmacher, S., & Bülthoff, H. H. (2002). View-Based Recognition of Faces in Man and Machine : Re-visiting Inter-extra-Ortho. *Lecture Notes in Computer Science*, *2525*, 651–660. https://doi.org/10.1007/3-540-36181-2_65

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. https://doi.org/10.1177/0956797611420575

Wang, T., Lavis, Y., Hall, G., & Mitchell, C. J. (2012). Location and salience of unique features in human perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes*. https://doi.org/10.1037/a0029733

Wang, T., & Mitchell, C. J. (2011). Attention and relative novelty in human perceptual learning. *Journal of Experimental Psychology. Animal Behavior Processes*, *37*(4), 436–45. https://doi.org/10.1037/a0023104

Ward-Robinson, J., & Hall, G. (1999). The role of mediated conditioning in acquired equivalence. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *52*(4), 335–50. https://doi.org/10.1080/713932712

Weibert, K., & Andrews, T. J. (2015). Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces. *Neuropsychologia*, *75*, 588–596. https://doi.org/10.1016/j.neuropsychologia.2015.07.015

White, D., Burton, A. M., Kemp, R. I., & Jenkins, R. (2013). Crowd effects in unfamiliar face matching. *Applied Cognitive Psychology*, *27*(6), 769–777. https://doi.org/10.1002/acp.2971

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of automatic face recognition software. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0139827

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin {&} Review*, *21*(1), 100–106. https://doi.org/10.3758/s13423-013-0475-3

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0103510

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings. Biological Sciences / The Royal Society*, *282*(1814), 20151292. https://doi.org/10.1098/rspb.2015.1292

Wickham, L. H. V., Morris, P. E., & Fritz, C. O. (2000). Facial distinctiveness : Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology*, *91*(1), 99–123. https://doi.org/10.1348/000712600161709

Wiese, H., Komes, J., & Schweinberger, S. R. (2013). Ageing faces in ageing minds: A review on the own-agebias in face recognition. *Visual Cognition*, *21*(9–10), 13371363. https://doi.org/http://dx.doi.org/10.1080/13506285.2013.823139#2013 Taylor & FrancisDownloaded by [Goldsmiths, University of London] at 04:26 10 September 2017

Wilkinson, C., & Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science and Justice*, *49*(3), 191–196. https://doi.org/10.1016/j.scijus.2008.10.011

Wilmer, J. B. (2017). Individual Differences in Face Recognition: A Decade of Discovery. *Current Directions in Psychological Science*, *26*(3), 225–230. https://doi.org/10.1177/0963721417710693

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The Cross-Race effect and eyewitness identification: How to improve recognition and reduce decision errors in eyewitness situations. *Social Issues and Policy Review*, *7*(1), 83–113. https://doi.org/10.1111/j.1751-2409.2012.01044.x

Wirth, B. E., & Carbon, C.-C. (2017). An easy game for frauds? Effects of professional experience and time pressure on passport-matching performance. *Journal of Experimental Psychology: Applied*, *23*(2), 138–157. https://doi.org/10.1037/xap0000114

Wisniewski, E. J., & Medin, D. L. (1994). On the Interaction of Theory and Data in Concept Learning. *Cognitive Science*, *18*(2), 221–281. https://doi.org/10.1207/s15516709cog1802_2

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. V. (1979). On training people to recognize faces. *Ergonomics*, *22*(3), 333–343. https://doi.org/10.1080/00140137908924617

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, *15*(8), 1234–1242. https://doi.org/10.1093/cercor/bhi006

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Young, A., Hellawell, D., & Hay, D. (1987). Configural information in face perception. *Perception*, *16*(July), 747–759. https://doi.org/10.1068/p160747

Yovel, G., Halsband, K., Pelleg, M., Farkash, N., Gal, B., & Goshen-Gottstein, Y. (2012). Can massive but passive exposure to faces contribute to face recognition abilities? *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 285–289. https://doi.org/10.1037/a0027077

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron*, *44*(5), 889–898. https://doi.org/10.1016/j.neuron.2004.11.018

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00562

Zhao, Y., Zhen, Z., Liu, X., Song, Y., & Liu, J. (2018). The neural network for face recognition: Insights from an fMRI study on developmental prosopagnosia. *NeuroImage*, *169*, 151–161. https://doi.org/10.1016/j.neuroimage.2017.12.023

Zimmermann, F. G. S., & Eimer, M. (2013). Face learning and the emergence of view-independent face recognition: An event-related brain potential study.

Neuropsychologia, *51*(7), 1320–1329.
https://doi.org/10.1016/j.neuropsychologia.2013.03.028