Training methods for facial image comparison: A Literature Review

Anne P. Hillstrom, Ph.D.

James Sauer, Ph.D.

and

Lorraine Hope, Ph.D.

University of Portsmouth, Department of Psychology

March 2011

This report was commissioned by and delivered to the Home Office Scientific Development Branch.

Contents

1. **Scope of the Review**

This literature review was commissioned to explore the psychological literature relating to facial image comparison with a particular emphasis on whether individuals can be trained to improve performance on this task. Surprisingly few studies have addressed this question directly. As a consequence, this review has been extended to cover training of face recognition and training of different kinds of perceptual comparisons where we are of the opinion that the methodologies or findings of such studies are informative.

The majority of studies of face processing have examined face recognition, which relies heavily on memory. This may be memory for a face that was learned recently (e.g. minutes or hours previously) or for a face learned longer ago, perhaps after many exposures (e.g. friends, family members, celebrities). Successful face recognition, irrespective of the type of face, relies on the ability to retrieve the to-be-recognised face from long-term memory. This memory is then compared to the physically present image to reach a recognition decision.

In contrast, in face matching task two physical representations of a face (live, photographs, movies) are compared and so long-term memory is not involved. Because the comparison is between two present stimuli rather than between a present stimulus and a memory, one might expect that face matching, even if not an easy task, would be easier to do and easier to learn than face recognition. In support of this, there is evidence that judgment tasks where a presented stimulus must be judged by a remembered standard are generally more cognitively demanding than judgments that require comparing two presented stimuli (Davies & Parasuraman, 1982; Parasuraman & Davies, 1977; Warm and Dember, 1998).

Is there enough overlap between face recognition and matching that it is useful to look at the literature recognition? No study has directly compared face recognition and face matching, so we turn to research in which people decided whether two non-face stimuli were the same or different. In these studies, accuracy of comparison is not always better when the comparator is present than when it is remembered. Further, all perceptual factors that were found to affect comparisons of simultaneously presented objects also affected comparisons of successively presented objects in qualitatively the same way. Those studies involved judgments about colour (Newhall, Burnham & Clark, 1957; Romero, Hita & Del Barco, 1986), and shape (Larsen, McIlhagga & Bundesen, 1999; Lawson, Bülthoff & Dumbell, 2003; Quinlan, 1995). Although one must be cautious in generalising from studies of object processing to studies of face processing (see, e.g., section comparing face processing to object processing), from these kinds of studies there is no evidence to suggest that there are qualitative differences in the perceptual aspects of how recognition and matching are done. As a result, this review will include studies of face recognition skill as well as face matching skill.

The distinction between face recognition involving memory and face matching not involving memory is clouded in many recognition studies which require observers to decide which of many presented faces matches a remembered face (e.g., eye-witness studies). And of course there are other forensic face-matching tasks that will require comparison to both presented and remembered comparators (e.g., deciding whether any person in a video showing a crowd is the target person). For this reason, too, we choose to include studies of face recognition as well as face matching in our review.

1.1 Organisation of the review

In order to develop training of any task, one must understand how the task is done, how flexible people are at doing the task in different ways (Sections 2 and 3), what stimulus, cognitive, and environmental factors affect how the task is done (Section 4), and how the skill varies in the human population (Section 5). This review covers all these facets of face processing. In addition, because there are so few studies of face recognition and face matching in particular (Section 6), we discuss how perceptual expertise is understood more generally and how people train perceptual skill in tasks that do not use face stimuli (Section 7). Our summary recommendations follow (Section 8). We start first with a general description of face recognition and matching in the forensic domain.

2. Forensic Face Recognition and Matching Behaviour

Eyewitness identification is a prime example of an applied face recognition task. After viewing a crime, a witness will often be asked to view a lineup (usually containing a police suspect a number of known-to-be-innocent ‘foils’), and decide if the perpetrator they saw committing the crime is present. That is, witnesses are essentially asked if they recognise anyone in the lineup as the offender. Unfortunately, the unreliability of eyewitness identifications, where witnesses have to rely on their memory of a face, is well documented (Wells & Olson, 2003). Poor identification accuracy has been observed in both laboratory and archival studies. For instance, in their analyses of police archival data Wright and McDaid (1996) noted that the mistaken selection of an innocent person from the lineup was about 20%. It is generally acknowledged that memory for faces can be error-prone, particularly when encoding conditions prevent the formation of strong, clear memories, for instance, when exposure duration is brief (Memon, Hope & Bull, 1993), when the perpetrator is disguised, or when a weapon is involved (Steblay, 1992).

2.1 Recognising and matching previously unknown faces

Forensic facial image comparison tasks vary between those that require people to decide whether a person in CCTV footage matches a photo of a suspect (looking for a match) and those that require people to decide whether a live person matches the photo on their passport (looking for a mismatch). Sometimes the person in the images is known to the observer (as when a police officer is searching footage for a suspect with whom he has personal experience) and other times the person is unknown to the observer.

Early research on face matching ability adopted Benton's neuropsychological tool (the Benton Face Recognition Test; Benton, 1980). In this test, which is typically used to measure particular forms of neuropsychological dysfunction, participants are required to identify which of a set of faces matches a particular target face. At the outset, the target face is matched to an identical photograph from a set of six alternatives. As the task progresses, changes to the lighting and viewpoint make the task rather more challenging. When people who show no signs of pre-existing cognitive dysfunction attempt the task, the error rate approaches 20%, suggesting that simply identifying and matching a target face can pose difficulties even without a memorial component.

The difficulty of recognition of unfamiliar faces has also been documented in laboratory-based studies (Hill & Bruce, 1996). In one of the early studies on spontaneous identifications based on prior exposure, Logie, Baddeley and Woodhead (1987) examined the ability of the general public to identify a live target in a town centre from a previously presented photograph. The photograph had been published in a local newspaper. Despite details on the precise location of the target, the spontaneous detection (i.e. identification) rate for the general public was very low and this was coupled with a high false recognition rate (i.e. false identifications of other 'innocent' passers-by).

Similar low recognition accuracy has also been documented in dynamic interactions where the target face is continually available to the witness. Kemp, Towell and Pike (1987) conducted a field study to examine whether credit cards bearing a photograph of the cardholder might serve to reduce credit card fraud. Including a photograph of the legal cardholder on a credit card (or indeed, other identity document) would seem to be a relatively foolproof method of ensuring the card is only used by the person entitled to use it. In their study, shoppers presented a credit card bearing a photograph of themselves to pay for half the transactions while for other transactions they presented a card bearing the photograph of another individual. When the photograph was of someone other than the shopper, it sometimes depicted an image of a person judged to resemble the shopper in appearance (a 'matched' foil). For other transactions, the photo depicted a person judged to be dissimilar in appearance to the shopper appeared on the card. In all conditions, the photographs were of a uniformly high quality and were no more than six weeks old.

Experienced checkout cashiers were required to either accept or decline the card depending on their verification of the cardholder's identity, and rate their confidence that the photograph appearing on the card was, in fact, that of the shopper. More than 50% of the fraudulent cards were accepted by the cashiers – despite the fact that cashiers were aware that a study was underway and acknowledged that they both spent longer examining cards and had been more cautious than usual. When the photograph resembled the shopper, only 36% of the cashiers correctly declined the card. Despite these high error rates, behavioural observations suggested that the cashiers had spent some time during the transactions deliberately comparing the appearance of the photograph and the shopper.

High error rates in the ability to match a target from video footage have also been documented. Typically, it has been assumed that difficulties in identifying faces from video recordings are largely due to the frequently poor quality nature of the recording, and that were highly quality recordings available such difficulties would not arise. While it is true that CCTV images may be of poor quality for a number of technical reasons (such as unsuitable lighting conditions, low image resolution, intermittent image sampling etc.), the assumption that this alone underpins low accuracy rates in face matching from CCTV has been challenged by research findings.

Bruce, Henderson, Greenwood, Hancock, Burton and Miller (1999) examined how well people were able to match faces extracted from a high-quality video-recording against high quality photographic images. The results revealed that overall accuracy was relatively poor (averaging only 70% across trials) even under these optimal conditions (i.e. identical photograph matching with high quality target images). Performance on the matching task was further degraded when the target expression or viewpoint was altered. Furthermore, the use of colour target images (as opposed to black-and-white images) did not appear to lend any particular advantage (or disadvantage) to performance on the matching task. Henderson, Bruce and Burton (2001) extended these findings. In a series of five studies they found that the ability to match a target appearing on video footage to another image of that same target was highly error-prone. Similarly, Davies and Thasen (2000) conducted two studies examining matching ability from both face and whole body CCTV images. In the first study, they examined identification ability from CCTV recordings taken in a large public space (a car park). Accuracy rates for matching were in the region of 30% despite the fact that participants had the opportunity to consult a constant still frame of the target. Davies and Thasen (2000) explain this low accuracy rate with reference to the change in perspective between the CCTV image and the photograph of the target. The CCTV images of the target were recorded from a height of six metres above ground level while the photographs were taken at eye-level. In the second study, the CCTV images were taken at close range to the target (such as might be produced by a surveillance camera in a bank or pay-point). Thus, participants saw high-quality full-face colour images of the target. However, matching results remained poor – despite the optimal conditions - only 56% of participants correctly matched the CCTV image to the target photograph.

The research results reviewed in this section indicate – with some consistency – that our ability to identify and match an unfamiliar face to reference image (either presented simultaneously or stored in memory) is surprisingly poor.

2.2 Recognising and matching previously known or familiar faces

In contrast, matching accuracy for known or familiar faces can be very accurate – even when the target images are of poor quality. To examine the impact of familiarity on face recognition, Burton, Wilson, Cowan and Bruce (1999) showed study participants surveillance video footage of a target (a lecturer) who would be known to some participants but not others. Three groups of participants were recruited. Two of the groups comprised students who were either familiar or unfamiliar with the target. The third group comprised police officers who were unfamiliar with the target but who were experienced in making identification judgements (average of 13.5 years of service). After being shown the video footage, participants were asked to identify the target from a set of high quality photographs. Results indicated a marked advantage for people who were personally familiar with the target – 73% of the poor quality image targets were recognised *when they were familiar.* Regardless of whether they were students or police officers, people who were unfamiliar with the target performed very poorly on the identification task. In a second study, Burton et al. (1999) explored whether this familiarity effect was due to the recognition of factors such as target gait or body shape. Participants were shown video surveillance clips of a familiar target. The clips were edited such that the body, face or gait of the familiar target was obscured. Results suggested that the advantage for familiarity was largely due to recognition of the face, rather than the recognition of other cues such as gait and body shape as identification accuracy was significantly worse when the face was obscured.

Bruce, Henderson, Newman and Burton (2001) extended this research and, in a series of studies exploring the role of familiarity, found that participants were able to correctly verify (or reject) a familiar target with a high degree of accuracy (over 90%) despite the use of poor quality video images. When participants were unfamiliar with the targets, the accuracy rate was significantly lower (56%). Subsequent experiments revealed that brief periods of exposure to the target do not necessarily generate sufficient familiarity to improve the recognition or matching of unfamiliar faces – unless some ‘deep’ or social processing has taken place (i.e. discussing the faces with another person).

2.3 Conclusions

In forensic face matching, faces will often be unfamiliar, and evidence to date suggested that accuracy will be poor. In contrast, in the cases where faces are familiar, recognition will generally be much better. In fact, recognition and matching of familiar faces is often so accurate that it may not be important to try to train it.

3. How People Recognize and Match

Psychological theories relevant to visual recognition in general and face recognition in particular all include a role for perception and decision making. The perceptual part of face matching is the accrual of evidence from the sensory system as to what the faces look like. As mental representations do not match physical representations of seen images, the accrual of evidence is selective, with some stimulus aspects being registered and other stimulus aspects being ignored. The decisional part is the way the evidence from the two faces is compared. Theories vary considerably in their emphasis on the perceptual (see Recognition by Components, Biederman, 1987; viewpoint-dependent, Tarr, Williams, Hayward & Gauthier, 1998; dual route, Hummel & Stankiewicz, 1996) or decisional aspects (see Signal Detection Theory, Green & Swets, 1966; General Recognition Theory, Ashby & Townsend, 1986; Slow/Fast Guessing Theory, Baranski & Petrusic, 2003) of recognition. We will discuss those aspects of recognition and matching separately.

3.1 Stimulus aspects of face processing

Nameable shape features of faces. Some research on face processing has aimed to identify what stimulus features underlie recognition or matching of faces. Most of these studies have taken photographs of people and, using image manipulation, changed one kind of feature while leaving all others the same. If the resulting faces are more difficult to recognise or match, the assumption is that that feature is relevant to face recognition.

The eyes are generally thought to be the most important feature for recognising faces, followed by the mouth and the nose (Walker-Smith, 1978; Haig, 1986; Fraser, Craig & Parker, 1990). The importance of these *internal features* is a replicable result for familiar faces (personally known or famous people), but when identifying unfamiliar faces observers are more likely to use *external features*, including hair and head shape (Bonner, Burton & Bruce, 2003; Bruce et al., 1999; Ellis, Shepherd & Davies, 1979; Young, Hay, McWeeny, Fude & Ellis, 1985). A more recent study of recognition of famous faces suggests that eyebrows should also be considered an important internal feature, perhaps even more important than the eyes (Sadr, Jarudi & Sinha (2003).

Jarudi and Sinha (2011) pointed out that the studies showing the importance of internal features for familiar faces have used conditions (large, high quality photos of faces) that are more representative of what we see when we are close to faces. They degraded images of whole faces, internal features alone, or external features alone to study the effect of distance on recognition of famous faces. Recognition of famous faces based on external features dropped in accuracy with increasing degradation (and so increasing distance) more gradually than recognition based on internal features. In fact, after a certain amount of degradation, recognition based on external features was more accurate than recognition based on internal features. Even at highest resolutions, recognition based on internal features was far less accurate than recognition based on the whole face, suggesting that external features are not ignored when recognising nearby faces.

Surface features. The kinds of features discussed above have been the object of study because they are nameable features for virtually everybody, it has been recognised that features that the brain uses in recognition may be qualitatively different than topologically chosen features. One line of research moving away from the elementary shape features of eyes, nose and mouth looks at the importance of surface features of the face. Surface features, such as reflectance (the amount of light reflected at each point in the face) and pigmentation, influence recognition for faces, perhaps more so for male faces than female faces (Bruce & Langton 1994; Hancock, Burton & Bruce, 1996; O'Toole, Vetter & Blanz, 1999).

Russell, Sinha, Biederman and Nederhouser (2006) extended this by studying normal and contrast reversed colour photos (effectively, colour negatives of normal colour photos) of unfamiliar faces. Shape-based features can still be extracted from these images, but surface features are changed radically. The researchers presented a pair of photos that varied only in spatial features (shape and arrangement of features) or surface features (pigmentation across the skin surface) or both, and asked observers to decide which of the two photos matched one presented less than 2 seconds before. Recognition of faces varying in spatial or surface features alone was less accurate than recognition of faces varying in both, even when the photos were not contrast reversed. This suggests both kinds of features contribute to face processing.

Gilad, Meng and Sinha (2009) followed this up by creating photos of famous faces that were negatively contrasted in most regions of the face but were normally contrasted in either the eye region or the mouth region. They compared performance at recognising these *contrast chimeric* faces to performance at recognising images of the same faces in full negative contrast, images of just the eyes or mouth in normal contrast, and image of the full face in normal contrast. Performance at recognising contrast chimeric faces focused on the eye region was substantially and significantly better than performance at recognising negatively contrasted faces and recognising the eyes alone. Contrast chimeric faces focused on the mouth region were not particularly well recognised. For this reason, they claimed that the surface features around the eyes are particularly important to face recognition.

The surface features studied can all be thought of as the relative luminance of different regions of the image of the face. (Think of this as the relative brightness of, say eye, eyelid and eyebrow). Although lighting changes, tanning, and sweating can make faces look different because of changes to the luminance distribution at different points around the face, the ordering of the luminance values of different regions of the face stays the same over these environmental changes. In contrast, it changes radically in luminance contrast. The research in this section shows that face processing is sensitive to relative luminance of regions across the face, with the region around the eyes being particularly important.

Configuration. *Another* line of research looking for non-elementary features influencing face processing investigated the importance of spatial relationships between the nameable features (e.g., distances between eyes, nose, and mouth). The earliest studies of the effect of configuration showed that people were worse at recognising faces when asked to choose based on the presentation of a scrambled set of face features than when asked to choose based on an image showing the same featural parts but in correct configuration (Tanaka & Farah, 1993).

Other studies have maintained *first order spatial relations (i.e.,* the coarsely similar arrangement of features that is seen in all faces), but have changed *second order spatial relations* by introducing slight distance differences like those that occur naturally (e.g., widely spaced versus narrowly spaced eyes). This, too, affected recognition of normally orientated faces (Haig, 1984; Hosie, Ellis, & Haig, 1988; Kemp, McManus, & Pigott, 1990; Leder & Bruce, 2000; Leder, Candrian, Huber & Bruce, 2001), particularly for familiar rather than unfamiliar faces (Brooks & Kemp, 2007). These studies generally manipulate the spatial relations of the internal features. A study that manipulated the spatial relations of external features (i.e., the ears) found that people were completely insensitive to that spatial relationship. Thus, spatial relations between internal features is more important than spatial relations between external features. This effect of spacing of features has been replicated in matching tasks as well (e.g., Freire, Lee & Symons, 2000; Le Grand, Mondloch, Maurer & Brent, 2001; Mondloch, Le Grand & Maurer, 2002). To further strengthen their findings, many of these studies tested matching using inverted faces that also differ only in spacing of features. It was much more difficult to discriminate inverted faces with this difference than upright faces with this difference, even though the amount of configural information was the same for upright and inverted faces. The conclusion reached is that configuration is integral to face processing, which is addressed more by upright faces than upside down faces.

Holistic processing. An overlapping body of research has suggested that recognition of faces is not based on elementary, topographically defined features or even on spacing of features, but instead is based on the face as a whole. When images of faces are divided in half, and the top half of each face is aligned with the bottom half of another face, recognition of the person represented by the top half of the composite face is inaccurate. This finding suggests that observers tend to process and recognise faces as wholes (Young, Hellawell & Hay, 1987; Hole, 1994). When the two halves of the faces are misaligned (e.g., the bottom half is shifted to the right, so that the outline of the two parts is disrupted), recognising the person in just the top half is much improved. The difference in accuracy (or response time, in some studies) has been labelled the face congruency effect. This effect has been replicated using a matching task for unfamiliar faces, where observers are asked to report whether the top half of two composite faces match, regardless of variation in whether the bottom halves match.

It is widely (but not universally) accepted that face processing has a strong holistic component. Recent research demonstrates that the holism is decisional rather than strictly perceptual (Wenger & Ingvalsen, 2002, 2003; Richler, Gauthier, Wenger & Palmeri, 2008). These studies used the composite face paradigm, and so to say face recognition is decisionally holistic means that the criterion an observer uses to decide that there is enough evidence that the top half of an aligned face matches what it should be compared to cannot help but be influenced by whether the bottom half of the face matches the bottom half of its own comparison face. This can occur even if the perceptual evidence that the decision depends on does not change, and the setting of a criterion is only partly conscious. To say that face recognition was not perceptually holistic in these studies means that (a) the evidence about the top half of an aligned face (i.e., the perceptual mental-representation that is based on the sensory information) was independent of information about the bottom half of the face.

The suggestion that face processing is entirely holistic has been challenged by recent work showing that when a group of participants were asked to complete both the composite face task and a face matching task, the correlation between the composite face effect and matching ability of observers was not significantly different from zero (Konar, Bennett & Sekuler, 2010). This suggested that face matching ability is not related to holistic processing, in that people might be good at holistic processing and poor at face matching or vice versa. However, a more recent study suggests that Konar et al.'s composite face task might have been susceptible to response biases (i.e., the tendency to say that faces matched more than they did; Richler, Cheung & Gauthier, 2011). When a different design of the composite face task was used to alleviate the potential for bias, and when a face recognition task was used instead of a face matching task, there was a correlation between performance on the composite face task and performance on face recognition (Richler et al).

Note that the main conclusion of Richler et al. was the importance of using the right test if one wants to find evidence for holistic processing. Although it was noted that one test used a face recognition task and the other used face matching task in the comparison being made, and that difference might have accounted for the difference in results, the possibility that face matching might rely less on holistic processing than face recognition was not tested.

It is important to understand that although face processing relies a lot on holistic processing, people who only have features to judge are still able to process faces to some degree (e.g., Van Belle et al., 2010; Williams, Moss & Bradshaw, 2004). Kimchi and Amishav (2010) have shown that there are two, reasonably independent routes to face processing: configural processing and featural processing. They used a face matching task for artificially created faces to show that matching was easy as long as either individual features were easy to discriminate or the configuration of features was easy to discriminate. Only when both were difficult to discriminate were the faces difficult to match. It is important to note that when the faces were judged to be the same, they were the same image of the face, not the same face depicted differently. Thus, observers could attribute any difference between the photos to differences between the faces; when forensic personnel do face matching, there is a lot of irrelevant variation in the photos that could contribute to the difficulty of comparing the phases.

A third study was more ecologically valid. People performed recognition judgments on either photos of famous faces or faces learned minutes before, where the photos at learning and test were not identical (Collishaw & Hole, 2000). Some photos in recognition testing were degraded in one of two ways. They were either blurred, which degraded feature representation but not configuration, or scrambled, which degraded configuration but not feature representation. A final set of faces was degraded in both ways. Blurring and scrambling lowered accuracy, and neither produced worse recognition than the other. Degrading both lowered accuracy to chance. Thus, some face recognition ability was retained when configural processing was impaired or featural processing was impaired, but not when both were impaired.

All conclusions drawn from studies comparing featural to configural processing are critically dependent on the assumption that if the brain processes based on features, those features are the ones that were manipulated (or that varied naturally) in the study. It is important to keep in mind that if researchers are manipulating the wrong kind of feature, then results supporting holistic processing could occur because the composite face task actually degrades the integrity of the features that are being used by the brain. For instance, studies have suggested that the brain areas that are specialised for face processing are using spatial frequencies of faces to do their processing (Yue, Tjan & Biederman, 2006) and questioned whether different bands or orientations of spatial frequencies are particularly important features for face processing (e.g., Collin, Therrien, Martin & Rainville, 2006; Goffaux & Dakin, 2010). Spatial frequency information is the periodic alternation between darkness and lightness that can be extracted from images using a mathematical technique called Fourier analysis. Any image has a broad range of spatial frequency information, with a particular spatial frequency indicating a particular number of cycles per second. There is no exact perceptual equivalent to a spatial frequency channel, but roughly speaking, high spatial frequencies contain information about quite fine-grained variation and low spatial frequencies contain information about quite gradual variation. See Figure 1 for an example of what a face would look like with the high spatial frequencies removed (labelled Low SF) or the low spatial frequencies removed (labelled High SF).

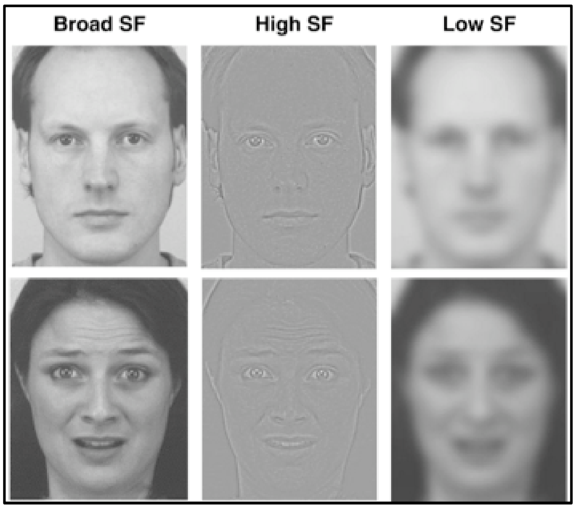

Figure 1. Image showing a face presented normally and with only high or low spatial frequency information. Figure from Vuilleumier, Armony, Driver & Dolan (2003). Nature Neuroscience, 6, 624-631.

Because composite faces preserve the representation of elementary features (eyes, nose, mouth), degraded performance on the composite face task is assumed to show that face processing is holistic or configural rather than featural. If the features the brain uses to process faces are eyes, nose, and mouth, then this is a logical conclusion. But if the features the brain uses are bands of spatial frequency (which can only be extracted across a whole image), then this conclusion is illogical, as composite faces disrupt spatial frequency bands.

Among the most important influences of spatial frequency on face recognition is that loss of high spatial frequency information (fine detail) impacts recognition of familiar and unfamiliar faces more than loss of low spatial frequency information

(Costen, Parker & Craw, 1996; Deruelle & Fagot, 2005; Fiorentini, Maffei & Sandini, 1983). Schyns and Oliva (1999) found the reverse, but that may be because they had a rather small set of faces to recognise, and so there may have been more differences between the faces to detect than can be assumed in most real-world comparisons. The difficulty in processing faces without high spatial frequency detail may be why observers attempting to recognise people at a far distance tend to rely on external features, such as clothing, more than faces (e.g., Tickner & Poulton, 1975). In assessing the importance of this literature, it is important to note that recognition of familiar faces is highly resistant to low resolution of CCTV footage (which would result in loss of high spatial frequencies) (Burton et al., 1999). Why familiarity would affect the role of low versus high spatial frequencies is unclear.

The importance of qualities of images (like spatial frequency information) that are impossible for people to perceive are particularly difficult for experimenters to study. One method that has been used to attempt to find such features uses a statistical analysis called principal components analysis to find out what varies in a large set of photos. It is beyond the scope of this report to explain how this works, but the goal of such analyses is to determine without a-priori hypotheses the physical dimensions along which images of faces vary and then to see whether these dimensions are important to the way faces are processed. The importance of dimensions is ascertained by correlating judgments about the faces with their position along the dimensions identified. Studies that have taken this approach have learned that distinctiveness and memorability of faces correspond to the most fundamental dimensions in a "face space" derived in this manner (Hancock et al., 1996; O'Toole, Abdi, Deffenbacher & Valentin, 1993), and that these two dimensions are independent. It is interesting that this technique has not resulted in dimensions that seem to correspond to the proposed features discussed so far, which adds to the uncertainty that the features that have been studied by researchers of face perception are the physical features that actually influence face processing.

3.2 The relationship between face and object processing

There has been a debate in the face perception literature about whether humans have evolved special face processing techniques and special brain areas to do that processing due to the critical importance of face information to survival in social environments. If one looks at a sample of people from the general population, there are differences in how they process faces and objects. For instance, individuating objects typically relies more on featural processing than holistic or configural processing (Tanaka & Farah, 1993; Yin, 1969). However, some claim that this difference relates more to the need to make quite subtle distinctions than to object class (Gauthier, Behrmann & Tarr, 1999; Gauthier, Tarr, Moylan, Skudlarski, Gore, & Anderson, 2000). Moreover, there is a body of research claiming that although faces are not processed differently in general from objects, because all adults have extensive practice at face perception and recognition, they can be considered experts at it. These researchers argue that the way people perceive faces is indiscriminable from the way experts make subtle discriminations among the class of similar objects related to their expertise (e.g., experienced bird watchers) (e.g., Diamond & Carey, 1986; Gauthier, Behrmann & Tarr, 1999; Gauthier & Tarr, 1997; Gauthier, Williams, Tarr & Tanaka, 1998). There are also claims that the brain imaging indicators of face processing also occur for other subtle discriminations made by experts (Gauthier, Skudlarski, Gore & Anderson, 2000; Tarr & Gauthier, 2000).

Despite the growing evidence that expert object processing is like face recognition, the claim is strongly debated in the face recognition field. There are other respected researchers (e.g., Kanwisher & Yovel, 2006; Nederhouser, Yue, Mangini & Biederman, 2007) who have argued quite cogently that there are flaws in the behavioural evidence for the claim and that brain imaging techniques that have the best spatial resolution find differences in the localisation and nature of brain functioning underlying face recognition and expert object recognition (Yue, Tjan & Biederman, 2006). Further, there are examples of persons with brain damage who have deficits in face processing while having spared visual expertise and object processing, and at least one person who is reported to have spared face processing but a deficit in object processing, even for a category of objects that he was once an expert at discriminating (Kanwisher & Yovel, 2006; Moscovitch, Winocur & Behrmann, 1997). This kind of “double dissociation” of abilities has often been used as an indicator that the two abilities are independent of one another.

Megreya and Burton (2006) examined the relationship between face processing and object processing in a series of six studies. Modest correlations between performance and test score were found for several standard tests of memory and perception (e.g. perceptual speed, visual short-term memory and figure matching ability). However, the most interesting finding indicated that performance on a face matching task was predicted by performance on an inverted faces task (i.e. when faces were presented upside down). In other words, accuracy on the face matching task correlated with accuracy on an inverted faces task such that people who performed well on one task tended to also perform well on the other. People typically perform more poorly when attempting to identify inverted faces and evidence suggests that inverted faces are processed differently than faces presented in the usual upright position. Megreya and Burton concluded that the underlying reason for the poor accuracy rates typically observed on face matching tasks when the target is unfamiliar (e.g. when trying to decide whether an individual appearing in CCTV footage matches an image of a suspect) may be due to the fact that unfamiliar faces are processed in a different way to familiar faces, perhaps even like objects are processed.

Wilhelm et al. (2010) presented a large set of face processing tasks to a broad sample of Caucasian persons and looked (using structural equation modelling) for underlying skills that would account for the variability in skill of those persons. They also included tests of memory, working memory, and object perception that did not test face processing, to determine whether the skills needed for face processing were unique to faces. They found a correlation between object perception and face perception as well as between object perception and face memory as explanations in skill variance in the population. However, they also found independence between those factors, leading them to conclude that object perception and face perception have important differences. The objects they used were houses, and if one believes that face processing is like expert object processing, the work should be repeated using people who are experts in discriminating a class of objects, to see if the skills at face discrimination and skills at perceiving that class of object might be so strongly correlated that one could conclude that they were the same. In addition, in the study that was done, no attempt was made to equate the similarity between houses with the similarity between faces, which could lead people to emphasise different aspects of perception.

To the degree that object processing and face processing are independent of one another, training on one would not benefit training on the other. Although the claim of independence is still under debate, there is enough evidence of independence that we would predict that training on subtle discriminations of objects other than faces would not effectively train face recognition.

3.3 The role of decisions in face processing

Face recognition and face matching tasks both require individuals to judge the similarity of two items. Recognition judgements assess the degree of match between a presented stimulus and an image in memory (ecphoric similarity: Tulving, 1981). Matching judgements require individuals to assess the degree of match between two physically present stimuli. Thus, before addressing specific variables relevant to face matching and face recognition, it may be helpful to outline a basic theoretical framework within which the effects of these variables can be interpreted. Two theories of recognition and perceptual discrimination provide suitable structures for this purpose.

Models of perceptual discrimination developed within the signal detection theory (SDT) framework (e.g., Green & Swets, 1966) hold that, when two stimuli are compared, this comparison generates a value which indexes the degree of match between the two items. This value falls somewhere along a continuum, with similar items generating values that fall towards the higher end of the continuum, and dissimilar items producing values (generally) falling towards the lower end of the continuum. Stimuli producing values exceeding a pre-set criterion are judged to be the same (match), while those resulting in values below the criterion are deemed to be different (mismatch). In this way, the comparison produces a binary response. When testing recognition, the same process occurs. A presented item is compared to an image of a previously viewed item stored in long term memory. This comparison generates a value[1] which falls somewhere along a familiarity continuum. Again, items producing values exceeding a pre-set criterion are judged to be previously seen (recognised), while those resulting in values below the criterion are deemed to be previously unseen. For both tasks, three points must be noted. First, a comparison generates a value which indexes the degree of match between two items. Second, this index is compared to a pre-set criterion to produce a binary

[1] This value has been referred to as a likelihood ratio or function of a likelihood ratio (e.g., Green & Swets, 1966; Stretch & Wixted, 1998), as an index of stimulus familiarity (e.g., Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996), or as a strength effect (e.g., Balakrishnan & Ratcliff, 1996).

classification (e.g., ‘match’ vs. ‘mismatch’ or ‘seen’ vs. ‘unseen’). Third, and crucially, the final decision reflects *both* the strength of the index *and* the placement of the response criterion.

Models of comparative judgement formulated within the sequential sampling framework assume that discrimination/recognition judgements are again reached by comparing two stimuli (Baranski & Petrusic, 2003). According to these ‘accumulator’ models, discrete units of information pertaining to presented stimuli are sampled and accumulated over time. These units can either favour the response that the two items are the same, or that they are different (some models include a third accumulator for ‘non-diagnostic’ information). Units of information favouring each alternative are stored in separate accumulators, each of which has a pre-set criterion. When the accumulated information exceeds the pre-set criterion for one of the available response alternatives, that response is given. The key consideration for our present purpose is that, as with SDT-based models, responses will be influenced by the strength of evidence supporting a particular response (i.e., the degree of the match between two stimuli) *and* the placement of the pre-set criteria. Thus, when assessing procedures and performance in face recognition and matching tasks, it is important to consider variables that will affect the degree of match between presented stimuli (i.e., perceptual features), and factors that will affect criterion placement (and make decision makers more or less conservative). Our review will cover both these areas, before presenting methods for improving task performance.

3.4 Conclusions

Research that looks for stimulus features of faces that are used in face processing has mostly focused on nameable features (e.g., eyes, nose, mouth, hairline), surface features and configural features (e.g., the set of distances between the features). There is also evidence that processing is at least in part holistic, in the sense that the whole face is processed as a unit rather than features being processed as separate sources of information about the face. The most rational conclusion is that all of these contribute to face processing. There remains a doubt as to whether researchers have identified and manipulated the features that the brain actually uses to process faces.

In order for training of face processing to be effective, one would not want the set of images that are used in training to have homogeneity of any feature that is important to processing. That does not mean that all features must vary together in all parts of training; it just means that at some point in training, the trainee should be exposed to variance in all important features.

Likewise, because face matching requires not only extracting stimulus information about the face but also requires making a judgment about that information, training should not ignore decisional factors.

4. Factors Affecting Recognition and Matching Performance

4.1 Stimulus factors

Recognition and matching tasks both require an individual to compare two stimuli and assess the extent to which they match. Even when the two stimuli depict the same target individual, performance on matching task will depend on the extent to which the stimuli provide clear images for the comparison process. In recognition tasks, this will depend in part on the conditions under which the memory was formed (as these will determine the quality of the memory). In matching tasks, this will depend on the quality of the images (static or dynamic) available for comparison. We now consider factors that will affect the quality of the stimuli available for matching, and contribute to task performance.

Constant and transient characteristics of the person. Stable factors (such as gender or age of the target) have little or no impact on ability to correctly identify or match the target. However, there are a number of well-documented factors that can serve to either impair or enhance recognition ability. For instance, distinctive faces are far more likely to be correctly identified than non-distinctive faces. Similarly, and perhaps due to their distinctiveness, attractive faces are also more easily identified than less attractive or more typical faces. The psychological mechanisms underlying these findings are relatively straightforward. When an encoded face is distinctive or atypical, it will not only attract more attention and greater processing resources but the distinctive feature is also more likely to benefit from an enhanced representation in memory (Ryu & Chaudhuri, 2007; see Brewer et al., 2005).

Unsurprisingly, disguises usually have a negative impact on identification ability (Cutler, Penrod & Martens, 1987; but see O'Rourke Penrod, Cutler, & Stuve, 1989). Simple changes, such as covering the head, wearing glasses, growing facial hair or even altering hair style slightly can significantly impair face recognition (Narby, Cutler, & Penrod, 1996; Shapiro & Penrod, 1986).

Recent work that is not yet published has found that cropping the photos of unknown faces so that only the internal features of the face are visible improved face matching for inexperienced matchers, but with training, matchers learned to ignore hair even when it was present (Burton et al., 2010a). This was a laboratory experiment in which people were to detect mismatches. In this study there was success in training face matching but when they reran the experiment using a task simulating forensic face matching more realistically, performance was worse, with false alarms being notable.

An extensive literature has documented the identification impairment that occurs when the perpetrator is from a different race or ethnic group to the witness. Research on own-race bias (also known as cross-race bias or other-race effect) typically demonstrates that witnesses are less accurate when attempting to identify a target from another race or ethnic group than when tasked with identifying a member of their own race (see meta-analysis by Meissner & Brigham, 2001). Specifically, research documents a higher correct identification rate from target present lineups and a lower false identification rate from target absent lineups when the witness and perpetrator are from the same race. This bias has been demonstrated in both laboratory and field studies (e.g. Wright, Boyd, & Tredoux, 2001) and has been observed across various combinations of ethnic groups (e.g. whites identifying blacks, blacks identifying whites etc.). Work by Chiroro and Valentine (1995) exploring a basic contact hypothesis suggested that everyday interactions with people of different races may reduce the effect – but not consistently. Other evidence suggests that the quality, rather than the quantity of cross racial interactions may be more important in reducing own race bias (Lavrakas, Buri, & Mayzner, 1976). Interestingly, a similar pattern of results has been demonstrated for gender and age such that a match between witness and target age and gender can promote recognition accuracy (e.g. Wright & Sladden, 2003; Wright & Stroud, 2002).

Taken together, these findings suggest a somewhat preferential processing mechanism for familiar stimuli. In this vein, McClelland and Chappell (1998) have argued that own race faces may benefit from more accurate and efficient processing due to their familiarity.

The effect of differences in age between the images. Face recognition and face matching tasks involve an assessment of the degree of match between two stimuli. While the age of a face, per se, will not affect recognition or matching performance, age-related changes in appearance reduce the degree of match between two images of the same individual (or between an individual and a photograph depicting them at a younger age), and impair face recognition performance (see Shapiro & Penrod, 1986, for a meta-analysis). Read, Vokey and Hammersley (1990) found that photographs of a target face taken after a two year delay were less likely to be recognised than photographs taken nearing the time of original encoding. Aging may affects structural/physiological aspects of a face. For example, as the collagen in skin tissue becomes less elastic, the shape of facial features alters. Alternatively, aging may be associated with more superficial appearance change. For example, hair length, density and colour change with age. While hair is, in a sense, a superficial feature, research shows that the encoding of hair-related details contributes to both recognition performance, and the phenomenology of recognising unfamiliar faces (O'Donovan & Bruce, 2001; Wright & Sladden, 2003).

At present, we are unaware of research investigating age-related changes in appearance in the face matching context but, given the fundamental similarity of the underlying comparison, we can assume similar effects will emerge. However, while research supports the deleterious effects of age-related changes in appearance on face recognition performance, exact effects of aging (on degree of match and performance) are impossible to quantify.

An additional applied issue, relevant to the effects of age-related changes in appearance on face recognition and face matching performance, needs to be highlighted. Often, after a witnessing a crime, a witness will wait weeks or months before viewing a lineup. Thus, the perpetrator's appearance during the crime and their appearance in the lineup may differ markedly. As mentioned above, such changes may impair recognition performance. For this reason, the U.S. Department of Justice's Guide (1999) for lineups recommends warning witnesses that the perpetrator's appearance may have changed since the original encounter. However, research clearly demonstrates that providing this instruction does not improve correct identification rates (Charman & Wells, 2006). It does, however, increase false identification rates. Thus, while theory and research suggest that age-related changes in appearance are likely to impair face matching performance, we strongly advise against the use of any 'appearance change' warning.

The effect of stimulus dynamism on performance. Face recognition and matching studies may require individuals to match two static representations of an individual, match a static to a dynamic representation (e.g., match CCTV footage to a mug shot, or an identification photo to a live individual), or match two dynamic representations (e.g., compare a physically present individual to CCTV footage). Researchers have argued that the use of dynamic images, as opposed to static images, may improve performance on face recognition and face matching tasks. Thornton and Kourtzi (2002) argued that dynamism may benefit face matching because a) face-processing mechanisms are designed to make use of motion cues, and b) more generally, dynamic stimuli provide a richer source of information upon which decisions can be based. More specifically, researchers have argued that, compared to static stimuli, dynamic stimuli may benefit face recognition and matching performance through the formation of more robust memorial representations; by facilitating the detection, encoding and recognition of specific/diagnostic facial features and expressions; and by providing specific information relating to the three-dimensional structure of a face (e.g., Christie & Bruce, 1998; Knight & Johnston, 1997; Lander, Christie, & Bruce, 1999). However,

Stimulus dynamism in face recognition. Despite the plausibility of these suggestions, the effects of stimulus dynamism (or motion) on face recognition are unclear (O'Toole, Roark & Abdi, 2002). The literature draws a distinction between two types of motion – rigid and non-rigid motion – and the effects these types of motion have on recognition performance. Rigid motion involves changes in whole-head position or orientation, while individual features remain still. In contrast, non-rigid motion refers to movements of facial features in relation to each other (e.g., fluid facial expressions). Research suggests that, compared to the use of static images, rigid movement benefits face recognition, regardless of whether the motion is present at study (e.g., Pike, Kemp, Towell, & Phillips, 1997) or test (e.g., Schiff, Banka, & Galdi, 1987). However, Shepherd et al. (1982) found no benefit of motion. The benefits of non-rigid motion are less obvious. Valentine and Bruce (1988) found little evidence that studying dynamic, compared to static, images improves later recognition. Other researchers have found that dynamic presentation at test can improve recognition of famous faces (e.g., Knight & Johnston, 1997; Lander et al., 1999), but only when testing conditions are sub-optimal (e.g., when recognition is based on photographic negatives rather than the original photographs). Finally, other research has shown no benefit of motion (rigid or non-rigid) on the recognition of studied but previously unfamiliar faces (e.g., Christie & Bruce, 1998; Pike, Kemp, Towell, & Phillips, 2003).

Stimulus dynamism in face matching. The research literature on the benefits of dynamism in face *matching* tasks is limited. Thornton and Kourtzi (2002) tested the effects of non-rigid facial motion on face matching using a sequential matching task. In two experiments, participants were presented with an initial stimulus (either a static or video clip depiction of a face) for 540ms. This stimulus was removed, and a second stimulus was presented. The second stimulus was always a static image of a face. Given the ease of the task (accuracy rates ranged from 82-96%), the authors used response time, rather than accuracy, as a measure of performance. Findings suggest a modest improvement in performance (i.e., reduced response time) for dynamic, compared to static, stimuli when the two stimuli depict the same individual but only when stimuli are not identical (either the second stimulus showed the same individual with a different facial expression, of the second stimulus was inverted). When identical images were presented at both phases, and when the images depicted different individuals at each phase, no motion-related advantage was found. Consistent with the motion-related advantages for recognition outlined above, the benefits of face dynamism for face matching tasks appear strongest when testing conditions make the task more difficult. Easy matching tasks - where faces are clearly identical or clearly different - show no motion related benefits.

Thornton and Kourtzi’s (2002) research was the first to demonstrate reliable performance differences between dynamic and static images using non-degraded images of non-famous individuals. Thus, these findings may be relevant to face matching tasks in applied settings - where most stimuli are not degraded, and most targets are not famous. However, recent face matching research has demonstrated that matching a live person to a photo (i.e., comparing dynamic and static stimuli) is no easier than comparing two photos of the same person (Davis & Valentine, 2009; Megreya & Burton, 2008). Similar patterns are reported in research requiring participants to match video footage to static images (Bruce et al., 1999). Two additional considerations merit further caution when applying Thornton and Kourtzi's findings to task performance in applied settings. First, Thornton and Kourtzi used a sequential matching task, which rendered their task more like a recognition task than many applied matching tasks (e.g., identity check at airports). As stated before, the simultaneous presentation reduces cognitive effort, resulting in an easier matching task. Results consistently demonstrate that task difficulty affects motion-related advantages. Thus, Thornton and Kourtzi's findings may overestimate the benefits of face dynamism in applied matching settings. It remains to be seen whether using simultaneous comparisons will have the same effect as using successive comparisons. Second, the researchers measured performance in terms of response time, rather than accuracy. For this reason, participants were instructed to respond as quickly as possible. Again, this procedure differs markedly from most matching tasks in applied settings where accuracy is paramount. Research consistently demonstrates that, in virtually all recognition and perceptual discrimination tasks, information processing and task performance vary depending on whether instructions emphasise speed or accuracy (see Vickers, Burt, Smith, & Brown, 1985). As speed increases, accuracy decreases. Given the consistent empirical and theoretical support for the speed-accuracy trade-off, the emphasis placed on speed in the current research limits its applicability to more applied settings.

In sum, across face recognition and face matching tasks, results suggest that, when testing conditions are difficult, motion-related cues may enhance task performance. However, results are not consistent. Additionally, there is virtually no consideration of the effects of (mis)matched stimulus modality. Potential effects of modality match can be explained with reference to the theoretical frameworks for recognition and perceptual discrimination outlined earlier. Recognition and matching tasks both require individuals to judge the similarity of two items. Recognition judgements assess ecphoric similarity (the degree of match between a presented stimulus and an image in memory). Matching judgements require individuals to assess the degree of match between two physical present stimuli. If stimuli are presented (or, in the case of recognition, learned and tested) in identical modalities, the degree of match is more easily assessed. Thus, in the face recognition context, if stimuli are dynamic at study, the use of dynamic rather than static test stimuli may improve performance by providing additional, diagnostic information (e.g., information relating to movement was encoded at study, and can be compared with movement cues present at test) . However, if stimuli are static at study, dynamic test stimuli are less likely to a) contribute diagnostic information (as no movement cues were encoded at study), and b) improve task performance. This factor may contribute to the inconsistent benefits of dynamic test stimuli reported for the recognition of famous (typically learned from dynamic stimuli) and non-famous (typically learned from static images) faces. Similarly, face dynamism may improve face matching performance if both stimuli are dynamic, but have little effect when one stimulus is static (as movement-related information cannot be compared). This is an empirical question.

4.2 Decisional factors and biases

As previously mentioned, face recognition and face matching tasks both require individuals to assess the degree of match two stimuli. Ideally, a high degree of match will result in a positive (recognise/match) response, while a low degree of match will result in a rejection or negative response. However, environmental and social factors can influence response criterion placement, and reduce the fidelity with which the eventual response indexes the assessed degree of match. For example, in a busy environment, quick decision-making may be a priority. When prioritising speed, decision makers will set a lower response criterion (in order to reach that criterion, and generate a response, more quickly). Thus, decisions will be based on a less thorough/compelling accumulation of evidence, and errors are more likely. Alternatively, an understanding of the relative consequences of different types of response error may affect criterion placement. Here, it may be useful to consider face recognition and face matching separately. A police lineup provides a suitable context for discussing how and awareness of the consequences of different response errors in an applied face recognition task can affect criterion placement and, inadvertently, increase error rates. If a witness wrongfully identifies an innocent suspect, the likelihood that this suspect will be charged increases. Further, mistaken identifications are a leading cause of wrongful conviction in Western criminal justice systems. If a witness is aware of the serious consequences of a mistaken identification, they may set a very conservative decision criterion (i.e., they will not identify a lineup member as the culprit unless they are certain). A conservative criterion reduces the risk of mistaken identification, but increases the risk of failing to identify the guilty party if they are in the lineup. Alternatively, research demonstrates that instructions that emphasise the importance of identifying the perpetrator lead witnesses to set lenient decision criteria: increasing the chance of a mistaken identification (Wells & Olsson, 2003).

Similarly, depending on the applied face matching context, some response errors are likely to be (or be perceived as) more costly than others. For example, at an airport security checkpoint, failing to correctly detect that a passenger matches a mug shot of a known or suspected terrorist could be disastrous. Thus, decision makers may set a lenient criterion (i.e., require less evidence) for a positive (match) response. In contrast, when checking identification at a nightclub, the consequences of admitting a minor may outweigh the consequences of denying entry to an adult. Thus, in this context, decision makers may set more conservative criteria when matching faces to identification documents.

In sum, error rates in matching tasks, and the *types* of errors made, will be influenced by factors that do not affect the degree of match between to-be-matched stimuli. Thus, improving performance on face matching tasks requires an awareness of variables that influence the degree of match between stimuli *and* variables that influence response criteria.

The effects of operator mindset on decision making. The success of an investigation depends to a large extent on the ability of the investigators to evaluate information and evidence accurately (Ask, Rebelius & Granhag, 2008).

Identifying, extracting and drawing accurate inferences from evidence requires the same accuracy and relies on the ability of investigator to remain objective and open to alternative interpretations (or possibilities) when reviewing and evaluating such evidence. Ideally, the evaluation of evidence should not be affected by external or contextual factors, such as time pressure, preconceptions, emotions, beliefs about likely suspects or sequence of events. Indeed, it is generally assumed that we make ‘hard-nosed’, objective decisions and judgments as a matter of course and that, in particular, experienced individuals will not be influenced by the vagaries of contextual factors. However, the results of research challenge this assumption and suggest that the evaluation of forensic evidence can be sensitive to external influences (Ask et al., 2008; Ask & Granhag, 2007; Dror, Charlton & Péron, 2006; Dror & Rosenthal, 2008).

At the outset of an investigation, police will be guided in their search and evaluation of evidence by preliminary or working hypotheses concerning the crime. For instance, how the crime was committed, who was involved and why it occurred. These working hypotheses may not be based on available, objective evidence – quite often, evidence of that type may not be available. Rather, these hypotheses may be based on expectations or script-based causal explanations (Ask & Granhag, 2005). In other words, investigation is hypothesis driven as investigators try to piece together any available evidence to formulate the most plausible account of the crime. Constructing theories which provide a causal structure for information or evidence is frequently a spontaneous cognitive response to ambiguous problems or situations (Kahneman, Slovic & Tversky, 1982; Nisbett & Ross, 1980).

In any investigation involving the matching of particular targets to prior information evidence, it is quite likely to be the case that officers will also have access to other details about the case and may be working with quite specific hypotheses concerning the suspects. Ideally, these hypotheses will be based on the triangulation of evidence from other sources, such as victims, witnesses, informers and so on. However, human cognitive processes are such that preconceptions, expectations, pre-existing schemas for particular crime types and other biasing tendencies may (unhelpfully) influence the evidence-evaluation process. These tendencies are fundamental to human information processing and have been documented throughout the psychological literature. The following section will examine several of these tendencies as they relate to the evaluation of forensic evidence.

Confirmation bias. Confirmation bias refers “to unwitting selectivity in the acquisition and use of evidence” (Nickerson, 1998, p.175). In other words, the tendency to favour information or evidence which confirms an initial or existing belief while avoiding or rejecting disconfirming evidence (Koriat, Lichtenstein, & Fischoff, 1980). In a face matching task, this may lead decision makers to set overly lenient (or overly conservative) criteria, and increase the risk of error, depending on their expectations.

Research by Darley and Gross (1983) provides a good illustration of the impact of prior expectations and preconceptions on ability to objectively evaluate actual evidence. In their study, two groups of people viewed a videotape of a child taking an academic test (the same child was viewed by both groups). One group was led to believe that the child came from a high socioeconomic background while it was suggested to the other group that the child's socioeconomic background was low. Both groups were asked to rate the academic ability of the child based on what they had seen of their performance in the video alone. Participants in the former group (high SES) rated the child's abilities higher than those who were led to believe that the child came from a low socioeconomic background. Darley and Gross argued that participants formed an advance hypothesis about the child's academic abilities on the basic of socioeconomic background and then unwittingly sought out evidence in the video recording that was consistent with this hypothesis.

Research consistently demonstrates that we prefer information biased towards our pre-existing beliefs or expectations (Hope, Memon, & McGeorge, 2004; Jonas, Shulz-Hardt, Frey, & Thelen, 2001) and attitudes, stereotypes and preferences (Lundgren & Prislin, 1998). Confirmation bias has been demonstrated relating to stereotypes about ethnicity (Duncan, 1976), clinical outcome (Swann, Giuliano, & Wegner, 1982); education (Foster, Schmidt & Sabatino, 1976) and gender (Oakhill, Garnham & Reynolds, 2005). The confirmation bias has also been robustly demonstrated across decision making in diverse domains from formal problem solving (Wason, 1968) to social interactions (Snyder & Swann, 1978) and across real life domains including public policy rationalisation (Tuchman, 1984), medical decision making (Elstein, Schulman & Sprafka, 1978) and judicial reasoning (Hope, Memon & McGeorge, 2004; Kalven & Zeisel. 1966; Pennington & Hastie, 1986, 1988, 1993).

More recently, research has examined how confirmation bias impacts on the evaluation of forensic evidence by experts. Early studies showed that even the interpretation of visual evidence could be biased by expectations. For example, Bruner and Potter (1964) showed participants a set of blurred images which were gradually brought into focus. They found that exposure to extremely out of focus images made it more difficult for participants to identify the image as it was brought into focus (i.e. early interpretations of the image inhibited subsequent correct recognition). This phenomenon has been replicated on numerous occasions with the same results suggesting that the initial hypotheses that people form to explain or understand ambiguous event may make it difficult for them to interpret subsequent detailed information. Thus, prior expectations clearly affect the perception and interpretation of visual stimuli.

In a series of studies examining the accuracy of fingerprinting experts Dror and his colleagues (see Dror & Charlton, 2006; Dror, Charlton & Person, 2006; Dror, Peron, Hind and Charlton, 2005) found that fingerprint matching decisions, including those made by expert forensic examiners were also biased by extraneous contextual information. Specifically, that visual information (fingerprints) was interpreted in a manner consistent with initial expectations. For example, in Dror et al. (2005) the difficulty of the matching task was varied and some participants were also given additional information about the crime, such as where the fingerprint was obtained. Some participants also saw emotional photographs that related to the scene of crime. Finally, some participants were subliminally primed with the words “guilty” and “same” during the matching task. Results indicated that both emotions (as aroused by background story and photographs) and subliminal messages did influence decision making in certain circumstances. Specifically, when the matching task was easy (i.e. the fingerprints were a clear, uncomplicated match) the extraneous contextual factors did not affect the accuracy of the decision-making. However, when the task was difficult and the fingerprints were not a clear unambiguous match, errors consistent with the contextual information were observed. When the fingerprints presented were ambiguous, participants in the control condition found a match for 47% of the trials whereas participants in the high emotion plus subliminal message condition found a match for 66% of trials (58% in high emotion only condition). Dror et al. (2005) concluded that top-down influences (i.e. contextual information) biased decision making when the task was ambiguous but did not over-ride bottom-up processing (i.e. the objective analysis of fingerprint attributes) when the task was clear-cut.

Dror et al. (2006) replicated these findings in a similar study using fingerprint experts. In this study, the experimenters selected fingerprints that had previously been evaluated by the experts in the normal course of their work. These fingerprints were then submitted for a second analysis by the same experts (who were not aware which fingerprints they would be tested on or when). Participants were asked to examine the target fingerprint alongside an exemplar print (a print obtained from a suspect) and were then provided with inaccurate contextual information concerning the print. The misleading information was designed to generate an expectation that the print was a non-match – participants were told that the print had been erroneously matched by the FBI as the Madrid bomber. Results indicated that fingerprint experts, with on average 17 years of experience, were just as susceptible to extraneous contextual information as non-experts. When presented with a different context, four out of five experts made different identification decisions to those they had made previously. In fact, three of those four experts decided that the fingerprint was now a definite non-match despite having identified those same prints as a definite match previously in the absence of contextual information. Clearly this research has important applications, given that in reality, fingerprints – like the components of other matching tasks including faces – are unlikely to be perfectly clear and obtained under optimal conditions. Previous sections have discussed the similarities and differences of face- and non-face matching processes. Nonetheless, this research demonstrates that the effects of confirmation bias extend beyond basic perception and interpretation, and affect matching decisions. Thus, confirmation bias is relevant to a consideration of decisional factors that may contribute to face-matching errors.

Several other studies have recently investigated the impact of the confirmation bias and associated effects on evidence evaluation in forensic settings. For example, Ask and Granhag (2005) presented both experienced police investigators and a student sample with case materials relating to the preliminary investigation of a homicide. Background information was also provided which suggested that either the suspect had a motive or that an alternative unknown offender committed the crime. Neither hypothesis had any basis in available evidence. Results indicated that the student sample demonstrated a clear confirmatory bias – participants who were made aware of a potential alternative perpetrator were less likely to view the main suspect as guilty. In this instance, police investigators did not appear to interpret the evidence in line with the background information concerning an alternative perpetrator. Instead, this group rated it likely the main suspect was guilty irrespective of background information. Ask and Granhag (2005) suggest this may be due to another commonly held preconception or 'guilt' bias which has been documented elsewhere (Baldwin, 1993; Leo, 1996).

Research has also shown that the sequential evaluation of different pieces of evidence can be distorted in favour of an initial hypothesis. Hope, Memon and McGeorge (2004) tracked the course of evidence evaluation in mock juror decision making and found that the effect of biased evidence evaluation was cumulative. Specifically, when an initial evaluation is biased (on the grounds of preconceptions or expectations) each subsequent evaluation biased in the favour of the previous (biased) evaluation. For jurors exposed to negative information about the defendant, this distortion process was exacerbated and the prosecution was more strongly favoured as the leader. In addition, evidence supporting the prosecution's case was more favourably evaluated or, alternatively, the evaluation of pro-defense testimony was distorted in favour of the prosecution.

Other research has also demonstrated that the confirmation bias is exacerbated by sequential information processing (Jonas et al., 2001) and that people use different cognitive processes when faced with sequential versus simultaneous information (Hogarth & Einhorn, 1992). Under sequential presentation, new items are immediately compared with prior or pre-existing beliefs (Edwards & Smith, 1996) and assessed relative to this prior belief. Jonas et al. (2001) have argued that sequential presentation involves a repeated consideration of a prior belief or evaluation and a concomitant increase in confidence in the veracity and reliability of this prior evaluation. This repeated, but biased, evaluations leads in turn to increased commitment to the belief or evaluation (see also Koehler, 1991; Schulz-Hardt, Frey, Lüthgens & Moscovici, 2000; Tesser, Martin, & Mendolia, 1995).

Decision-makers who have been biased by their expectations or preconceptions are typically unaware that their decisions have been distorted and tend to retain an "illusion of objectivity" despite their selection attention to particular information (Pyszczynski & Greenberg, 1987). In other words, decision makers tend to report that their decision making has been unbiased and objective.

Asymmetrical scepticism. A second related tendency which can influence evidence evaluation is asymmetrical scepticism. Rather like confirmation bias, asymmetrical scepticism is a naturally occurring tendency for people to scrutinise information which threatens previously held beliefs or preconceptions more rigorously than information which is positive with respect to an existing belief or value (Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998; Ditto, Munro, Apanovitch, Scepansky & Lockhart, 2003; Lord, Ross & Lepper, 1979). Ask and Granhag (2007) examined the occurrence of asymmetrical scepticism in criminal investigations and found that experience interviewers judged the reliability of witness statements different depending on whether the statements confirmed – or disconfirmed – prior hypotheses held by the investigators concerning the case. Specifically, the results indicated that although the witness statements were produced under the same circumstances by witnesses with the same characteristics “investigators subjected the disconfirming (vs. confirming) statement to stricter scrutiny and hence found stronger grounds for questioning its reliability” (Ask et al., 2008).

In their most recent examination of bias in the evaluation of criminal evidence, Ask et al. (2008) considered the extent to which different types of evidence might be susceptible to extraneous biases. Specifically, they considered the extent to which the perceived ‘elasticity’ of the evidence varied with objective nature of the evidence where ‘elasticity’ refers to the degree of ambiguity associated with the piece of evidence. In other words, if a piece of evidence is open to multiple interpretations, decision makers tend to give weight to the interpretation most consistent with their initial hypotheses or preferences. In the Ask et al. study, police trainees were given case materials (homicide) which contained information suggesting that a particular suspect was guilty. The purpose of this information, as in previous studies, was to set up a prior belief or hypothesis concerning that particular suspect. After participants had indicated their views on the case, new evidence was then introduced which was either consistent or inconsistent with the guilt-related suggestion. Furthermore, this evidence varied in terms of its perceived elasticity. In one condition, participants were provided with DNA evidence which is typically associated with a low degree of elasticity given the limited possibilities for subjective lay interpretations. In a second condition, participants were provided with a visual image (pictures taken from a CCTV security camera) which was deemed to have a moderate degree of elasticity (i.e. a moderate possibility of subjective interpretation). In the final condition, participants received details of witness evidence which was ascribed a high degree of elasticity on that grounds that witness evidence can be open to a number of biases and different interpretations based on both witness and situational factors (Wells & Olson, 2003). Analyses revealed asymmetric scepticism in the evaluation of evidence - participants rated the disconfirming (as opposed to confirming) evidence as less reliable and generated more arguments supporting this point of view. Furthermore, this scepticism was exacerbated for highly 'elastic' witness evidence.

Other potential sources of biases. Many 2-alternative cognitive judgments show effects of the frequency of the alternatives occurring, with infrequent targets reducing the likelihood of people spotting them when they occur (see, e.g., Whitman & Geller, 1972; Wolfe, Horowitz & Keener, 2005). It is interesting to note that for forensic tasks, sometimes the target of a facial comparison is a match (e.g., looking through CCTV footage for a suspect) and other times the target is a mis-match (e.g., checking whether people are carrying their own passport or someone else's by inspecting the photo). Regardless, the majority of studies of face matching have used target frequencies at or close to 50%, whereas the likelihood of a target in forensic matching is much lower. So it is important to understand whether frequency of target presence does change the care with which people inspect faces. Bindemann, Avetisyan and Blackwell (2010) tested the frequency of targets in a task where people compared photos of unfamiliar people taken with different cameras to find mismatches. In an Infrequent target condition, mismatch targets (1 pair out of 50 shown) were spotted as often than in a Frequent target condition (25 pair out of 50 shown). They concluded that the infrequency of mismatching occurring in forensic settings will not be a problem in any task that requires someone to match two photos of faces. Unfortunately, there is a problem with what they did. In an attempt to equate the amount of task experience gained before the mismatch trials that were compared, the last image pair compared was always a mismatch, which meant that in the infrequent condition, it was the only mismatch encountered. In the infrequent condition, that meant that participants had made 49 matched pairs before encountering the 1 mismatched pair. As they had been told that there would be mismatches on 2% of trials, by the time the mismatch occurred, participants were likely have a strong expectation to encounter a mismatch. The reason for this is that they would have encountered 49 mismatched paired, and so would likely have made a string of "match" responses, even if they made erroneous "mismatch" responses on some trials. Historically, reducing recency of an alternative response in a forced choice task increases expectancy for it to occur (Rabbitt & Vyas, 1979). This confound of overall expectancy and recency makes it unclear whether people in the study maintained their standard of checking for mismatches even when they were infrequent, or alternatively that they simply expected a mismatch more at the end of the study than at the beginning, leading to a less careful inspection, and so a potential false alarm.

A variety of stressors can affect decision making (Dror, Busemeyer, & Basola, 1999) and also increase the chances that biased information will impact negatively on judgements and decision making (Kruglanski & Freund, 1983). For instance, time pressure has been consistently shown to impair decision making and lead to increased bias and selectivity information processing (Edland & Svenson, 1993). Time pressure has also been shown to increase reliance on stereotypes and heuristics allowing less time for careful and considered processing of evidence (Bodenhausen, 1990; Dijker & Koomen, 1996). Ultimately, time pressure can severely restrict the ability to generate and consider alternative hypotheses, thereby limiting the flexibility of the decision maker to consider other outcomes or possibilities.

4.3 Conclusions

The psychological research literature reviewed above clearly demonstrates that both the matching and identification of a previously unfamiliar face from static images, live scenarios or dynamic footage is a difficult and error-prone task. These findings are counter-intuitive to the lay assumption that matching a face to either a suspect photograph or the actual live suspect should be a simple, objective and non-fallible task. In other words, people expect to be able to do this task with a high degree of accuracy. However, research consistently demonstrates poor performance on these tasks – even under optimal conditions.

However, by contrast, accuracy rates for the identification and matching of people with whom the observer has interacted previously – even from poor quality footage – are typically high.

When a video is one of the images to be involved in a comparison, it is not important whether the dynamic video or stills taken from it are used. Matching performance is equivalent for the two situations. When dynamic stimuli show an advantage, it can be attributed to having seen the face from multiple viewpoints, not because dynamism helps people to understand the structure of the face.

Recognition is more difficult when the images show the faces from different viewpoints. Some of this is due to the absence of critical face features in one or both images, and that difference may not always be overcome. There is also likely to be a need for mental rotation of the mental representation of one or the other face in order to align the faces enough to compare them.

We could not find enough literature to see how differences in age affect face comparison, other than to say accuracy is lower when faces show different ages. Likewise accuracy is lower when any of a number of transient characteristics change (e.g., clothing, hairstyle, hair colour, facial hair, etc.). Observers can be trained to ignore some of these irrelevant variances between images, but it is likely that they will continue to have some effect, even if it is only in the effort required to compare the images.

In addition to the effect of stimulus characteristics, performance may vary with social and/or environmental pressures influencing the matcher's decision criteria (e.g., expectations, perceived consequences of different error types).

5. Individual Differences in Face Processing Skill

There have been a few attempts to find neurobiological differences correlated to face processing ability. Schretlen, Pearlson, Anthony and Yates (2001) found a correlation between face matching ability, perceptual speed and cerebral volume. Zhu et al. (2010) tested whether face processing has a genetic component. Participants were 189 pairs of monozygotic or dizygotic twins ranging in age from 7-19. They performed recognition tasks and matching tasks that used successive rather than simultaneous displays. The matching tasks compared performance on upright, unfamiliar faces to performance on houses, inverted faces, or composite faces. Performance benefits for upright faces over the comparator stimuli were significant for all tasks, and were larger for monozygotic twins than for dizygotic twins. Since monozygotic twins share more genetic material than dizygotic twins and since there was no reason to suspect large differences in environmental factors between twins in the sample, the conclusion is that there is a genetic component to face processing ability. The amount of variance accounted for by genetics was less than or equal to 40% for the tasks (less than or equal to 60% for older twins), so the conclusion to reach is that face processing skill may not be completely malleable, due to genetic contributions, but there is also a lot of room for performance improvements due to training or environmental factors.

Several behavioural tests have been developed to find individual differences in face-processing skill, including the Benton Face Recognition Test (Benton, 1983), Warrington's Recognition Memory Test (1984), Bielefelder's Famous Faces Test (BFFT, Fast, Fujiwara & Markowitsch, 2005), the Cambridge Face Perception Task (Duchaine Germine & Nakayama, 2007), and the Cambridge Face Memory Task (CFMT, Duchaine & Nakayama, 2006). These have been criticised for the images showing more than the face, which means that they could incidentally measure more than just face processing ability (Duchaine & Weidenfeld, 2003; Herzmann et al., 2008). Tests that use famous faces rather than recently learned faces have been criticised as measuring semantic knowledge about people as well as strictly defined face processing (Hertzmann et al., 2008).

The CFMT (Duchaine & Nakayama, 2006) shows three photos of a face (showing only the interior part of face) and then presents a test display of multiple faces where the participant has to identify which is the target face. Difficulty of the task increased as the test progressed. In the easiest part of the test, the photo of the target face was identical to one of the photos used during learning. The intermediate section of the test showed the target face using a different photo. Finally, the difficult section of the test used a different photo and imposed visual noise over all photos so that it was harder to see. Performance on this task resulted in mean accuracy of 80.4%, ranging from 59% to 99%. Scores were more variable in the difficult section than in the intermediate section, which in turn were more variable than in the easy section. That this test measures face processing ability was inferred from the fact that when faces were inverted, performance on the test decreased (mean accuracy = 58.4%).

Russell, Duchaine and Nakayama (2009) investigated whether some people could be considered "super-recognisers" (i.e., people who demonstrate exceptionally good face recognition). They used the Cambridge Face Matching Test, but because this test failed to discriminate well between people with better than average ability, they modified it. To improve high-end discrimination, the researchers added 30 more difficult items to the original 72. In this last section, the faces to be learned were the same, but the displays for testing recognition were different. Photos showed faces in quite different orientations (profiles rather than ¾ views), added the external features of the faces to the photos to be chosen between, or showed faces with a different expression. Four self-selected super-recognisers and an age-matched group of control participants completed this test and a test of identification of famous faces that used photos of the famous people before they were famous. Performance on the CFMT was better for each of the super-recognisers than for all control participants. One control participant matched super-recognisers on accuracy at recognising the famous faces, but the rest had lower accuracy. Thus, super-recognisers do exhibit superior performance on face perception tests involving memory. The study then compared super-recognisers to new control participants on the ability to rank order six morphed faces in their similarity to a standard photo. For each photo, the standard was morphed with another face with a varying proportion of inclusion of the standard face. Importantly, each of the six photos morphed the standard face with a different other face. Thus, although similarity could be ranked, there was variability in the faces. This is a face-matching task. As a group, the super-recognisers were more accurate than the control group but there were 4 or 5 individual control participants whose performance matched or bettered that of super-recognisers. Thus, super-recognisers had very good face-matching ability, but their face-matching ability was not as superior to controls as their face-recognition ability. When the test was repeated with inverted faces, there was almost no difference between the groups, with two super-recognisers performing below the average performance of the control group and two performing quite well. A third experiment in the study compared performance of control participants on the recognition and the matching test, and found a strong correlation between scores on the two tests (r=.667).

Herzmann et al. (2008) raised the concern that each of the tests people were using to assess face processing ability used a single face processing task, when little is known about the reliability of most of the tests and the degree to which any individual test captures all facets of face processing skill. To address these limitations, they developed a battery of tests aimed at limiting the influence of non-face information on responding. These tests aimed to assess both the speed and accuracy of face processing using multiple measures. The performance of a large group of young adults revealed that the battery of tests captured more about differences between people than any individual test.

Wilhelm et al. (2010) exploited this battery of tests to show that there are differences between how face perception and face memory skills vary in a broader sample of people (broader in terms of age but still all Caucasian), and demonstrated that these skills are largely independent of speed of face processing as well. The study further demonstrated that both the perception and memory skills underlie face recognition, whereas perception plays a dominant role in face matching. This means that evidence from face recognition studies must be carefully considered to determine the degree to which it informs about perception rather than memory.

Hildebrandt, Sommer, Herzmann and Wilhelm (2010) retested the Wilhelm et al. (2010) model of face processing skills using a population that varied continuously from age 18 to age 88, with a total of 452 participants. The contribution of the factors (face perception, face memory and speed) had the same relationship to task performance over the lifespan. Speed decreased with age from about age 30. Decreases in face memory ability were smaller, beginning at about age 45-50, and decreases in face perception ability were also smaller, beginning at about age 60.

In forensic facial image comparison, it is common for extraneous details to vary between images. Any test of skill that is influenced by matching clothing or other extraneous details is likely to provide less information about forensic face matching skill than is ideal. This has been addressed in the Glasgow Fact Matching Test (Burton, White & McNeill, 2010b). The test requires comparison of images that differ in camera source (one is a still image extracted from a video and the other is a still photograph taken in the same session). In both images, the face and hair appear against a uniform background, with no clothing showing. The photos were predominantly of university-age Caucasian individuals. Performance on the test was evaluated in a sample of 300 people (40% male) from the Glasgow area ranging in age from 18-80. With this more ecologically valid test, mean accuracy was 89.9%, ranging from 62% to 100%. Median accuracy was 92%. The cut-off score between the top 10% of scorers and the rest is 98%. Thus, this test like the CFMT, does not have good high-end discriminability, and may test comparisons that are not particularly difficult. However, the test was shown to be reliable, and only had a small correlation between accuracy of responses and time to complete the test (r=.177). The participants also completed a face recognition test, using faces from the same database, a test of matching of line drawings of familiar objects, and a test of visual short-term memory. Scores on the GFMT correlated most strongly with scores on the object-matching task (r=.420), less strongly but still significantly with scores on the recognition memory test (r=.285) and not at all with scores on the short-term memory test.

5.1 Conclusions

The main aim of examining individual differences in face processing skill is to understand how much skill varies and to determine the degree to which this variance is specific to face processing. The studies reported here, none of which include people diagnosed with conditions associated with impaired face recognition, show there is substantial variation in face processing ability in the general population, and that there is a portion of that skill that is not explained by general cognitive abilities or object perception abilities. The studies also show that face perception ability varies somewhat independently of an underlying face memory ability.

Tests of face processing ability tend to show ability to be skewed toward skilful performance, which suggests that the tests are less able to discriminate different levels of fairly skilful performance than to discriminate different levels of fairly unskilled performance. The fact that it requires particularly difficult recognition tests to elicit variance in performance among skilled highlights the importance of choosing stimuli for training that require subtle discriminations. Thus, if training were to use recognition tasks, it would be important to select stimuli that would clearly train the perceptual part of the task.

The twin study suggests that genetics will be a limit to the malleability of skill, but it also demonstrates a substantial non-genetic component, showing that environmental and training factors ought to be influential. Finally, given that face perception is far more important than face memory to facial image comparison, the study of face processing across the life span shows that there is no reason to question whether face matching ability is likely to degrade with age in forensic employees during the ages that people are typically employed. This assumes, of course, that (lens-corrected) visual acuity of employees remains generally good.

6. Training of Face Matching and Recognition

6.1 Effects of practice occurring outside the lab

Extensive naturally occurring experience with particular face types impacts how those faces are processed. This is seen in the other-race effect, where faces of people from one's own race are recognised more accurately than people from another race (Meissner & Brigham, 2001). It is also seen in the other-age effect, where faces of unfamiliar children are more easily recognised by preschool teachers with many years of experience working with children than by age and gender matched adults with very little experience with children (Kuefner, Macchi Kassia, Picozzi & Bricolo, 2008). For non-teachers, inverting faces lowered accuracy more for adult's faces than for children's faces, but for teachers, inversion had the same effect for both types of faces. De Heering and Rossion (2008) extended this to demonstrate that extensive experience made face processing more holistic. They studied pre-school teachers with many years of experience working with children compared to age and gender-matched adults with little experience with young children. The two groups showed a typical composite face effect when shown adult faces: a greater tendency to be influenced by the bottom half of the face when judging the top half of aligned faces rather than misaligned faces. Preschool teachers showed a composite-face effect of the same magnitude when shown faces of children, but adults with little experience with children showed less effect of alignment. What this suggests is that both groups have gained sufficient experience with adult faces to process them well through largely holistic processing. For child faces, however, only the teachers have had enough exposure to have the same accuracy and holistic processing as seen for adult faces.

To date, there is mixed evidence from very few studies about whether people with experience of the forensic identification process (such as experienced police officers) perform any better than untrained, inexperienced individuals when required to match or identify a previously unknown face. In Burton et al. (1999), police officers demonstrated the same low accuracy rates as student participants when attempting to identify and match an unfamiliar target face.

Wilkinson and Evans (2009), on the other hand, report an effect of forensic experience on matching of faces between photos and CCTV footage. They tested two academics who frequently were hired to act as expert facial image comparers in court cases and a control group who were members of the public. When the person in the video was not wearing a cap, experts were 85% correct over target-present and target-absent lineups, making 3% of possible false positives and 8% misses of targets. The general public, on the other hand, was 53% correct overall, making 10% of possible false positives and 39% misses. When the person in the video was wearing a cap, experts were 81% correct overall, making 2% false positives and 25% misses. The general public were 37% correct overall, making 10% of possible false positives and 59% misses.

There were a number of potentially important differences between the two studies. People with different forensic experience were used. Sample sizes were different, and in fact, it would be quite risky to generalise from a study using only two participants. The perspective of the CCTV footage was different. In one study, still images from CCTV footage were used whereas in the other, dynamic footage was used. And the two studies had a different number of choices in the lineup of photos (one versus five).

A final difference deserves a bit of explanation. Because Wilkinson and Evans were conducting a pilot study, they used themselves as the expert participants. This almost certainly led to a difference in task motivation for the experts, who would wish to perform well (e.g., perhaps to demonstrate the need to get funding to follow up the pilot) compared to the controls, who because they were not vested in the outcome of the study would not be particularly motivated to perform as well as possible. It would be better to wait for the outcome of Wilkinson and Evan's follow up study than to act based on this pilot study alone.

Because of the many differences between the two studies, it is impossible to completely reconcile the results. Burton et al. (1999) was the more careful and complete study. However, Wilkinson and Evans (2009) does give hope that experience will improve performance. It has to be said, though, that any study that compares pre-existing skill level rather than manipulating it suffers from the inability to say whether the job led to the skill or the skill led to the job.

6.2 Training face recognition in the general population

A few studies of the effect of training on face recognition or face matching have targeted samples of people from the general population.

Beginning with training of face recognition, Seipel's PhD thesis (2002) tested the effect of three training methods on subsequent lineup performance. In all training sessions, participants decided which of two faces presented matched one that had just been presented. Training differed by how the two alternative faces differed. Participants trained to focus on configural relationships were shown alternative faces that differed only by spacing between features. They did not perform more accurately on a lineup task in which target-absent responses were allowed (29% correct after training versus 13% correct before). False alarms were reduced (21% after versus 50% before). Participants trained to focus on features were shown alternative faces differing only by a feature. They were more accurate after training (42% correct after training versus 13% before). False alarms were reduced (25% after versus 50% before). Finally, participants trained to focus on the interior part of the face were shown alternative faces that differed in both configuration and features, where all changes were confined to the internal region of the face. This training method did not improve lineup performance reliably (17% correct after training versus 13% before; false alarms: 50% both before and after). Of the three methods, then, training people to compare features was the most effective.

The ineffectiveness of training people to attend to internal parts of the face is interesting. Familiar faces are recognised and matched more accurately than unfamiliar faces, and people tend to rely more on information from the internal features for familiar faces than for unfamiliar faces. This correlation in trends has led some researchers to assume that training people to focus on internal features of unfamiliar faces might improve performance. The results presented by Seipel (2002) suggest that this may not be the case. The ineffectiveness of training individuals to attend to internal features is supported elsewhere. Firstly, when researchers allowed untrained observers to look where they wish during a face matching trial, they found that, provided that there was a sufficient amount of time to process the display, those who gazed more at internal features were no more accurate than those who gazed at external features at least as much as at internal features (Fletcher, Butavicius & Lee, 2008). Secondly, in Burton et al. (2010a), training to focus on internal features improved performance a little in a laboratory study but not in a more realistic setting where the critical photos were being compared to live people.

Tanaka & Pierce (2009) demonstrated that giving practice at individuating people of another race improved recognition more than giving practice at categorising them by race. Each participant was trained for almost four hours distributed across five days, where one “other race” was the subject of individuation training and the other “other race” was the subject of categorical training. The two races used, counterbalanced across training, were Hispanic and African American. Participants were mostly Caucasian, but some were Asian or First Nation Aboriginal. In training, each face was presented alone, and participants were to report its name or category, depending on the race. Before training, recognition performance was assess using different faces from the same dataset, and was found to be equivalent for the two groups. Both kinds of training improved accuracy for target absent trials: False positives dropped from 40% before training to 32% after categorisation training, and dropped from 38% before training to 32% after individuation training. Individuation training, but not categorisation training, increased accuracy for target present trials. There were misses on 42% of trials both before and after categorisation training. Misses dropped from 44% of trials before individuation training to 34% of trials after training. Thus the benefit of individuation training over categorisation training was in reduction of misses.

Turning our attention to training of face matching, Burton et al. (2010a) report an unpublished study of face matching where the goal was to decide if inspection of passport photos would benefit from training and/or different types of photos. In the lab, participants compared a critical photo to either another photo or a video, with the goal of detecting critical photos that did not match. T The training was designed to increase focus on internal region of the face. In the basic laboratory experiments, training improved performance significantly. However, in a simulation of a passport-checking environment, this effect was not replicated. Notably, feedback about accuracy did improve performance substantially.

6.3 Training of those poor at face recognition

Face recognition ability varies in the population because of developmental problems, brain damage, and natural variation. A number of studies have aimed to rehabilitate those who are poor at recognition. Improvements have been seen in children with autism (Tanaka et al., 2010), adults with brain injury (Powell, Letson, Davidoff, Valentine & Greenwood, 2008), adults with developmental or acquired prosopagnosia (Francis, Riddoch & Humphreys, 2002; DeGutis, Bentin, Robertson & D'Esposito, 2007), long term alcoholics (Hochhalter et al., 2001), and the elderly (da Silva & Sunderland, 2010). Training techniques that have proved effective in these studies include having trainees engage in semantically enriched visualisation during learning (Francis et al., 2002), rewarding correct recognition (Hochhalter et al., 2001), caricaturing faces during learning, having trainees make semantic associations during learning, or focussing attention on face parts during learning (Powell et al., 2008), providing immediate rather than delayed feedback about accuracy (da Silva & Sunderland, 2010), focussing attention on spatial relationships between face parts (DeGutis et al., 2007), and permitting practice on attending to the eyes, processing faces holistically, and comparing faces differing in expression or viewpoint (Tanaka et al., 2010).

6.4 Conclusions

There is insufficient empirical evidence to suggest that people with experience of the identification process (such as experienced policed officers) perform any better than untrained, inexperienced individuals when required to match or identify a previously unknown face.

Surprisingly little work has attempted to change face recognition or face matching performance of people from the general population through training, and what has been done has had at best modest effect. The work done with adults from the general population suggests that giving feedback is quite helpful, probably because without training observers are probably not confident about their responses. To a lesser degree the research suggests that it may be helpful to use training that focuses observers on particular ways that faces might differ.

Note that only the Burton et al. (2010a) training has given observers explicit instruction about strategies to take. The other studies have shaped the desired behaviour by presenting stimuli that require the kind of comparison thought to be useful. In general, training is most effective when it combines explicit and implicit training, so this may need further consideration. The degree to which explicit training is useful largely depends on how much control observers have over their perceptual behaviour. If useful conscious strategies are available, then explicit training is called for, perhaps in conjunction with implicit training.

Another limitation of all the studies is that they have not tested whether the effect of training persist. Before any conclusions can be reached about the efficacy of training, it is critical to ensure that that perceptual behaviour has been (relatively) permanently changed rather than merely primed for a short duration.

It is important not to draw too strong a conclusion from the studies in which individuals with face-processing deficits were trained for at least two reasons. First, training in these studies improves recognition to within a “normal” range, meaning that performance before training is particularly poor and performance after training is merely above an acceptable threshold to allow adequate everyday functioning. But of the studies that give distributional details of performance, none have found training to result in performance that equates to skilled performance. Secondly, reliance on studies involving abnormal populations is problematic as most of the original deficits displayed by participants in such studies resulted from brain damage or abnormal brain functioning. In general, individuals who have had brain damage can relearn some tasks, but are sometimes found to do them differently than neurologically intact people. This learning of alternative behaviour can also happen for developmental disorders as well. One documented example of this for face processing is that children born with bilateral cataracts have quite limited visual experience in infancy. Recent studies have shown that when there is even 3 months of childhood with these cataracts present, which deprives the infants of early visual experience, even 10 years after cataract removal the now adolescents or young adults can recognise faces but are unable to do configural processing of faces (Le Grand, Mondloch, Maurer Brent, 2001; Mondloch, Robbins & Maurer 2010). Thus, some aspects of face processing, if not learned during critical periods in childhood, may never be teachable.

7. Perceptual Skill and Expertise, Generally

7.1 Description of perceptual expertise

Expertise arises through practice and experience. Accepted hallmarks of perceptual expertise include basing responses on holistic processing or at least on larger parts of the stimulus than in novice behaviour (Busey & Vanderkolk, 2005; Goldstone, 1998). Expert behaviour is also largely automatic – fast, unconscious, and difficult to unlearn (Goldstone; Ro, Friggel & Lavie, 2009). Those with perceptual expertise do not necessarily hold more items of information in working memory, but each object is represented in more detail in an expert's working memory than in a novice's (Sun, Zimmer & Fu, 2010). Experts tend to make more fine-grained groupings of members of the class of objects that is the subject of expertise than novices would (Johnson & Mervis, 1997; Tanaka & Taylor, 1991). Finally, experts are more able to generalise learned processing to newly encountered instances of the class for which one has expertise (Gauthier, Williams, Tarr & Tanaka, 1998).

7.2 Perceptual skill training

As expertise arises through practice, it is informative to consider studies aimed at training perceptual skill. Tanaka, Curran and Sheinberg (2005) used an everyday class of stimuli (birds) which some people are expert at identifying (bird watchers) but most people have rudimentary skill at identifying. They trained two groups of unskilled people to identify owls and wading birds. One group was trained to classify them at a species level (i.e., label them as owls or wading bird) and the other was trained to classify them at a family level (e.g., label them as barn owl, eastern screech owl). The number of times they saw the pictures was equal for the two kinds of judgments. After training, the group that had made the more fine-grained classification during training was more accurate at classifying new examples of owls and wading birds according to species. Thus, it was not mere exposure to examples that led to expertise at classifying birds, it was practice at individuation rather than classification. Of course the aim of training at image comparison is to be able to individual faces, not categorise them. Tanaka and Pierce (2009) showed that practice at individuating faces rather than classifying them led to better individuation for new members of the classes of faces (see also Gauthier et al., 1998; Yue et al., 2006).

Another approach has been taken to improve training in difficult real-world perceptual tasks (e.g., spotting tumours in X-rays, or inspecting aircraft for signs of wear). In these situations, an image to be inspected was shown to the trainee superimposed by either a static or dynamic display of where and how long an expert gazed as they inspected the image. The trainee was instructed to follow the expert's gaze, but was not told what decision the expert made. Doing this increased trainees' and shortened the required duration of training to achieve threshold performance (Litchfield Ball, Donovan, Manfield & Crawford, 2010; Nalanagula, Greenstein & Gramopadhye, 2006; Sadasivan et al., 2005). This kind of display is effective only if the eye movements and the image match (i.e., the eye movements were not from inspection of another display; Litchfield et al).

Because the trials in which oculomotor demonstration (i.e., exposure to expert's eye movements) was used were also the trials on which performance was measured, these studies did not demonstrate generalisation of learning to new images. But recent unpublished work has demonstrated generalisation of training to new displays, at least in the short term (Pfister & Hillstrom, in preparation).

Combining this technique with explicit instructions about how to do a task has proved a more effective way of training judges of gymnastic performance than simply giving explicit instructions (Page, Lafferty & Holder, under review).

It is important to understand why a training technique works, not just that it does work. Why would oculomotor demonstration train perceptual skills? We contend it is because of learning mechanisms that are best seen in early childhood but that carry over into adulthood. In particular, very young children learn what is interesting or important in the world from their caregivers, and gaze seems to be particularly important in that. It can be seen in behaviour called joint attention. As a caretaker interacts with a very young child and the world, the child tends to follow the caretaker's eye movements to develop a joint understanding of the actions they are engaged in and the importance of the objects in the environment (Tomasello, 1995). The importance of other people's eye movements is believed to be inherent from birth, and joint attention develops by about the ninth month (Mundy & Crowson, 1997). In adulthood, other people's gaze continues to direct people's attention (Friesen & Kingstone, 1998), and people seem to implicitly assess the value of objects according to whether they are looked at by others (Bayliss, Paul, Cannon & Tipper, 2006). We speculate that this sensitivity to gaze makes showing someone else's eye movements a particularly rich non-verbal training method of where to find and how to process critical information in the visual world.

In the oculomotor demonstration training technique, gaze is far more abstracted than in joint attention studies: the trainee is seeing not the expert's eyes but a dot representing their gaze. However, there is evidence that people respond to gaze of a virtual entity much as they respond to real gaze (Schilbach et al., 2005). Thus, it is the representation of intended gaze, rather than gaze itself, to which people seem to respond. Further, even if the following of gaze is more deliberate in this kind of training than in real joint attention, experience with joint attention will have trained people to evaluate carefully what others gaze at.

However, for two reasons it cannot be guaranteed that this technique would improve face matching or face recognition performance. First, the technique has only been tried for instructing individuals inexperienced at a task. Adults, on the other hand, have such extensive experience in identifying faces that it is unclear whether oculomotor demonstration would add anything to what they already know. Second, recognition of faces can be automatic, and faces are often recognised with a single fixation (or very few fixations). This might limit the effectiveness of demonstrating gaze.

In its favour, forensic face matching if far less automatic than face recognition in day to day life. This can be demonstrated by the time taken to do it. More importantly, it seems that skilled behaviour at face matching may be partly implicit behaviour (i.e., behaviour that people cannot always say how they do), in which case adding an implicit training tool to explicit instruction may be useful.

7.1 Conclusions

There are recognisable hallmarks of expert behaviour. Although there are few studies of training of perceptual expertise, there are two techniques we address. The first is to provide the opportunity to practice individuation of objects relevant to expertise. This is a straightforward technique that has one important feature: the training must be equal for all subclasses of stimuli so that people do not become expert at one class and inexpert at the other(s). The second technique, oculomotor demonstration, is not a proven technique, but shows promise in the tasks to which it has been applied. It is likely to be more useful for tasks that have a strong implicit component, and we believe face matching has that characteristic.

8. General Summary

This review has explored research and theory on face recognition and face matching ability. We have identified, based on an understanding of the psychological processes underlying these skills, various causes of error in these tasks. We have built on previous work, and psychological theory to inform training practices. Based on this review, we make the following recommendations.

1. Existing studies of face processing training have not been sufficiently thorough and methodologically sound to firmly recommend implementation of any single approach to face-processing training.
2. Giving trial-by-trial feedback during training is recommended. If trainees erroneously match two stimuli during training, then highlighting the differences between the faces may be beneficial if there is a way to highlight the differences in the faces during feedback, then do it. If people erroneously fail to match two faces, then it may be beneficial to draw the trainee's attention to differences between the photos that were irrelevant. However, these recommendations require more research to determine the extent of their efficacy.
3. During training, it may be beneficial to require trainees to rate their confidence in their judgments. There are two reasons why this might be beneficial. First, it allows a better basis for assessing whether they are discriminating well, in that errors made with confidence need remediation more than errors made with uncertainty. Second, confidence will play a role in the judgments they make on the job. In many forensic face-matching scenarios, although they may be asked to rate their confidence through something like describing the amount of evidence their judgment is based on (Wilkinson & Evans, 2009). One would hope that confidence is calibrated to actual performance, but trainees may require additional feedback about performance to determine whether their confidence is warranted.
4. Configural or holistic processing will play a role in on-the-job performance, and for this reason it may be important to train individuals to identify configural or surface-feature differences. Configural processing might be trained using blurred photos of faces or using photos in which the spacing between features

is manipulated. It is best to use the kinds of stimuli that will be encountered, when possible. This means that training using blurred images might be more appropriate than training using manipulated photos because use of manipulated photos requires that the two images are exactly the same both in face features and in irrelevant features (e.g., lighting, viewpoint). It would be highly unlikely in a forensic context to be asked to compare such similar photos.

5. Although face matching is partly achieved through holistic processing, in forensic settings there is probably a stronger contribution from explicit use of featural comparison than would occur in face processing in more mundane contexts. Therefore, we suggest that part of the training is focussed on featural comparison. It might be worth using composite face stimuli for this part of training. To our knowledge, all matching tasks using composite face stimuli have used identical images, such that if the top half of the faces matched, the entire top half of the photos matched. For realism, it seems like a good idea to use different photos of same faces. However, further research is necessary to determine the efficacy of training in featural comparison.
6. During training, the use of challenging stimuli will be critical. Laboratory studies find quite high accuracy for simple stimuli. Applied studies find quite low accuracy for real-world stimuli. Trainers should ensure that the set of stimuli used include lighting differences, viewpoint differences, differences in distance from the faces, and other-race faces. All of these factors can affect performance, and it is important for trainees to encounter such factors during training.
7. We were unable to locate a literature on comparing unfamiliar faces across ages to judge whether the older face is from the same person as the younger face. We would expect that explicit training about how faces typically change with age would be helpful, but it is not possible to fully support this suggestion due to an absence of relevant literature.
8. The literature appears to suggest that matching performance for static-static comparisons is comparable to matching performance for static-moving comparisons. If anything, moving images produce better accuracy because of the multiple views of the person provided. Obviously, static images can

usually be extracted from dynamic footage (e.g., the comparison of CCTV to photos might be trained by using stills from the CCTV footage).

9. Explicit instruction during training about maladaptive heuristics people use when making judgments could prove useful.
10. It might be useful to explore the use of oculomotor demonstration training. This is a promising new technique for training perceptual skill. However, further training is necessary to evaluate this technique in forensic contexts.
11. Face matching tasks can be difficult. The current review demonstrates that laboratory studies often find better performance than studies set in more realistic contexts. When evaluating any training programme, care should be exercised when predicting on-the-job performance based on laboratory performance achieved by the end of training. Field trials, more closely matched to actual on-the-job tasks is highly recommended.

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2,* 43–63.

Ask, K., & Granhag, P. A. (2007). Motivational bias in criminal investigators' judgments of witness reliability. *Journal of Applied Social Psychology*, 37, 561-591.

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'Elasticity' of Criminal Evidence: A Moderator of Investigator Bias. *Applied Cognitive Psychology, 22*, 1245-1259.

Balakrishnan, J., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 615-633.

Baldwin, J. (1993). Police interview techniques; establishing truth of proof. *British Journal of Criminology, 33,* 325-352.

Baranski, J. V., & Petrusic, W. M. (2003). Adaptive decision processes in perceptual comparisons: Effects of changes in the global difficulty context. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 658-674.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review, 13,* 1061-1066.

Benton, A. L., Hamsher, K. S., Varney, N. R., & Spreen, O. (1983)*. Contributions to neuropsychological assessment.* New York: Oxford University Press.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94(2),* 115-147.

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*, 378-386.

Bodenhausen, G. V. (1990). Second-guessing the jury: Stereotypic and hindsight biases in perceptions of court cases. *Journal of Applied Social Psychology, 20,* 1112-1121.

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

Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness Identification. In N. Brewer & K. D. Williams (Eds.), *Psychology and Law: An Empirical Perspective.* (pp. 177-221). New York: Guilford.

Brooks, K. R., & Kemp, R. I. (2007). Sensitivity to feature displacement in familiar and unfamiliar faces: Beyond the internal/external feature distinction. *Perception, 36*, 1646-1659.

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

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

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

Burton, A. et al. (2010a) *Bilateral Australia: Solving the passport problem: Re-designing photo-ID to improve recognition.* ESRC End of Award Report, RES-000-22-2519. Swindon: ESRC.

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

Burton, A. M., White, D., & McNeill, A. (2010b). The Glasgow Face Matching Test. *Behavior Research Methods, 42*, 286-291.

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

Busey, T., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45,* 431-448.

Charman, S. D., & Wells, G. L. (2006). Applied lineup theory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), Handbook of eyewitness psychology: Memory for people (pp. 219–254). Mahwah, NJ: Lawrence Erlbaum.

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

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

Collin, C. A., Therrien, M., Martin, C., & Rainville, S. (2006). Spatial frequency thresholds for face recognition when comparison faces are filtered and unfiltered. *Perception & Psychophysics, 68*, 879-889.

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

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high pass and low pass spatial filtering on face identification. *Perception & Psychophysics, 58,* 602-612.

Cutler, B.L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identifications: the role of system and estimator variables. *Law and Human Behavior, 11*(3), 233-258

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44*, 20-33.

da Silva, L., & Sunderland, A. (2010). Effects of immediate feedback and errorless learning on recognition memory processing in young and older adults. *Neuropsychological Rehabilitation, 20,* 42-58.

Davies, D. R., & Parasuraman, R. (1982). *The Psychology of Vigilance.* London: Academic Press.

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

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

DeGutis, J. M., Bentin, S., Robertson, L. C., & D'Esposito, M. (2007). Functional plasticity in ventral temporal cortex following cognitive rehabilitation of a congenital prosopagnosic. *Journal of Cognitive Neuroscience, 19,* 1790–1802.

de Heering, A., & Rossion, B. (2008). Prolonged visual experience in adulthood modulates holistic face perception. *PLoS ONE, 3*(5), e2317.

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

Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavourable medical diagnoses. *Personality and Social Psychology Bulletin*, *29*, 1120–1132.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, *75*, 53–69.

Dijker, A. J. & Koomen, W. (1996). Stereotyping and attitudinal effects under time pressure. *European Journal of Social Psychology, 26,* 61-74.

Dror, I. E., Busemeyer, J. R. & Basola, B. (1999). Decision Making Under Time Pressure: An independent test of sequential sampling models. *Memory and Cognition*, 27, 713-725.

Dror, I. E . & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification, 56,* 600-616.

Dror, I. E., Charlton, D., & Peron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International, 156*, 74-78.

Dror, I. E., Peron, A., Hind, S., & Charlton, D. (2005). When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, *19,*799-809.

Dror, I., & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences, 53*, 900-903.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychologgy, 24*, 419-430.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44,* 576-585.

Duchaine, B. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia, 41*, 713-720.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology, 34,* 590-598.

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

Elstein, A. S., Schulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning.* Cambridge, Mass: Harvard University Press.

Fast, K., Fujiwara, E., & Markowitsch, H.-J. (2005). *Der Famous Faces Test*. Göttingen: Hogrefe.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12,* 195-201.

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

Foster, G., Schmidt, C., & Sabatino, D. (1976). Teacher expectancies and the label "learning disabilities." *Journal of Learning Disabilities, 9,* 111-114.

Francis, R., Riddoch, M. J., & Humphreys, G. W. (2002). 'Who's that girl?' Prosopagnosia, person-based semantic disorder, and the reacquisition of face identification ability, *Neuropsychological Rehabilitation, 12,* 1- 26.

Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception, 19***,** 661-673.

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

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze*. Psychonomic Bulletin & Review, 5,* 490-495.

Gauthier, I., Behrmann, M., & Tarr, M. J. (1999). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience, 11,* 349-370.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology,* 11, 219–224.

Gauthier, I., Skudlarski, P., Gore, J., & Anderson, A. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*, 191–197.

Gauthier, I., & Tarr, M.J. (1997). Becoming a 'Greeble' expert: Exploring the face recognition mechanism. *Vision Research, 37,* 1673–1682.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C. & Anderson, A.W. (2000b). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12,* 495–504.

Gauthier I, Williams P, Tarr MJ, & Tanaka J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. *Vision Research, 38*, 2401–2428.

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

Goffaux, V and Dakin, SC (2010). Horizontal information drives the specific signatures of face processing. *Frontiers in Perception Science , 1* , 1 - 14. , Article 143.

Goldstone, R. L. (1998). Perceptual Learning. *Annual Review of Psychology, 49,* 585-612.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* New York: Wiley.

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

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

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory and Cognition 24,* 26-40.

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

Hildebrandt, A., Sommer, W., Herzmann, G., & Wilhelm, O. (2010) Structural invariance and age-related performance differences in face cognition. *Psychology and Aging, 25,* 794-810.

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

Hochhalter, A. K., & Joseph, B. (2001). Differential outcomes training facilitates memory in people with Korsakoff and Prader-Willi syndromes. *Integrative Physiological and Behavioral Science, 36*, 196-204.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief adjustment model. *Cognitive Psychology, 24,* 1–55.

Hole, G. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23,* 65–74.

Hope, L., Memon, A. & McGeorge, P. (2004) Understanding pre-trial publicity: Pre-decisional distortion of evidence by mock jurors. Journal of Experimental Psychology: Applied, 10, 111-119.

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

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 93_121). Cambridge, MA: MIT Press.

Jarudi, I. N., & Sinha, P. (2011). *Recognizing degraded faces: Contribution of internal and external features.* Unpublished paper downloaded from http://web.mit.edu/bcs/sinha/papers/jarudi_sinha_2005.pdf.

Johnson, K., & Mervis, C. (1997). Effects of varying levels of expertise on the basic level of categorisation. *Journal of Experimental Psychology: General, 126,* 248-277.

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*, 557–571.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kalven, H., & Zeisel, H. (1966*). The American Jury.* Boston: Little Brown

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of face. *Philosophial Transactions of the Royal Society B, 361*, 2109–2128.

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

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

Kimchi, R., & Amishav, R. (2010). Faces as perceptual wholes: The interplay between component and configural properties in face processing. *Visual Cognition, 18,* 1034-1062.

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

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110,* 499–519.

Konar, Y., Bennet, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21,* 38-43.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107-118.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay inferences: Effects of impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 19,* 448-468.

Kuefner, D., Macchi Cassia, V. M., Picozzi, M., & Bricolo, E. (2008). Do all kids look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 811–817.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recogniton of famous faces

Larsen, A., McIlhagga, W., & Bundesen, C. (1999). Visual pattern matching: Effects of size ratio, complexity, and similarity in simultaneous and successive matching*. Psychological Research, 62,* 280-288.

Lavrakas, P. J., Buri, J. R., & Mayzner, M. S. (1976). Perspective on recognition of other-race faces. *Perception & Psychophysics, 20*, 475-481.

Lawson, R., Bülthoff, H. H., & Dumbell, S. (2003). Interactions between view changes and shape changes in picture - picture matching. *Perception, 32*, 1465 – 1498.

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

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

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

Leo, R. A. (1996). Inside the interrogation room. *Journal of Criminal Law and Criminology, 86,* 266-303.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest X-Ray inspection. Journal of Experimental Psychology: Applied, 16, 251-262.

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

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098–2109.

Lundgren, S. R. & Prislin, R. (1998). Motivated cognitive processing and attitude change. *Personality and Social Psychology Bulletin, 24,* 715-26.

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

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

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

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

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339-354.

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

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

Mondloch, C. J., Robbins, R., & Maurer, D. (2010). Discrimination of facial features by adults, 10-year-olds, and cataract reversal patients. *Perception, 39,* 184-194.

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

Mundy, P., & Crowson, M. (1997). Joint attention and early social communication: Implications for research on intervention with autism, *Journal of Autism and Developmental Disorders, 27,* 653 - 676.

Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*, 289–300.

Narby, D.J., Cutler, B.L., & Penrod, S. (1996). The effects of witness, target, and situational factors on eyewitness identifications. In Sporer, S.L., Malpass, R. & Koehnken, G. (Eds.). *Psychological Issues in Eyewitness Identification.* Mahwah, NJ: Lawrence Erlbaum.

Nederhouser M, Yue X, Mangini MC, & Biederman I. (2007). The deleterious effect of contrast reversal on recognition is unique to faces, not objects. *Vision Research.* 47, 2134–2142.

Newhall, S. M., Burnham, R. W., & Clark, J. R. (1957). Comparison of successive with simultaneous color matching. *Journal of the Optical Society of America, 47,* 43-54.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220.

Nisbett, R. E., & Ross, L. D. (1980). *Human inference: Strategies and shortcomings of social judgment*. Edgewood Cliffs, NJ: Prentice-Hall.

Oakhill, J., Garnham, A. & Reynolds, D. (2005). Immediate activation of stereotypical gender information, *Memory & Cognition, 33,* 972-983.

O'Donovan & Bruce, V. (2001).

O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior*, *13*, 385-395.

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

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.

O'Toole, A. J., Roark, D. A., & Adbi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Science, 6,* 261-266.

Page, J., Lafferty, M. E., & Holder, T. (in review). A comparison of expert and novice judges' and coaches' visual search when analysing the handspring vault. *Human Movement Science*.

Parasuraman, R., & Davies, D. R. (1977). A taxonomic analysis of vigilance performance. In R. R. Mackie (Ed.), *Vigilance: Theory, Operational Performance, and Physiological Correlates* (pp. 559-574). New York: Plenum Press.

Pennington, N. & Hastie, R. (I 986). Evidence evaluation in complex decision making. *Journal of Personality & Social Psychology,* 51, 242-258.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 521–533.

Pennington, N. & Hastie R. (1993). A theory of explanation-based decision making. In G.A. Klein, J. Orasanu, J. Caldenvood and D. MacGregor (eds) *Decision Making in Action: Models and Methods*. Norwood NJ, Ablex Publishing.

Pfister, R., & Hillstrom, A. P. (in preparation). *Expert's Eye: Presenting Eye-movement patterns To Train Visual search in Complex Geophysical Displays.*

Pike GE, Kemp RI, Towell NA, & Phillips KC. (1997). Recognizing moving faces: the relative contribution of motion and perspective information. Visual Cognition, 4, 409–437.

Powell, J., Letson, S., Davidoff, J., Valentine, T., & Greenwood, R. (2008). Enhancement of face recognition learning in patients with brain injury using three cognitive training procedures, *Neuropsychological Rehabilitation, 18*, 182- 203.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. *Advances in Experimental Social Psychology,20,* 297–340.

Quinlan, P. T. (1995). Evidence for the use of scene-based frames of reference in two-dimensional shape recognition. *Spatial Vision, 9,* 101-125.

Rabbitt, P., & Vyas, S. (1979). Signal recency effects can be distinguished from signal repetition effects in serial CRT tasks. *Canadian Journal of Psychology, 33,* 88-95.

Read, J. D., Vokey, J. R., & Hammersley, R. (1990). Chnaging photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy confidence relationship. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*, 870-882.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. Psychological Science

Richler, J. J., Gauthier, I., Wenger, M. J. & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 328–342.

Ro, T., Friggel, A., & Lavie, N. (2009). Musical expertise modulates the effects of visual perceptual load. *Attention, Perception & Psychophysics, 71,* 671-674.

Romero, J., Hita, E., & Del Barco, L. Jimenez (1986). A comparative study of successive and simultaneous methods of colour discrimination. *Vision Research, 26,* 471-476.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability*. Psychonomic Bulletin & Review. 16,* 252-257.

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

Ryu, J. J., & Chaudhuri, A. (2007). Differences in attentional involvement underlying the perception of distinctive and typical faces. *Perception, 36*, 1057-1065.

Sadasivan, D., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2005 (pp. 289-300).* New York, NY: ACM Press.

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

Schiff, W., Banka, L., & Galdi, G. D. (1986). Recognizing people seen in events in dynamic mug shots. *American Journal of Psychology, 99*, 219-231.

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction, *Neuropsychologia, 44*, 718-730.

Schretlen, D. J., Pearlson, G. D., Anthony, J. C., & Yates, K. O. (2001). Determinants of Benton facial recognition test performance in normal adults. *Neuropsychology, 15*, 405-410.

Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology, 78,* 655–669.

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

Seipel, A. H. (2002). Face-processing practice: A new tool for improving eyewitness identification. Dissertation Abstracts International, *Dissertation Abstracts International: Section B: The Sciences and Engineering, 63*, 2086.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139-156.

Shepherd, J.W. *et al.* (1982) *Identification Evidence: A Psychological Evaluation*, University of Aberdeen Press.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*, 1202-1212.

Steblay, N. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413-424.

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1397-1410.

Sun, H., Zimmer, H. D., & Fu, X. (2010). The influence of expertise and of physical complexity on visual short-term memory consolidation. Quarterly Journal of Experimental Psychology. In press. Downloaded from http://dx.doi.org/10.1080/17470218.2010.511238.

Swann, W. B., Jr., Giuliano, T., & Wegner, D. M. (1982). Where leading questions can lead: The power of conjecture in social interaction. *Journal of Personality and Social Psychology, 42,* 1025-1035.

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise, *Psychological Science, 16,* 145-151.

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

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective & Behavioral Neuroscience, 9,* 22-131.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457-482.

Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M. D., & Schultz, R. T. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: The Let's Face It! program. *Journal of Child Psychology and Psychiatry, 51,* 944–952.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3*, 764–769.

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

Tesser, A., Martin, L., & Mendolia, M. (1995). The impact of thought on attitude extremity and attitude-behavior consistency. In R. E. Petty & J. A. Krosnick

(Eds.), *Attitudes strength: Antecedents and consequences* (pp. 73–92). Mahwah, NJ: Erlbaum.

Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception, 31(1),* 113−132.

Tickner, A. H., & Poulton, E. C. (1975). Watching for people and actions. *Ergonomics, 18,* 35–51.

Tomasello, M. (1995). Joint attention as social cognition. In Moore, C. and Dunham, P., (Eds.), *Joint Attention : Its Origins and Role in Development.* Lawrence Erlbaum Associates.

Tuchman, B. W. (1984). *The march of folly: From Troy to Vietnam.* New York: Ballantine Books.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning & Verbal Behavior, 20*, 479-496.

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition, 16*, 556-566.

Van Belle, G., de Graef, P., Verfaillie, K., Busigny, T., & Rossion, B. (2010). Whole not hole: Expert face recognition requires holistic perception. *Neuropsychologia, 48*, 2609–2620.

Vickers, D. Burt, J., Smith, P., & Brown, M. (1985). Experimental paradigms emphasizing state or process limitations: I. Effects of speed-accuracy tradeoffs. *Acta Psychologica, 59,* 129-161.

Walker-Smith, G. J. (1978). The effects of delay and exposure duration in a face recognition task. *Perception & Psychophysics, 24,* 63-70.

Warm, J. S., & Dember, W. N. (1998). Tests of vigilance taxonomy. In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.) *Viewing Psychology as a Whole: The Integrative Science of William N. Dember.* (pp. 87-112). Washington, DC, US: American Psychological Association.

Warrington, E. K. (1984). *Recognition Memory Test*. Windsor, U.K.: NFER-Nelson.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20,* 273-281.

Wenger, M. J. & Ingvalson, E. M. (2002). A decisional component of holistic encoding*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 872–892.

Wenger, M. J. & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1106–1118.

Whitman, C. P., & Geller, E. S. (1972). Sequential effects of stimulus probability and prediction outcome on choice reaction time. *Journal of Experimental Psychology, 93*, 373-387.

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces – One element of social cognition. *Journal of Personality and Social Psychology, 99,* 530-548.

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*, 191-196.

Williams, M. A., Moss, S. A., & Bradshaw, J. L. (2004). A unique look at face processing: The impact of masked faces on the processing of facial features. *Cognition, 91,* 155-172.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature, 435(7041)*, 439-440.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology Public Policy and Law, 7*, 119-133.

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology, 10*, 75-84.

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica, 114*, 101-114.

Wright, D. B. & Stroud, J. N. (2002). Age differences in lineup identification accuracy: people are better with their own age. *Law and Human Behavior* , 26, 641–654

Yin, R. (1969). Looking at upside down faces. *Journal of Experimental Psychology, 81,* 141–145.

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5*, 418-441.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14*, 737-746.

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

Yue, Tjan & Biederman, I. (2006). What makes faces special? *Vision Research, 46,* 3802–3811.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology, 20*, 137-142.