Unfamiliar face matching in the applied context

Jennifer McCaffery

PhD

University of York

Psychology

September 2016

Abstract

Matching unfamiliar faces is a difficult task. Despite this, ID checks are the primary screening method for individuals wishing to access countries, employment and a range of financial and medical services. Those we might consider experts, such as passport officers, are no better at the task than general population. Individuals with superior unfamiliar face matching have been identified, but the range of ability remains large across expert and general populations alike. Even individuals with superior face recognition skills have not been consistently found to have superior unfamiliar face matching abilities. This suggests that unfamiliar face matching ability may be highly specific. It may also suggest that the unfamiliar matching tasks carried out in the lab are different from ID checks in the applied context. It is the aim of this thesis to investigate the nature of unfamiliar face matching in the applied context and identify ways in which performance might be predicted. In Chapters 2 and 3 participants are required to match unfamiliar faces shown with a passport context and to check the validity of the accompanying biographical information. The presence of a passport context biases viewers to identify face pairs as the same and presence of a face pair biases and reduces accuracy when checking biographical information. These findings demonstrate that applied error rates in unfamiliar face matching may well have been underestimated. In Chapter 4, a battery of tasks is used to identify predictors of unfamiliar face matching ability. The results show that unfamiliar face matching is positively associated with other face identity tasks. However, same and different unfamiliar face matching also associate with more general measures of local processing and space perception. These findings are tested in Chapter 5 and the theoretical implications of these results and methods for optimising unfamiliar face matching performance are discussed.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Thank you to Professor Mike Burton for giving me the opportunity to complete this PhD and his patience, guidance and good humour.

Thank you also to the ERC for the funding and supporting research in the applied context. I am also very grateful to the PhD students and post docs at Aberdeen and York who provided more support than I could ever have expected.

Finally, and most importantly, I want to thank Heléna Krajewski without whose love and support I would never have stayed the course and my mum who helped me begin.

This thesis is for Margaret McCaffery, Edna Farrelly and all those who were never given the opportunity; they are many.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. The research was funded by a studentship from the ERC.

Chapters 2 and 3: Experiments 3 and 4 from Chapter 2 and Experiment 7 from Chapter 3 have been published in Applied Cognitive Psychology. McCaffery, J. M., & Burton, A.M. (2016). Passport checks: interactions between matching faces and biographical details. *Applied Cognitive Psychology*. Portions of this data was also presented in a poster at the York Applied Face Recognition Meeting, 2015.

Chapter 4: Dr David Robertson provided the adapted version of the Before They Were Famous Test (BFWT: Russell, Duchaine & Nakayama, 2009). Portions of this data was presented at the York Psychology Postgraduate Research Day, 2015 and the North East Person Perception Workshop, 2015.

Chapter 6: The relationship between unfamiliar face matching and the Face Recognition Unit in the unfamiliar face matching model was informed by discussions held in the FaceVarLab and in particular by the work of Professor Mike Burton and Dr Robin Kramer.

Chapter 1 – General Introduction

1.1 Introduction

Our ability to perceive, learn and recognise faces has dominated face perception research for decades. However, in several important occupational contexts it is accurate *face matching,* rather than *face recognition*, which is critical. For example, at UK Border Control we rely on an officer's ability to decide whether an individual's face matches the face-photo on their passport. In policing and the criminal justice system, people are regularly required to decide whether the face of a suspect matches that of an individual caught on CCTV footage. The detection of fraud is a key performance indicator for both Her Majesty's Passport Office and the UK Border Agency (Her Majesty's Passport Office, 2013; UK Border Agency; 2013) In the financial year 2013-14, 5.7 million passport applications were processed, and of these 0.15% (over nine thousand) were detected to be fraudulent (Her Majesty's Passport Office, 2014). In the same year, 241 million passengers travelled via a UK airport (Civil Aviation Authority, 2014). 2014 also saw the detection of fraudulent passports within the EU/Schengen become greater than the detection of fraudulent passports from outside the EU/Schengen area (9,968 and 9,400 respectively; Frontex, 2015). These volumes show just how important identifying occurrences of fraud are and the size of the task facing governments and agencies. Outside the forensic or security context, proof of identity is required for a wide range of commercial and social transactions including employment, accessing benefits and applying for loans. In 2015-2016 identity theft is estimated to have cost UK economy £5.4 billion (Experian, PKF Littlejohn & University of Portsmouth, 2016). These figures only represent fraud that is identified. The true cost of fraud, in both commercial and human terms, may never be known. However, it is unlikely to be less than estimated.

It is concerning then that a widely used test of unfamiliar face matching, the short form of the Glasgow Face Matching Test (GFMT: Burton, White & McNeill, 2010), has shown a mean performance of 81.3% (SD = 9.7) with a range of 51%–100%. If these figures were reflected in real passport checking rates, this would mean that of the 241 million passengers travelling via a UK airport in 2014, passport officers would not be able to identify if 45 million of those passengers were travelling with their own passport, potentially allowing fraudulent passport holders into the country. The substantial individual differences found by Burton et al., (2010) could also mean that dependent on the locations of personnel some airports could be at risk of even greater volumes of fraud being missed. However, this would only be the case if individuals working in forensic and security settings were as vulnerable to error as the naïve viewers tested by Burton et al., A recent study in which passport officers completed the short form of the GFMT found that their mean performance was 79.2% (SD = 10.4) which did not differ significantly from the normative scores (White, Kemp, Jenkins, Matheson, Burton, 2014a). The passport officers also showed large individual differences in performance but no relationship between experience, or time on the job, and accuracy. This was surprising, as it would not be unreasonable to expect that individuals who complete unfamiliar face matching tasks day to day would show some improvement over time. Passport officers appear equally vulnerable to errors in unfamiliar face matching as the general public.

2016 saw the Federal Bureau of Investigation (FBI) identify identity theft as a key threat to transnational security (FBI, 2016) and the UK launch a Joint Fraud Task Force (Home Office, 2016). It has become increasingly clear that any methods available to reduce the threat and impact of identity fraud should be identified and implemented. Given that unfamiliar face matching appears difficult for naïve and experienced viewers alike, any research that has identified why this task might be so difficult and how performance might be improved would be extremely valuable.

The applied context of face processing has been fundamental to the progression of research, but its relationship to theoretical work has been far from comfortable. In their review of face recognition as a specific ability, Wilmer, Germine and Nakayama (2014) identify how the lack of a relationship between face recognition ability and more general measures of intelligence led to face recognition becoming a relatively neglected area of research. The George Washington Social Intelligence Test (GWSIT: Hunt, 1928) was designed to identify relationships between social intelligence and general intelligence across a range of students and individuals employed in different industrial groups. The test included an unfamiliar face recognition task, which ultimately dissociated from the other tasks in the test and from the measure of general intelligence. Instead of being recognised as a valuable measure of independent factors crucial to general and particularly social ability, the GWSIT was cited as an invalid test (Campbell & Fiske, 1959) and discarded with other measures of social intelligence (Kihlstrom & Cantor, 2000).

This relegation of face recognition as an invalid measure of social intelligence resulted in little theoretical investigation until the 1970s. In 1975, Bahrick, Bahrick and Wittlinger published a study investigating familiar face recognition over 50 years. Their purpose was not to investigate face recognition specifically, but long-term memory, and they identified that photographs in a yearbook could provide a uniform set of stimuli that might be remembered years later by members of that year. Recognition performance remained at 90% even when participants had graduated 15 years prior. This was a particularly interesting finding when compared to work being carried out in eyewitness research, which was driven by concerns about the accuracy of eyewitness identification. Many studies showed that viewers were extremely susceptible to changes in pose, dress and appearance when recognising unfamiliar people (for review see: Wells & Olson, 2003). Patterson and Baddeley (1977) for example recorded a reduction in recognition rate to less than chance when changes were made in images of unfamiliar faces to pose, dress and appearance. This concern was later confirmed when more than 75% of the 100 people convicted prior to the advent of forensic DNA, and later exonerated by DNA tests, were found to be convicted primarily on the basis of mistaken eyewitness identification (Wells, Small, Penrod, Malpass, Fulero & Brimacombe, 1998; Scheck, 2000).

Eyewitness research led more theoretical face recognition researchers to pay closer attention to their methods and findings. Applied face recognition moved from being a barrier to face research to an enabler, which has led to over 40 years of extremely fruitful research into face recognition. In contrast, face matching is a much more recent area of investigation, driven from a theoretical perspective. Researchers began using face matching tasks to investigate how well viewers could identify images of individuals across different types of viewing conditions without the burden of memory (Bruce et al., 1999). It had been assumed that problems in recognising unfamiliar individuals were due to issues with memory; however, the 80% accuracy achieved in the unfamiliar face matching task demonstrated that matching unfamiliar faces was difficult, even when the faces were shown simultaneously without time constraint. Since unfamiliar face matching may well be a much more frequent occurrence in the forensic or security settings than unfamiliar face recognition it is perhaps surprising that it has received very little consideration in the applied context. Researchers have begun to investigate the performance of those individuals that we might consider experts at the task e.g. police officers or passport officers (Burton, Wilson, Cowan & Bruce, 1999; White et al., 2014a). However, there has been less investigation of the face matching tasks carried out in the applied context. Precisely why face matching in the applied context is so difficult is the focus of this thesis. This introduction aims to provide an overview of face matching and its relationship with face recognition, and the research that has been carried out into how individuals succeed and fail in face matching tasks and how performance can be improved. Fundamentally I review the assumptions that have compromised the progress that has been made in face matching to date: that familiar and unfamiliar faces are processed in the same way; that face recognition and face matching call on the same set of processes; and that everyone is an expert with faces.

1.2 Familiarity

An expectation that participants would excel at unfamiliar face matching may in part be due to the lack of memory demands in the task but also because participants excel at familiar face matching as they do at familiar face recognition. As discussed earlier, individuals have been shown to recognise familiar individuals in photographs with 90% accuracy even if the photographs were taken 15 years previously (Bahrick et al.,1975). However, these high levels of ability do not translate to recognising unfamiliar faces. Klatzky & Forrest (1984) asked participants to view sets of familiar and unfamiliar faces and then view these sets mixed with faces they had not been shown previously. As they viewed these mixed sets, the participants were asked to indicate whether they recognised the faces from the earlier part of the experiment. They were almost 90% accurate in identifying familiar faces they had been shown before but only 70% accurate in identifying the unfamiliar faces they had been shown before. Even when face recognition tasks are carried out using poor resolution in CCTV footage from different viewpoints, viewers who were familiar with the individuals shown maintained a much higher recognition accuracy than viewers who were unfamiliar with the individuals shown (Burton et al., 1999a).

Clutterbuck and Johnson (2002) demonstrated that while participants matched familiar faces at high degrees of accuracy, they were only 80% accurate when matching unfamiliar faces. These high levels of accuracy for familiar faces are retained even when the quality of images are degraded as those found in CCTV, whereas unfamiliar face matching accuracy falls even further. In a face matching task where participants were required to match photographs with video stills in which the faces were degraded, participants averaged over 90% accuracy for familiar faces and less than 70% for unfamiliar faces (Bruce, Henderson, Newman & Burton, 2001). Unfamiliar face matching like unfamiliar face recognition is highly susceptible to changes between the images being matched e.g. viewpoint, expression and time (Bruce et al.,1999a; Megreya, Sandford & Burton, 2013). Even viewed in good lighting conditions and with unlimited time, individuals perform poorly on both recognition and matching tasks with unfamiliar faces

(Klatzky & Forrest, 1984, Megreya & Burton, 2006). Participants also perform poorly whether the faces to be recognised or matched are viewed as still photos, video clips or live individuals (e.g. Klatzky & Forrest, 1984; Megreya & Burton 2006; Burton et al., 1999a; Bruce et al., 1999; Brown, Deffenbacher & Sturgill, 1977; Kemp, Towell & Pike, 1997). Eyewitness testimony has also been shown to be highly error prone (for review see Davis & Valentine, 2015). Even in sorting tasks where participants are asked to sort a mixed pile of photographs into two identities; recognising that there are only two identities is perfectly straightforward when participants are familiar with the identities. However, when participants are unfamiliar with the identities the task is exceptionally difficult, with participants sorting the photographs into seven piles on average (Jenkins, White, Van Montfort, & Burton, 2011). Across all types of face processing tasks and across a wide variety of viewing conditions familiarity is a powerful moderator of face processing abilities.

1.3 Within Person Variability

The difficulties in recognising, matching and sorting unfamiliar faces are due in part to within person variability and secondly due to the image bound nature of unfamiliar face processing (Hancock, Bruce & Burton, 2000). Within person variability describes the ways in which images of the same person can vary. Jenkins et al., (2011) demonstrate that there is a great deal of variation not only between photos of different individuals but also within photographs of the same individual.

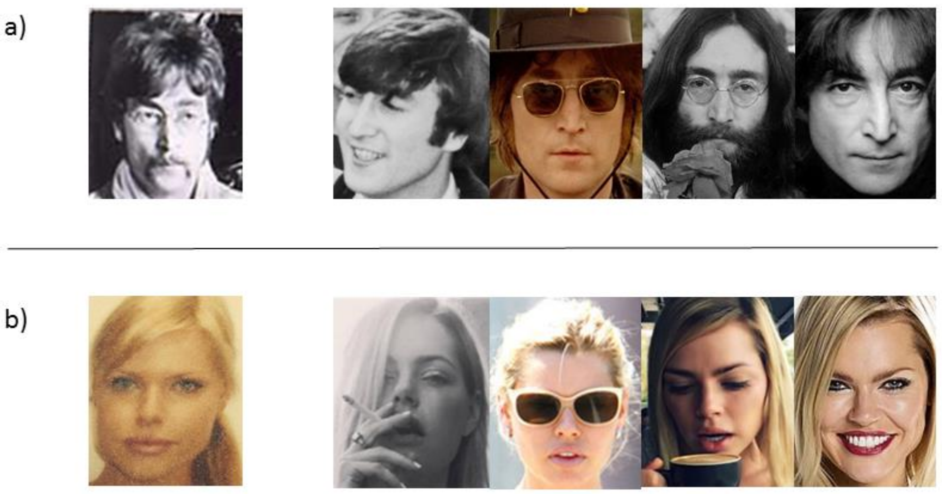

Figure 1.1. Photos taken from ID (passport) on the left and images of the same individuals in different contexts on the right. The individual in the top line will be familiar to a great

number of people as he is John Lennon and very well known. The individual in the bottom line is a celebrity in Australia but will be unfamiliar to the majority of people in the UK.

As figure 1.1 shows, it is very easy for viewers familiar with John Lennon to match the passport photo on the left with each of the photographs on the right taken of John across his lifetime. However, for viewers who are unfamiliar with the woman shown on the bottom row, it is much more difficult for them to match the passport photo on the left with all of the photographs on the right. The images differ in terms of lighting, pose, expression, age, quality, and whether individuals are wearing glasses or holding objects that occlude or obscure part of the face.

Even when images of the same individual are highly constrained by photo ID requirements e.g. neutral pose and distance from the camera, viewers have been shown to have great difficulty identifying three photo IDs from the same unfamiliar individual as the same person (Bindemann & Sandford, 2011). Our ability to recognise familiar individuals across a wide variety of images and instances is generally so high and so intuitive it is very difficult for us to imagine that anyone unfamiliar with the individual could possibly be unable to recognise or match images of them (Ritchie, Smith, Jenkins, Bindemann, White & Burton, 2015). This systematic bias may well explain why researchers have continued to conflate familiar and unfamiliar face processing in their research and missed the large role that within person variability plays in our day to day interaction with faces (for review see Burton, 2013).

1.4 Image Bound Processing

Unfamiliar face processing is also more difficult than familiar face processing because it is image bound (Hancock et al., 2000). During unfamiliar face recognition tasks viewers very often learn an identity from a single image and are asked again to recognise the identity from a target image during testing. In the applied context, this learning mechanism is frequently used with single images of individuals shown on watch lists for border officers to remember and recognise at a later date. Unfamiliar face matching is often as constrained with viewers trying to match a single image with a target image or person shown alone or in an array; both in the lab and the applied context. Figure 1.2 provides two examples of the task. Each example has a token image on the left where the identity is known and two comparison images on the right, one of which is the target (another image of the known identity) and the other the foil (an image of a different identity). In most applied face matching tasks there would usually only be a token image and a comparison image, however border control officers may also be provided with an array or candidate list of images by a face recognition system from which they must identify the target (Heyer & Semmler, 2013).

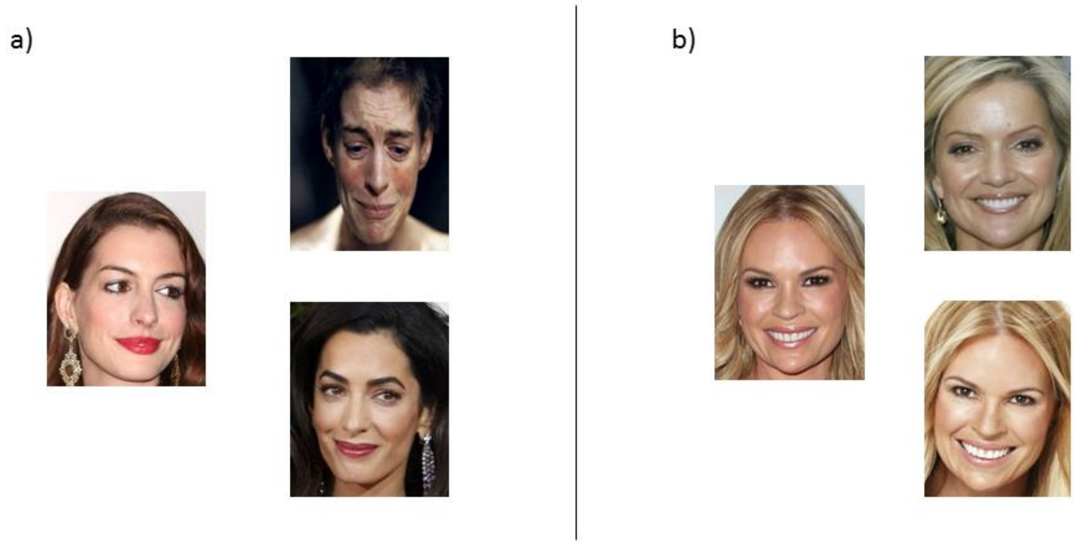

Figure 1.2. Two examples of a face matching task. In example a) the token image on the left is a photograph of Anne Hathaway; the two comparison images on the right are the target image of Anne Hathaway (above) and the foil image of Amal Clooney (below). In example b) the token image on the left is a photograph of Sonia Krueger; the two comparison images on the right are the foil image of Sandra Sully (above) and the target image of Sonia Krueger (below).

The image bound nature of unfamiliar processing can be seen in the contrast between the two versions of the task. In task a) a photograph of Anne Hathaway is provided as the token image. Anne Hathaway is a global celebrity who is also well known for playing the role of Fantine in the 2012 Les Miserables film (Bevan, Fellner, Hayward, Mackintosh & Hooper, 2012). She is particularly recognised for the scenes in which she has her head shaved as she is shown in the target image, the higher image of the two comparison images. The foil image below is a photograph of Amal Clooney who is well known for her human rights work as well as marrying George Clooney, a global celebrity. The target image of Anne Hathaway varies a great deal in pose, hairstyle, makeup, lighting and expression from the token image. If we were not familiar with her as an individual, we might be more likely to select the image of Amal Clooney as the target. However, because we are familiar with her we are much more likely to select the image above as the target.

Matching familiar faces can be achieved in at least three ways. We could be comparing the ID photo with both of the comparison photos to see if the individual were the same, however this would not account for higher levels of performance for familiar individuals. We could be performing an almost an image based recognition task where we are searching our memory for an identical match or similar match with any of the images we store of Anne Hathaway and overriding the matching part of the task with the recognition of the image task – 'I recognise the person in the token image as Anne Hathaway. Does either of the images on the right match an image I have seen before of Anne Hathaway?' We may have seen the film and recognise the image as taken from the scene. Thirdly, we could be comparing each of the ambient images with a higher-level representation of Anne Hathaway (Bruce & Young 1986; Burton et al., 1999a) which also contains all versions of Anne Hathaway we have seen (Kramer, Ritchie & Burton, 2015) and links to our biographical knowledge of Anne Hathaway (Bruce & Young, 1986). In this instance, we are not looking for a matching image but that the comparison image is a good fit with a representation of Anne Hathaway's face, a fit that could also be supported by our biographical knowledge that she appeared in this way in this film even though we have not seen the images. In the second and third methods we bypass the matching task through a simple recognition of the target image as Anne Hathaway – a task that can be achieved whether we are familiar with Amal Clooney or not.

Being familiar with both identities, Anne Hathaway and Amal Clooney would also provide greater confidence in our decision since we recognise both of the individuals shown in the comparison images. Alternatively, if we were not familiar with Anne Hathaway but we were familiar with Amal Clooney this provides a further strategy for completing the task, but only if we know that the target identity is present in the comparison images. We could bypass the matching task again through deduction; the individual on the bottom right is Amal Clooney – neither the token image on the left or the comparison image above is Amal Clooney so by default these images match. However, if we do not know that the target identity is present in the comparison images then knowing who one of the images only reduces the chances of being wrong. A matching task will still have to be carried out between the token image and the other comparison image and this will have to be done like task b) at the level of the images.

It is clear that being familiar with at least one identity in a matching task offers a number of strategies for providing the correct response – the majority of which use recognition. However, task b) demonstrates how difficult this task can be with unfamiliar identities and even when the photographs appear to be taken under very similar conditions. The token image on the left shows Sonia Krueger an Australian celebrity who is not well known outside Australia. Of the comparison images, the image below is the target image, showing Sonia Krueger. The foil image below shows Sandra Sully an Australian celebrity who is not well known outside Australia but looks very similar to Sonia Krueger. If we were familiar with either of these celebrities, we would know that these identities are very similar in their appearance but that this photograph of Sandra Sully is taken from a viewpoint in which they look even more similar than usual. As a viewer we are reliant solely on comparing these three images alone and are disadvantaged by the images chosen. Neither unfamiliar face recognition nor matching provides us with this type of advantage or the option to use one of multiple strategies to match or recognise a face. Taken together the image bound nature of unfamiliar face processing and within person variability not only demonstrate how difficult unfamiliar face processing is but also how powerful familiarity is as a moderator of our face processing abilities.

1.5 Models of familiar face recognition

There have been a number of models proposed to explain or illustrate the mechanisms underlying familiar face recognition. Examining these models can provide a useful framework for identifying the limitations and qualitative differences of unfamiliar face recognition and matching.

Cognitive Model

One of the most influential cognitive models of unfamiliar face recognition was proposed by Bruce and Young (1986) (see figure 1.3). On viewing an unfamiliar face, a pictorial code or description of the viewed image of a face is generated. This code includes both specific and abstract information including face shape, lighting conditions, pose and, is, Bruce and Young (1986) assert, the level at which recognition can be achieved using unfamiliar faces in recognition yes/no tests in the laboratory. However, Bruce and Young are also very clear in noting that the pictorial code does not provide enough information for faces to be recognised despite changes in age, weight, hair colour etc. and it is for this reason that accuracy in recognising unfamiliar faces is far less than achieved when recognising familiar faces (Bruce, 1982). These greater levels of accuracy are achieved through structural encoding - a higher level of coding achieved through repeated exposure to a face, whether as an image or in reality. It is this detailed information, the structural code, which leads to the creation of the Face Recognition Unit

(FRU). As semantic information is learned about the individual, e.g. age, gender or occupation a Person Identity Node (PIN) is created which is accessed via the FRU enabling perceptual and cognitive information about a person to be combined. It is only when a PIN is created that true familiarity is achieved – when we know to whom a face belongs. The model also accommodates the particular difficulties viewers experience with retrieving a name despite being able to recall many other details about a person, also known as the tip of the tongue phenomenon (Yarmey, 1973). Here, the names are only accessible via the semantic information held about an individual, rather than directly, reflecting the theories of Warren and Morton (1982) and findings of Young, McWeeny, Ellis and Hay (1986). Though it is not the focus of this thesis, it is also important to note that the routes for recognising expression and identity are also shown as separate pathways, as is the dominant view but which is not without contention (for review see Calder & Young, 2005).

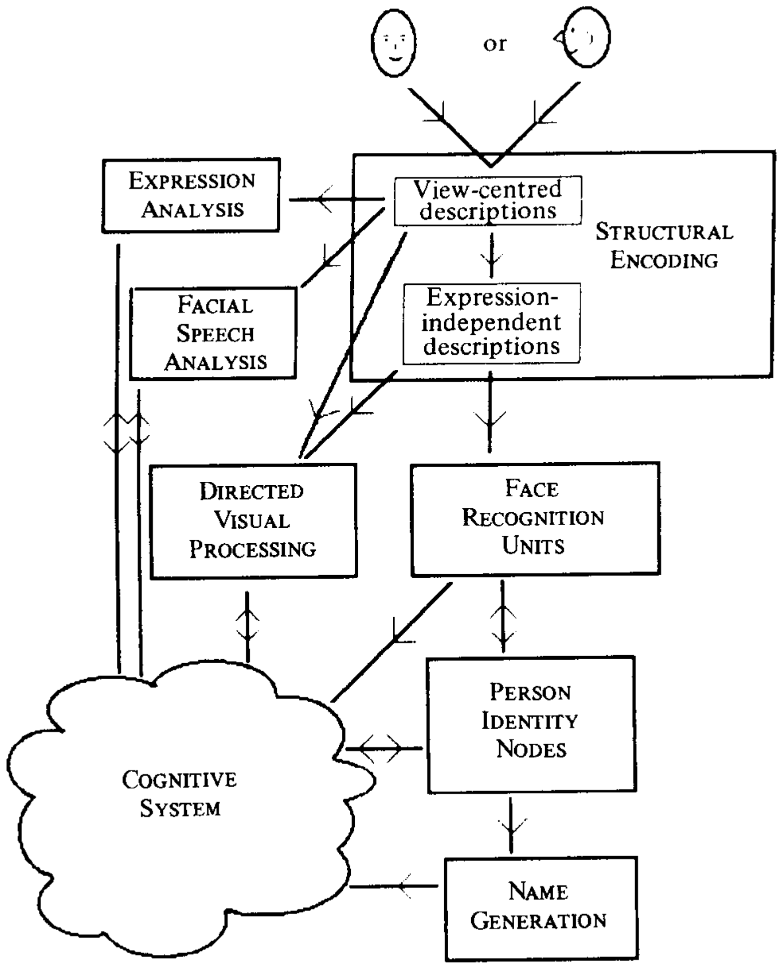

Figure 1.3. A cognitive model of face recognition, reproduced from Bruce and Young (1986).

Perceptual Model

Bruce and Young (1986) provided a functional framework for familiar face recognition specifically focusing on the relationships between different cognitive functions. This model has become the dominant account enabling researchers to investigate and test hypotheses within a working context. However, little was said regarding how these functions took place, in particular with regard to perception and the structural encoding of faces. Perhaps the most influential perceptual model of face recognition comes from Valentine (1991). Valentine proposed that faces were encoded as points in a multidimensional space. Two possible models were identified, the norm based model and the exemplar based model. In the norm based model faces were proposed to be located in space relative to an average or norm that is constantly updated as more faces are learned. In the exemplar model, an average or norm is not extracted but the faces are placed in space relative to each other. Since the multidimensional space is defined as the central tendency of the dimensions used, and these dimensions are gained through holistic encoding of each face, the models can be used interchangeably. Within both models, faces seen more frequently or faces that are more typical will appear closer to the central tendency than faces that are more distinctive or the faces seen less frequently, as shown in figure 1.4.

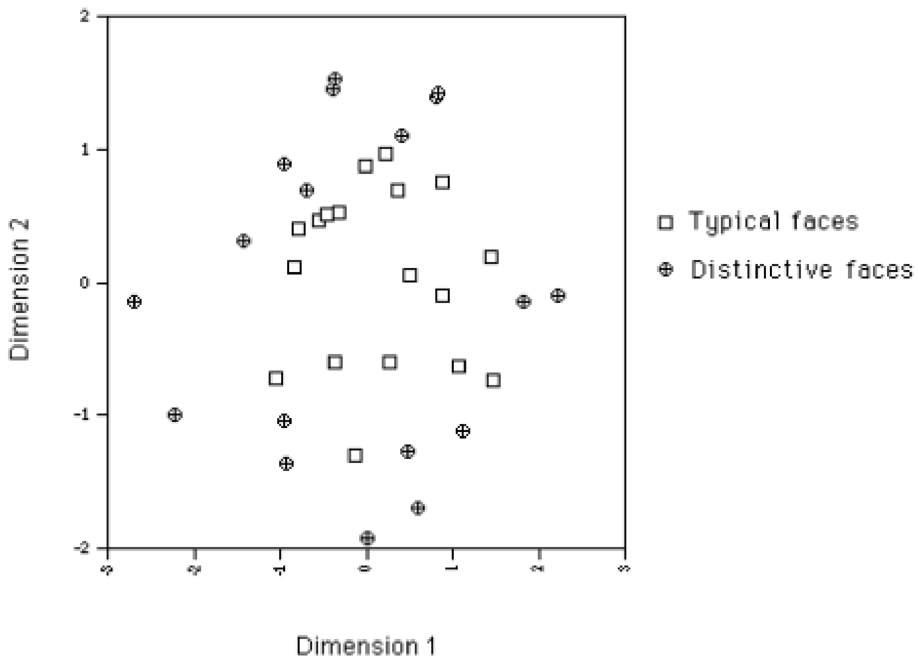

Figure 1.4. A perceptual model of face recognition, reproduced from Valentine (2001; citing Johnston, Milne, Williams and Hosie, 1997).

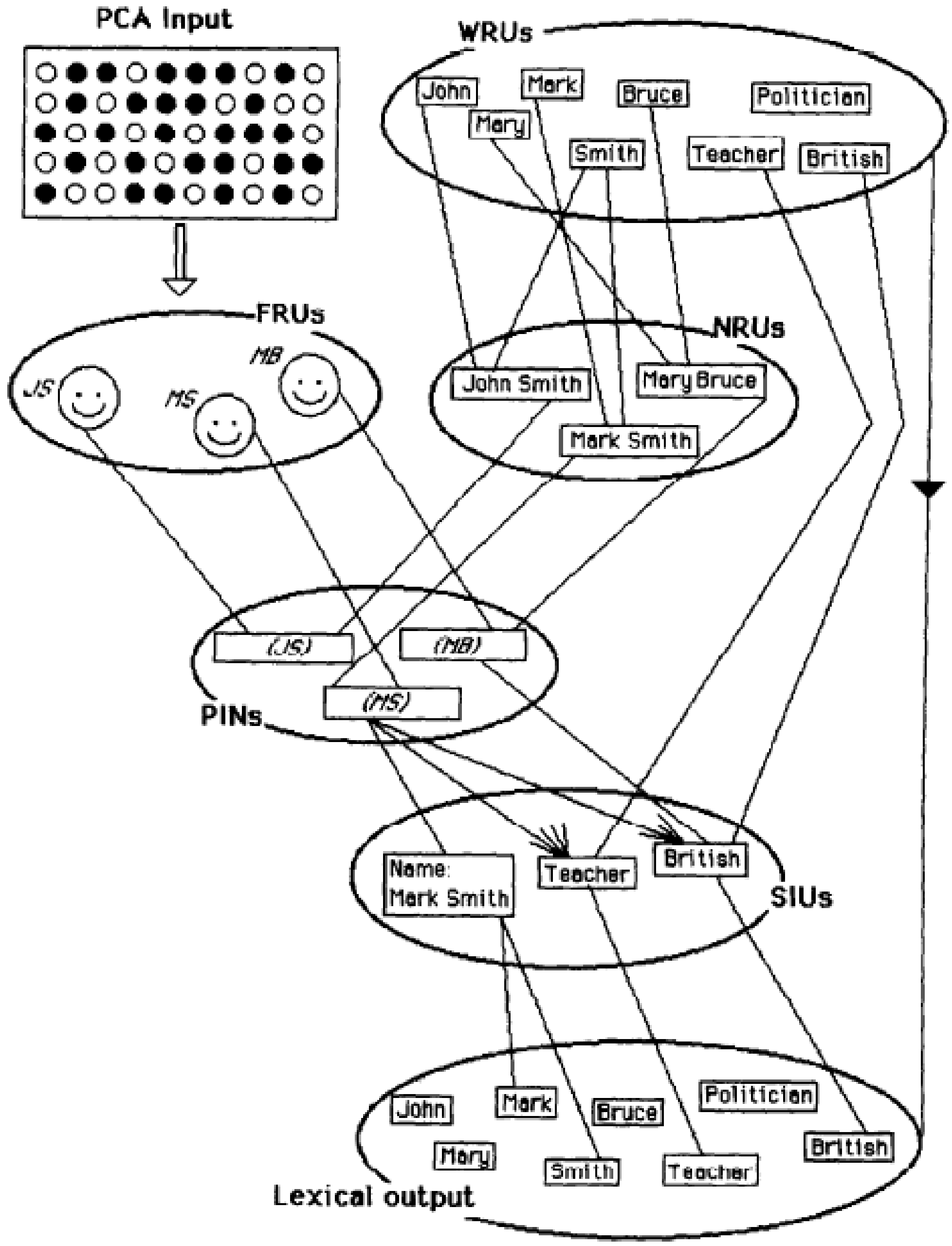

Figure 1.5. IAC model of face recognition with PCA units, reproduced from Burton et al., (1999b).

The interactive activation and inhibition (IAC) model developed by Burton, Bruce and Johnston (1990), clarified, and extended by Burton and Bruce (1993) and Burton, Bruce and Hancock (1999b) brought together the perceptual and cognitive models proposed by Bruce and Young (1986) and Valentine (1991). First developed in 1990 the (IAC) extended the cognitive model presented by Bruce and Young. The box and arrow model did not provide a great deal of information about how information is pooled together about an individual and might infer that only serial processing was possible, with one type of processing only taking place after another is complete. The IAC proposed a cascade architecture for accessing and storing semantic information where semantic knowledge is presented as a hierarchy in which name is not the predominant association, but a name can also be accessed without first retrieving all other information about the individual.

As can be seen in figure 1.5 the FRU for each individual is located within a pool of other FRUs connected to each other by inhibitory connections. The PIN is no longer where the semantic information is held but where all other familiar perceptual triggers are held, such as voice, sound of name, written name. These PINs are also located in a pool with other PINs connected by inhibitory connections. The semantic information is held within Semantic Information Units (SIUs) which hold information such as age, gender and name all as separate pools within the SIU connected via inhibitory connections within each pool. Connections are made between PINs and SIUs via excitatory connections allowing bidirectional associations to take place. Burton and Bruce (1993) give the example of Prince Charles whose PIN might be activated by an image – the FRU, which in turn would activate many SIUs e.g. age = 60 – 70, gender – male, occupation – Royal, Name – Charles at the same time. Conversely, the occupation – Royal, would trigger a large number of SIUs and related PINS and FRUs as connections are processed in parallel. The bigger SIUs being more immediate in their activation, but potentially requiring greater lengths of time processing to inhibit the other Royals that are triggered before the much smaller SIU for the name Charles and then Prince Charles is attended to. This suggests that the more common, frequently heard/seen or familiar the information relating to an individual, the more easily their PIN / FRU and associated SIUs will be activated since the lowest number of pools would require activation (Burton & Bruce, 1992, Carson & Burton, 2001). The relatively unique nature of names would require multiple sub pools within the SIUs across which connections must be located and activated, delaying processing.

Burton et al., (1999b) further extended the IAC model to include a perceptual module. Using a face space model, Burton et al., (1999b) identified that image pixels provided dimensions that could be used via principal component analysis to create a multidimensional space in which individual faces could be located. This perceptual module precedes the cognitive modules as the structural coding preceded the FRUs in the Bruce and Young (1986) model. However, the cascade architecture importantly allows for multiple types of information to be accessed and retrieved without waiting for each type to be returned before moving on to the next.

Neural model

These models, and in particular the Bruce and Young (1986) model, are still the dominant account of human familiar face recognition and have formed the basis for investigating and mapping the neural systems underlying familiar face recognition. Figure 1.6 shows how Gobbini and Haxby (2007) mapped their distributed neural system for face recognition (Haxby, Hoffman & Gobbini, 2000) using the Bruce and Young model.

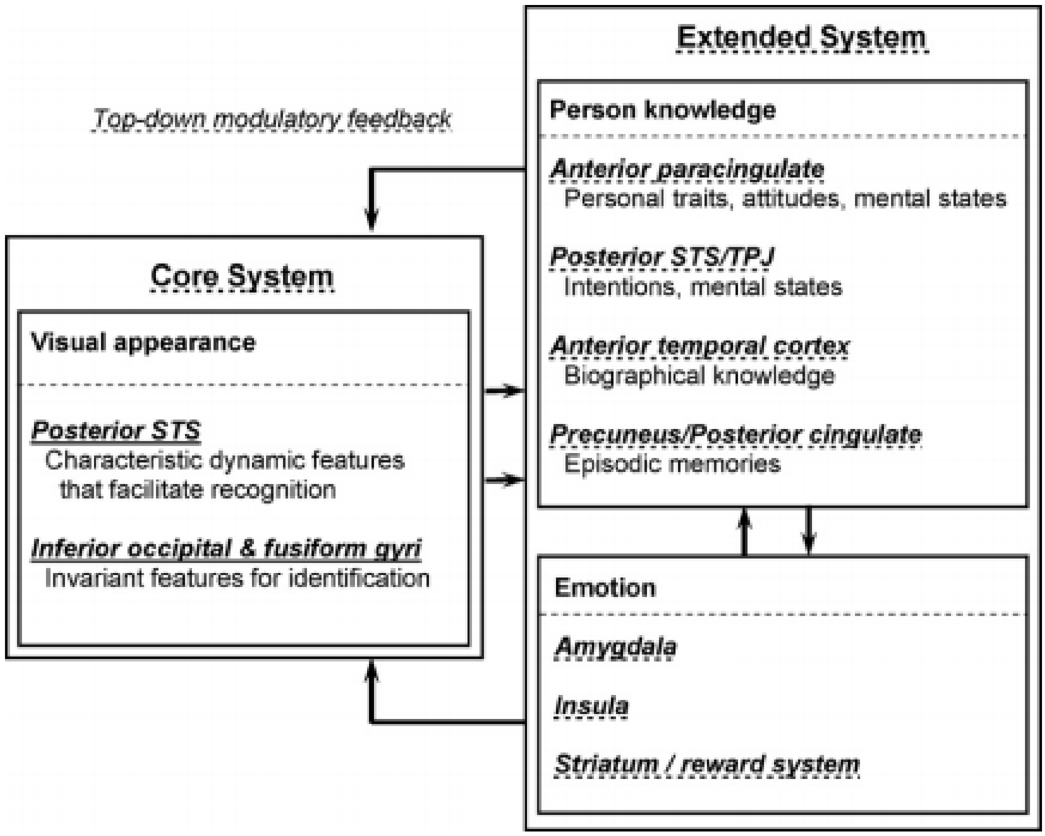

Figure 1.6. A model of the distributed neural areas that mediate familiar face recognition. (Gobbini & Haxby, 2007).

This model clearly identifies the structures involved in perceptual (the core system) and cognitive (the extended system) familiar face recognition and has recently been extended by Duchaine & Yovel (2015) to define the face selective areas more specifically in relation to dynamic as well as static faces.

Taken together one could argue that these models tell us very little about unfamiliar face recognition, let alone unfamiliar face matching. The recognition of unfamiliar faces is often done with very little, if any, semantic information. Therefore, the extended or cognitive system that is so important to familiar face recognition is relatively redundant in the recognition of unfamiliar faces. Bruce and Young (1986) are clear that unfamiliar face recognition uses pictorial codes only. The perceptual models used by Valentine (1991, 2001) and Burton et al., (1999b) are perhaps more useful as one would presume a face would have to be represented in multidimensional space to be recognised whether the face was familiar or learned (unfamiliar). However, these models and this use of them does not explain the difference in accuracy achieved when recognising familiar and unfamiliar faces. If each face is only a point in space, then both familiar and unfamiliar faces are points in space, how does this account for the difference in accuracy between recognising or matching familiar and unfamiliar faces? The model does not represent the difference between the rich structural codes of familiar faces and the relatively limited pictorial codes of unfamiliar faces.

However, Burton, Kramer, Ritchie and Jenkins (2016) have used the PCA method developed by Burton et al., (1999b) to map within person variability for faces. When running a PCA analysis on a number of celebrities, Burton et al., (2016) mapped multiple instances of each individual's face in multidimensional space creating face regions for each individual rather than a point in space. This allows the variability within each individual to co-exist with the variability between individuals. It was also clear that the variance within individuals was idiosyncratic which would allow a rich representation of that individual to be held by the face region specifically for that individual. Even using 30 images not all of the variability within each celebrity was captured suggesting that as in the applied setting, greater exposure to an individual would create a richer representation and more accurate recognition of old or new images of that individual. This within and between person variability model provides a working explanation of why unfamiliar face recognition is less accurate than familiar face recognition. With only limited images, face regions for unfamiliar people will capture only a very small amount of the within person variability for that person. This allows greater room for error if there is a great deal of variability between the image or individual shown for comparison and the token image. Alternatively, if a face is viewed in highly variable conditions this should also lead to a richer face region and it has been shown that the more variability there is in a set of faces for an individual to be learned, the better they are remembered (Ritchie & Burton, 2016). Familiar face regions will be much richer encapsulating changes in viewpoint, hair, makeup and lighting for example. If unfamiliar face matching is able to use the perceptual systems used by unfamiliar face recognition; the within and between person variability model may also provide a working model of unfamiliar face matching. This model may allow unfamiliar face matching to be better understood and identify ways in which unfamiliar face matching can move away from being so image bound.

1.6 The specific nature of face processing

The discarding of the GWSIT (Hunt, 1928) because of the dissociation between the face recognition test and other measures was not the last time face recognition was dropped from more general measures of ability. As Wilmer et al., (2014) note, a face recognition subtest was introduced as part of the third version of the Wechsler Memory Scale (WMS-III: Wechsler, 1997). As the WMS is one of the most widely used measures of memory the face recognition subtask added was an unfamiliar face recognition task in which participants were required to learn a set of faces and subsequently identify them as 'old' (seen before) or new (not seen before). However, again, because of its dissociation with the other subtasks (Millis, Malina, Bowers & Ricker, 1999) the face recognition subtask was dropped from the WMS-IV (Wechsler, 2009). This dissociation from other general measures such as memory and intelligence would certainly suggest that face recognition is quite specific.

Behaviourally, recognising faces has also been seen to be quite different from recognition of other objects. Yin (1969) demonstrated that when stimuli were learned and presented upright for recognition, participants identified faces with much more accuracy than houses, aeroplanes or men in motion. However, when the stimuli were learned and presented inverted, participants identified faces with much less accuracy than houses, aeroplanes or men in motion. This susceptibility to inversion, or the inversion effect (IE), has been shown repeatedly for faces when being recognised whether familiar or unfamiliar and when being matched (Valentine & Bruce, 1986; Freire, Lee & Symons, 2002; Mondloch, Le Grand & Maurer, 2002). The IE has been associated with faces being recognised or matched configurally (using the relations between features) or holistically (as a gestalt or whole) rather than analytically like objects (Maurer, Le Grand & Mondloch, 2002; Harel, 2016). Performance on memory tasks for faces and classes of objects has also been compared. Dennett, McKone, Tavashmi, Hall, Pidcock, Edwards & Duchaine (2012) developed the Cambridge Car Memory Test (CCMT) in the same format as the widely used Cambridge Face Memory Test (CFMT: Duchaine & Nakayama, 2006) so that unfamiliar recognition of cars and faces could be compared. Overall performance for recognising individual cars was lower than performance for recognising individual faces. This suggests that faces are better recognised than objects.

Within the brain imaging literature there is also significant evidence for the specific nature of face recognition. Structures such as the Fusiform Face Area have been identified as highly specific and preferentially responsive to face stimuli (Sergent, Ohta & MacDonald, 1992; Kanwisher, McDermott & Chun, 1997; Kanwisher & Yovel, 2006; Kanwisher, 2010). This work has built on the studies in neurological and clinical research of 'Face blindness' or prosopagnosia that have provided valuable insights into the complex processes underlying face recognition. First documented in the 19th century, John Hughlings Jackson identified a patient in 1876 who often mistook her niece for her daughter. However, he also diagnosed her as experiencing a more generalised imperception disorder – rather than a condition exclusively affecting faces (Finger, 2001). It was only during the mid-20th century that the term prosopagnosia was used by Joachim Bodamer to describe patients with specific impairments in processing visual information about faces. His case notes describe three separate individuals, with three very different experiences of face processing. The first patient (S) was able to see and classify a face as separate from other objects but was unable to name the owner of the face, whether newly learned or members of his family, nor was he able to see the face as a whole, pointing only to individual features and naming them. The second patient's (A) experience with faces was even more limited with faces appearing as a blur apart from a focal point about the size of an eye that he could move about the face to see individual characteristics. Patient A was also unable to identify individuals by their faces from his past or the present, known or unknown, or even his reflected image in a mirror. The third patient's (B) case was very different with this patient able to recognise and name faces; his own, his family and the newly acquainted nurses but all images of faces appeared grotesquely distorted with features squashed together or misaligned (Ellis & Florence, 1990).

It is important to note that these cases had definitive differences but were presented under the umbrella term of prosopagnosia. Even at this early stage, the differences between the patients' experience of processing faces hinted at what has become a vast and wide-ranging area of research. While prosopagnosia has been identified as heterogeneous with many subtypes, e.g. individuals born with prosopagnosia (congenital prosopagnosia) and those who became prosopagnosic as a result of a brain injury (acquired prosopagnosia) there is strong evidence that prosopagnosia is a disorder that is specific to faces (Susilo & Duchaine, 2013a; Susilo, Yovel, Barton & Duchaine, 2013b). Together with the brain imaging and behavioural literature, this research supports a face specificity hypothesis (Kanwisher & Yovel, 2006), underlines the importance of having face specific models for recognition, and potentially face matching. The specific nature of face recognition and face matching would also suggest that in the applied context individuals could only be selected for face recognition or face matching roles using face specific tasks and programs. If face recognition and matching are also very automatic processes it might also suggest that training may not be available that can improve performance. However, not all researchers agree that face recognition is a specific process using neural systems that have been optimised solely for faces.

1.7 The expertise hypothesis

The expertise hypothesis proposes that the system used for recognising and matching faces is not optimised 'out of the box' for faces rather it is a general system for recognising and matching individual objects within categories that has become very good at face recognition and matching because faces are the most viewed category of objects. Diamond and Carey (1986) ran a seminal behavioural study in which novices and dog experts were required to recognise dogs and human faces in upright and inverted conditions. The dog experts showed an inversion effect for the dogs which was as large as the inversion effect for faces. The novices showed no such inversion effect for the dogs but a large inversion effect for the faces. Diamond and Carey (1986) proposed that it was not that faces were special or used specific systems or processes but that it was expertise that was special. Expertise allowed viewers to use configural processing when viewing the dogs; a process which was disproportionately affected by inversion. Evidence of the expertise hypothesis has also been found in the neuroimaging literature. Gauthier, Skudlarski, Gore, and Anderson (2000) used FMRI to demonstrate that bird and car experts when carrying out tasks with faces, familiar objects, cars and birds showed a very strong correlation between a behavioural test of expertise for birds and cars, and activation in the right FFA. This research suggested that familiar objects (familiar through years of study and interaction) were using structures of the brain that had been specifically associated with face recognition (e.g. Kanwisher, 1997). Not only this, but Gauthier, Tarr, Anderson, Skudlarski and Gore (1999) also demonstrated that when participants were required to learn a novel category of objects 'greebles' for a recognition task, as expertise in greebles increased so did activation of the FFA. Gauthier et al., (1999) demonstrated that the FFA was also activated through recognising unfamiliar or learned objects. Gauthier's work is not without its critics. For example, other researchers have suggested that the activation of the FFA by the greebles is due to their similarity to faces rather than the viewer's expertise (Brants, Wagemans & de Beeck (2011). However, there may be evidence in the clinical and neurological fields that supports a more general view of face recognition.

Autism, and Williams syndrome (WS) are disorders with distributed cognitive deficits and benefits that have also been shown to affect face recognition and face matching. Individuals with WS, a rare genetic disorder, have been described as having a complex profile of abilities and impairments (Donnai & Karmiloff-Smith, 2000). Despite having a significantly reduced IQ, averaging around 56 (Bellugi, Klima & Wang, 1996), individuals with WS are hyper social and have relatively spared language processing skills but poor visual spatial processing skills (Martens, Wilson & Reutens, 2008; Bellugi, Lichtenberger, Jones, Lai, & George, 2000; Wang & Bellugi, 1994). Perhaps most importantly, individuals with WS have been shown to perform as well as controls without disorders on the Benton Facial Recognition Task (BFRT: Benton, Hamsher, Varney & Spreen, 1983; Bellugi, Lichtenberger, Mills, Galaburda & Korenberg, 1999). This has been claimed by many as an example of the specific nature of face recognition that is spared despite other impairments. However, these findings should be taken with caution; Duchaine and Weidenfeld (2003) have shown that the BFRT can be completed when all the face information is removed. More recent research has shown that adults with WS in a face matching task when matched with chronological age controls performed comparably when the matching task required featural analysis but poorly when the task required configural analysis (Karmiloff-Smith, 1997). In a face matching task for typically developing children and children with WS, autism and Down's syndrome (DS) researchers found that the children with WS and high functioning children with autism were impaired in identifying configural changes to stimuli but comparable with typically developing children in identifying featural changes. The low function children with autism and the children with DS were poorer at identifying both types of changes. Interestingly only the children with WS gained comparable scores as the typically developing children on the BFRT. The children with autism, and downs syndrome also had lower scores on the BFRT than the children with WS and the typically developing children (Dimitriou, Leonard, Karmiloff- Smith, Johnson & Thomas, 2015). These results could support a number of theories. One explanation could suggest the cognitive deficits associated with WS support the dissociation of the two types of face processing. Face recognition appears to be more specific than face perception, which is also associated with visual spatial processing. The results could also be unreliable since the BFRT has been shown to be unreliable. However, it is unclear why the performance on the BFRT would be reduced for individuals with autism if this were the case.

Autism is traditionally defined as a triad of impairments social deficits, communicative impairments and restricted and repetitive behaviours and interests with other traits such as spatial cognition being relatively spared (World Health Organization, 1992; Edgin & Pennington, 2005). In addition to the social difficulties imposed by their lack of understanding of social cues and rules, individuals with autism also experience deficits in face recognition and face perception making social interaction increasingly difficult (for a review see Weigelt, Koldewyn & Kanwisher, 2012). In their review, Weigelt et al., (2012) found that individuals with autism were not processing faces in any qualitatively different way, but were poorer at the task, particularly in tasks requiring memory. This suggested that their deficit was face specific. However, in a follow up study Weigelt, Kolewyn and Kanwisher (2013) demonstrated again that individuals with autism had deficits in face recognition relative to controls but not face perception. Individuals with autism were no less accurate than controls for recognising cars or places. This again suggests a domain specific deficit and a functional dissociation between face memory and face perception. However, they also found that individuals with autism were also impaired in perceiving and recognising bodies, suggesting that the domain specifically affected was social rather than facial.

These findings are supported by earlier work by Wolf, et al., (2008) who found that children with autism spectrum disorder (ASD) were impaired in their ability to recognise faces. The children in the study were also impaired in their ability to discriminate featural and configural information in the eye regions during a face discrimination task. Their ability to discriminate featural and configural information in the mouth region however was preserved as was their ability to recognise cars, and their discrimination of featural and configural changes in houses was superior to typically developing controls. Taken together these findings suggest that individuals with autism do not have general deficits in memory or configural or featural processing, or a more specific deficit for faces but for the most socially salient stimuli – in this case, the eyes. This deficit that has led to the development of the ‘eye avoidance hypothesis’ to explain the low levels of face recognition experienced by individuals with autism (Tanaka & Sung, 2016).

The deficits for individuals with WS and autism suggest a functional dissociation between face memory and face perception. Individuals with autism are impaired at face recognition and potentially only disadvantaged in face perception due to the social rather than facial aspects of the task, since they have strong visual spatial skills and no problems with identifying configural changes in either the mouth regions of faces or houses. In contrast, individuals with WS appear not to be impaired in face recognition but are impaired in face matching potentially due to their poor visual spatial skills. These deficits also suggest that face processing may not be as specific as the face specificity hypothesis might suggest. However, they do not necessarily support the expertise hypothesis. Individuals with autism appear to be highly skilled at processing houses and places but is this an effect of experience and therefore expertise? It is difficult to establish, as individuals with autism are not a blank slate and have years of experience in their daily life with houses and places that is very difficult to quantify. The same could be said of typically developing individuals and faces, but it might be reasonable to expect that individuals employed in roles where they are required to recognise and match faces day to day have greater experience of those tasks than the average individual. This being the case are they more expert and therefore better at matching or recognising faces?

1.8 Expertise in the applied context

As discussed earlier, research has indicated that time spent matching faces does not necessarily lead to an improvement in face matching ability. White et al., (2014a) found that passport officers who match faces day to day professionally were no better at a standardised task of unfamiliar face matching (the GFMT, Burton et al., 2010) than the general population, regardless of their time in the role. These were not isolated findings in the applied context. Heyer (2013) tested passport examiners on a one to many unfamiliar face matching task using photograph arrays and found that the examiners were not significantly more accurate than the novice participants (82.6% correct against 81.3% correct). Heyer also did not find any effect of training or duration of experience i.e. time on the job.

These findings are not limited to passport officers, face matching using photographs or even face matching tasks alone. Kemp et al., (1997) found that the supermarket staff accepted fraudulent Photo-ID on over 50% of trials during a photo and live individuals face matching task, despite being aware that they were part of a study. Police officers with experience in forensic identification performed as poorly as naïve participants in an unfamiliar face recognition task using photos and CCTV footage (Burton et al.,1999a).

We might expect supermarket cashiers to perform less accurately on a face matching task than passport or police officers since in 1997 most of their day-to-day work involved validating identity via signatures. Police officers and passport officers would be more likely to carry out verification using photo ID but even so are no more accurate in unfamiliar face recognition tasks than naïve viewers are. It could be assumed that police and passport officers would have received training in face matching or face recognition and that studies with these individuals also show that training does not improve ability. However, this might not always be the case. Individuals with extensive training in facial identification have also been tested and found to be no more accurate in unfamiliar face matching than untrained viewers. Lee, Wilkinson, Memon and Houston, (2009) compared the performance of naïve viewers and trained individuals who had completed an MSc in Human Identification on an unfamiliar face matching task using photographs and CCTV footage. Even with unlimited time and the opportunity to manipulate and replay the film as they wished, the trained participants fared no better than untrained participants did.

Wilkinson and Evans (2009) found that facial imagery experts, who had years of experience in providing facial imagery analysis in court were more accurate than naïve participants when matching CCTV footage with a watch list array of identities. However, since the facial imagery experts were also the authors of study, it is difficult to be sure that the study was as blind or free of confounds as the authors might suggest. White, Phillips, Hahn, Hill and O'Toole (2015) ran a much more controlled study in which forensic facial identification examiners attending a Facial Identification Scientific Working Group (FISWG) policy meeting were invited to take part in a series of unfamiliar face matching tasks. Policy makers who did not complete facial matching as part of their day-to-day role were also invited to take part as an age and education matched control group, and finally students completed the tasks as a further naïve group. The facial identification examiners were significantly better at the face matching tasks than the students were across all three tasks; however, they were only better than the policy makers on the GFMT. White et al., (2015) identify that motivation could account to some degree for the facial identification examiners' success. The policy makers were significantly better than the students on the GFMT and only significantly worse than the facial identification examiners on the GFMT. Given that they had no training in face recognition or matching and were not assigned policy roles as a result of superior face matching ability it could be that they achieved their success through motivation. This motivation could also have been shared by the facial identification examiners.

Motivation aside, the examiners were superior in their unfamiliar face matching ability on the most widely used task in the study, suggesting that there may be an advantage in training and experience. White et al., also identified that the examiners performed with greatest accuracy when they were given unlimited time for response suggesting that they took an analytical approach. In addition, the examiners also had a reduced inversion effect, which could suggest that this approach might have included detailed featural analysis that would have been affected to a lesser degree by inversion than configural processing. These are interesting findings and suggest that some, but not all, types of experience may help improve unfamiliar face matching. They also provide some evidence regarding the ways in which experts might behave differently when compared to naïve viewers. While further investigation is required to ensure the validity of these isolated findings, they also provide some support for the experience hypothesis. These findings suggest that training and experience in unfamiliar face matching can result in changes in behaviour that improve face matching performance.

1.9 Super recognisers

While there is some support that experience can improve unfamiliar face matching, it is limited, only being found in one study to date. A more consistent finding is the presence of large individual differences in face matching ability. In their original validation of the GFMT, Burton et al., (2010) found a wide range of performance on both the long form (range 62% – 100% with a mean of 89.9%, SD = 7.3) and the short form (range 51% - 100%, with a mean of 81.3%, SD = 9.7). A wide range of performance was also found when passport officers were invited to complete the short form of the GFMT (range 58% - 95%, with a mean of 79.2%, SD = 10.4; White et al., 2014a). This range in ability is not limited to the GFMT, as passport officers in the same study also had a wide range of performance in a photo to photo matching task where the images were taken 2 years apart (range 59.5% - 92.9%, with a mean of 80.3%, SD = 8.3). The wide range in performance was also found in the photo to live person matching task (range 70.1% - 100%, with a mean of 89.9%, SD = 8.7). It is also important to note that as shown in figure 1.8 performance in unfamiliar face matching was dissociated from time on the job.

These ranges in performance suggest there are individuals who excel at unfamiliar face matching regardless of experience. Russell et al., (2009) were the first to investigate face recognition in high performers by testing members of the public who contacted their lab to report their superior face recognition abilities. Russell et al., (2009) identified that these individuals performed significantly better than controls across familiar and unfamiliar face recognition tasks and face perception tasks and termed them 'super-recognisers'. This seminal study was the first to identify that individuals could excel across a range of face recognition and face perception tasks. It was only later that individuals with super recognition skills were tested on face matching tasks. Using the lowest score achieved by the super recognisers in the study by Russell et al.,(2006) as a cut off for identifying individuals as super recognisers, Bobak, Dowsett and Bate (2016a) found that super recognisers also had superior unfamiliar face matching skills. These findings were also replicated in more applied versions of face matching in arrays and recognising faces from video footage (Bobak, Hancock & Bate, 2016b).

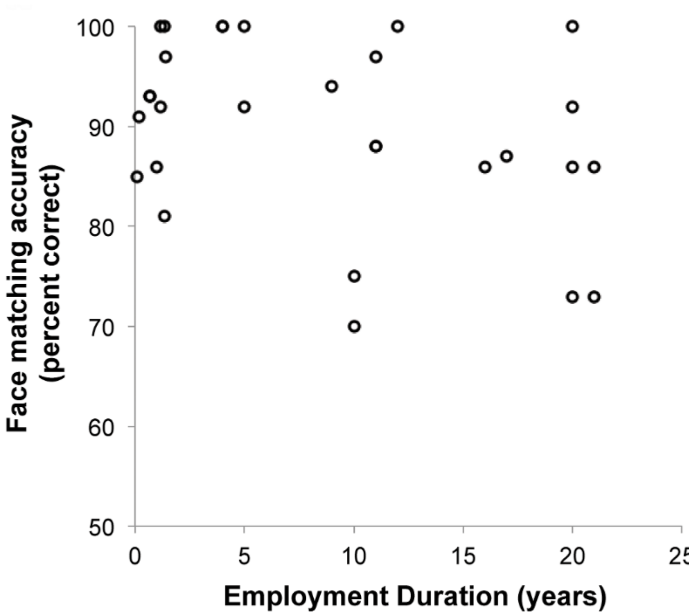

Figure 1.8. Performance on Person-to-Photo test as a function of Employment Duration (White et al., 2014).

As shown in the range of performance in the passport officers there are also individuals within what might be thought of as expert populations who have superior face processing skills, regardless of experience. A number of individuals within the Metropolitan Police identifiers team were identified as having superior face recognition skills by Davis, Lander, Evans and Jansari (in press) based on their performance across a range of in house tests, on the job suspect identification rates, and using the criteria adopted by Bobak et al (2016a). Davis et al (in press) confirmed the status of these individuals as having superior face recognition skills and face matching skills when they were found to perform at a significantly higher level than other members of the police and controls across familiar and unfamiliar face recognition tasks and unfamiliar face matching tasks. Davis et al (in press) confirmed the findings of an earlier study that had also found that a group of police identifiers who had been identified as having superior recognition skills were found to have superior performance across both familiar and unfamiliar face matching tasks (Robertson, Noyes, Dowsett, Jenkins & Burton, 2016). Taken together these studies suggest that high performers in one particular face task have superior performance across a range of face tasks regardless of whether those faces are familiar or unfamiliar or the task requires faces to be recognised or matched. Davis et al (in press) also suggest that this ability in lab based tests also translates to the applied context since the super recognisers identified within the Metropolitan police also had superior performance on the job with higher than average suspect identification rates. These findings would appear to support a face-specificity hypothesis.

However, Davis et al (in press) also provide a note of caution, not all of the super recognisers performed at a superior level across all of the tasks. One super recogniser performed very poorly in the GFMT, so much so that they were asked to repeat the task in case of confounds that were specific to the testing day, but their performance was again very poor. In the most recent study with super recognisers, Bobak, Bennetts, Parris, Jansari and Bate (2016c) again performing a range of face recognition, face matching and face perception tasks, showed that super recognisers had heterogeneous face processing skills. Although the scores for the super recognisers in general were higher than the controls for unfamiliar face matching, only two of the six super recognisers significantly out-performed the controls on the task.

These later studies suggest that performance across face processing tasks is related but as Davis et al (in press) note, super recognisers are not super human. In general, they perform very well across face processing tasks, but their performance may not always be superior. There are isolated findings, which do warrant further investigation. However, in general face processing skills appear to be associated, specific and unaffected by experience. The super recogniser literature also notes that numbers of super recognisers studied to date are low, perhaps too low for generalisations to be drawn and applied at this stage across the population. Further research is required to identify if face processing skills are, as the super recogniser literature suggests, both related and specific in the general population.

1.10 Improving Unfamiliar Face Matching

Davis et al (in press) predict that 95% of the population will perform below the level of super recognisers in the GFMT. Given that passport officers and border control officers may not have been recruited because of their face matching ability, it is reasonable to assume that 95% of the individuals working in those roles are not super recognisers. Recruiting super recognisers to these roles would provide greater security in the long term; however, improving performance of those currently in the role would provide even greater benefit in the short term. Since unfamiliar face matching was identified as a difficult task in the 1980s, a number of studies have investigated ways in which performance can be improved. Two main approaches have been taken; those which have made changes to aspects of the task and those which have sought to make changes to the individuals carrying out the task.

Changing the task

Unfamiliar face matching in the applied context occurs across a wide range of viewing conditions e.g. CCTV footage, live line-ups, IDs and social media. This high level of variability when images of unknown individuals are being compared impairs performance (Hancock et al., 2000; Bruce et al., 1999). The standardisation of passport photos with regard to neutral pose, forward facing, and advice regarding size of the face, distance from the camera and the lighting conditions sought to reduce variability in images (International Civil Aviation Organisation, 2015). Though this type of image was originally introduced to allow e-passports to store digital images and biometric measurements of the face that could be read by facial recognition systems (House of Commons, Committee of Public Accounts, 2007) it has also improved viewing conditions for human operators comparing ID images.

Standardising images within the ID context aims to improve face matching by reducing the variability between two images in a highly image bound task. A different approach would be to reduce the image bound nature of the task by providing multiple images of an individual or video footage. The additional images and video might allow multiple pictorial codes of an individual to be combined into a structural code creating greater familiarity and accuracy. However, research using moving images e.g. CCTV footage has not improved performance over and above static images for either face recognition or face matching (Burton, et al., 1999a; Bruce et al., 1999; Bruce et al., 2001). This suggests that unfamiliar faces may well only be stored as pictorial codes that cannot be improved. In contrast, research has also shown that moving images of the face only, such as rotating, moving side to side or in social motion e.g. speaking, have improved unfamiliar recognition performance over static images (Pike, Kemp, Towell & Phillips, 1997; Lander & Bruce, 2003). These findings suggest that a focus on the face rather than the individual and that more controlled variation e.g. controlled viewpoint and specified movement may allow the beginnings of a structural code to be created. They also suggest that unfamiliar face matching might also benefit from high quality moving images of the face that could be provided in the form of a 3d image or video embedded in the passport. However, given that unfamiliar face matching and unfamiliar face recognition may not be associated, this cannot be assumed and requires further investigation.

A number of studies have investigated improving unfamiliar face matching specifically. White, Burton, Jenkins & Kemp (2014b) found that matching a photo to an array of four images identified as showing the same individual improved performance for unfamiliar face matching when compared to performance matching a photo to a photo. White et al., (2014b) also found that participants were better at matching unfamiliar faces when an average image was used instead of a photo. In both cases, the aim was to encapsulate more of the within person variability for each individual than a single photo could do. Interestingly, White et al., also found that multiple photos also created greater accuracy than using an average image, suggesting that multiple images may enable the creation of a richer structural code than an averaged image. Research into unfamiliar face recognition has also shown that learning faces with multiple images improves unfamiliar face matching. Participants who learned new identities through multiple images prior to completing a matching task were significantly more accurate when matching photos for those identities (Ritchie & Burton, 2016). Importantly the gain in performance is greatest when the variability between the photographs used for training is also greatest e.g. images taken months apart as opposed to same day. This suggests that if photo ID was redesigned to include multiple photographs of an individual as White et al., (2014b) recommend the greatest levels of accuracy would be achieved if those photos were highly variable.

It is also important to note that multiple images of the same individual can be useful in face matching, but multiple faces can be detrimental. Bindemann, Sandford, Gillatt, Avetisyan & Megreya, (2012) demonstrated in an unfamiliar face matching task that if participants were required to match 2 target faces with faces from an array rather than a single face with an array their performance was significantly compromised. Even when the participants were told that only one of the target faces was in the array, their performance remained compromised (though less so). Heyer (2013) found that decision aids provided with facial recognition systems that were designed to improve matching performance for human operators did not necessarily achieve their aim. When a new image of an individual was entered into the system, a candidate list of best same faces would be provided for the human operators to use for matching. However, there was no guidance on the optimum size of candidate list so individuals across agencies were working with different candidate list sizes ranging from 7-10 candidates to over 250. Heyer (2013) found that as the candidate list size increased, performance decreased.

These studies demonstrate that multiple images are only useful when they are clearly identified as showing the individual to be matched and allow variation to be combined. Multiple identities surrounding a token image of an individual reduce the attention available for the individual and potentially interfere with the encoding of that individual's face. Multiple comparison identities in an array or candidate list create greater opportunities for error and misidentification. However, adding multiple photos of the named individual to ID documents may well increase unfamiliar face matching performance, as could an additional 3d image, face average or small embedded video of the individual's face in motion. Making any of these changes would require changes to the design and build of ID documents/systems and could be costly and technically difficult. Ensuring the individuals who carry out the unfamiliar face matching task are performing optimally may well be easier to implement.

Changing the Practice

There are ways that face matching for particular individuals can be improved. Familiarity can be rapidly increased through a sorting task. Viewers are given multiple images of two identities and told to sort them into a pile for each identity, prior to completing the face matching task. If viewers are not told how many individuals are in the mixed set then they will be highly inaccurate in their sort, once told the number they become very accurate. After completing the sort viewers become more accurate in face matching tasks for that identity (Andrews, Jenkins, Cursiter & Burton, 2015). Familiarity can also be improved through a face learning task as shown in (Ritchie & Burton, 2016). Dowsett, Sandford and Burton (2016) demonstrated that face matching accuracy could shift from as little as 50% to 90% overall as participants were given extra images of one of the individuals in an unfamiliar face matching task. However, as Dowsett et al., (2016) also showed these improvements were specific to the identity and did not generalise to any improvements for performance on other identities.

Improving unfamiliar face matching generally (rather than for specific identities) appears to be a much more difficult task. As discussed, individuals working as facial identification examiners, who may have received training, have been shown to have superior unfamiliar face matching performance when compared with controls (White et al., 2015). Training programs are provided in unfamiliar face matching, many of which are based on the Facial Identification Scientific Working Group (FISWG) guidelines (FISWG, 2012). However, their content is proprietary and these training programs are often unvalidated, which makes their effectiveness unknown. An evaluation of one method of training used in facial image comparison did not find an advantage for training (Towler, White & Kemp, 2014). Face shape comparison is a common component of face matching training. This method requires viewers to classify the faces viewed as being one of seven face shapes and to use this classification as a tool when comparing faces with each other. Towler et al., (2014) found neither agreement in face shape classification nor any improvement in participants' face matching ability post training. This suggests that traditional components of training programs may not be as successful as supposed.

However, less structured and potentially more informal methods of training have been shown to improve performance in unfamiliar face matching tasks. Trial by trial feedback has been shown to improve unfamiliar face matching ability. White, Kemp, Jenkins and Burton (2014c) showed that immediate feedback provided to participants after each trial on an unfamiliar face matching task improved performance on those identities and was generalised to unfamiliar face matching with new identities. The feedback provided only told participants whether they were correct or incorrect and that the faces shown were either the same individual or different individuals. This suggests that viewers may be able to adjust their matching strategies without specific guidance as to what needs to be changed e.g. 'You were incorrect, look again at the features of the faces, the noses are clearly different'. However, this kind of information may also prove useful. Dowsett and Burton (2015) have demonstrated that working in pairs may improve unfamiliar face matching performance in naïve viewers. They found that poor performers improved in accuracy when matching with high performers; improvement that was retained even when the poor performers continued face matching later alone. It is not known to what extent the participants discussed their reasons for their decisions, but the high performers may well have not only disagreed with the poorer performers but also shared the reasons for their disagreement e.g. ‘The noses are clearly different’. Whatever the participants working in pairs shared, the level of improvement in performance was around 10%, the same level of improvement as found in the feedback study conducted by White et al (2014c). This suggests that there are ways of improving performance, but that these methods are limited in the amount of improvement that can be realised.

The increased performance through working in pairs and receiving feedback may also be due in part to motivation. In their study with facial identification examiners, White et al (2015) noted that motivation could have played a role in their superior performance since the control policy group also performed as well as the experts on two of the three face matching tasks. This policy group were not selected because of their face matching skills but may well have been more motivated than the student controls through aspects such as professional pride or competition with colleagues. Motivation has been tested as a route to improving unfamiliar face matching performance. Moore and Johnston (2013) invited participants to participate in an unfamiliar face matching task where if they performed above average they would receive their choice of a food incentive. Chocolates and sweets were left in full view while the participants completed the task and those in the incentivised condition were 10% more accurate than those participants who were in the non-incentivised condition. Interestingly, when the prevalence rate of different faces was dropped from the usual equal split between same faces and different faces seen to just 2 different face pairs and 30 same face pairs the motivated participants became almost 30% more accurate for mismatch trials and just 2% more accurate for match trials.

When the prevalence of same and different face pairs is at 50/50 motivation is an improving factor but like feedback and working in pairs, limited. However, the change in prevalence showed that motivation could have an even greater impact of performance particularly when in the non-incentivised condition, performance dropped from 82% to 57% in the mismatch condition. Keeping employees incentivised in the applied context is difficult and the implications of failing to do so are stark. However, the reality is that the lower prevalence of fraudulent documents is the most likely scenario in the applied context.

These potentially low levels of performance day to day only highlight the need to recruit and retain the best performers or ‘super recognisers’ and use feedback and working in pairs in training to provide any limited increases in performance for those already in the roles. This review of the unfamiliar face recognition and face matching literature provides three main themes that can guide further research into unfamiliar face matching in the applied context. First, the literature reviewed provides greater support for the face specificity hypothesis than the expertise hypothesis, suggesting that unfamiliar face matching is a face specific process. Secondly, the literature reviewed also provides some support for a dissociation between unfamiliar face *recognition* and unfamiliar face *matching* which suggests solutions for improving one may not necessarily translate to the other. Finally, caution must be taken when using any kind of test e.g. the BFRT (Benton et al., 1983) to ensure it is tapping the underlying mechanisms required.

1.11 Aims & Overview

The aim of this thesis is to investigate unfamiliar face matching in the applied context. In light of the themes emerging from the review of the literature, three objectives have been identified:

1. Identify the nature of the unfamiliar face matching task carried out in an ID context;

2. Identify measures and tasks that predict success in unfamiliar face matching; and

3. Test these predictions through manipulations of the stimuli.

Together, the experiments in this thesis suggest that unfamiliar face matching is different in the applied context when compared to the lab. Face matching in the lab usually consists of faces being matched in isolation, however in the applied context, one of the faces would usually appear in some sort of ID with associated biographical data. I show in chapters 2 and 3 that, in this context, faces are more likely to be identified as showing the same individual. Not only this, but the face pair interferes with participant's ability to check biographical data. Invalid data is frequently missed and even more so if the face pair shows the same individual. This confirms that face matching in the applied context is not only different from tests carried out within the lab but also more difficult. In chapter 4, a large battery of face processing and more general tasks completed by members of the general population demonstrate that success in unfamiliar face matching can be predicted to a large degree by success in other face processing tasks. This finding lends greater support to the face-specificity hypothesis and does not support a dissociation between face recognition and face matching.

However, unfamiliar face matching is also unusual in that performance on same and different face pairs was dissociated. Part of that dissociation may be explained by the way same and different face matching associate with unfamiliar face recognition and more general tasks of local processing and spatial relations. This was explored through a number of unfamiliar face matching tasks in which featural and configural processing was isolated or disrupted.

Within the applied context, these findings suggest that individuals with superior unfamiliar face matching skills could be recruited from a pool of super recognisers or via an unfamiliar face recognition task. However, in order to ensure optimum performers in unfamiliar face matching roles, applicants should be screened on their performance in same and different face pair matching as well as their performance in unfamiliar face matching overall.

Chapter 2 – Unfamiliar face matching in passport frames

2.1 Introduction

It is now well established that unfamiliar face matching is a difficult task, on which viewers are prone to errors (e.g., Bruce et al., 1999; Johnston & Edmonds, 2009; Megreya & Burton, 2006). This difficulty is typically demonstrated by lab-based experiments in which viewers are asked to make same/different decisions to two photographs, however it is also observed when matching a live person to a photo (e.g. Davis & Valentine, 2009; Kemp et al., 1997; Megreya & Burton, 2008). Furthermore, expert observers, including passport officers, are typically no more accurate than the general population (White et al., 2014a).

While these findings appear to be important for informing practical uses of face recognition, the experiments typically do not capture an important aspect of day-to-day experience with checking photo ID. In reality, viewers rarely see faces in isolation (Heyer & Semmler, 2013). Instead, photos are usually embedded in documents such as passports, driving licenses, or other photo-ID. These documents usually contain important information about the bearer (e.g. name, age, address), and people checking ID are often required to confirm some of these details - for example age-checks for alcohol sales. This raises two questions: first, does the presence of biographical data within a document context make any difference to people's face matching ability; second, does the presence of faces affect a viewer's ability to check the biographical data?

Although most laboratory-based face recognition work has used images of isolated faces, some previous experiments have presented these in the context of a real document. Kemp et al., (1997) asked volunteer shoppers to show a mocked-up payment card to cashiers, while Bindemann and Sandford (2011) showed participants photos of university identity cards. However, neither of these experiments required participants to check the data shown on these cards, and neither provide a direct comparison between seeing faces within a document context and seeing them alone. In the following experiments, I therefore set out to test whether these factors influence one's ability to match faces. The intention is to provide an important missing link between laboratory-based face matching experiments, and the demands of day-to-day operation by professionals checking ID.

There are some good reasons to hypothesise that there may be an interaction between face matching and checking document data. First, any observer required to make *both* a judgement on a face, and a judgement about the likely veracity of document information is under greater cognitive load than someone simply required to make a judgement about faces. Of course, task flow could be designed to attempt to optimise viewers' performance, but in a typical ID card (e.g. passport), information about the carrier's face and personal information is simultaneously present and so it is plausible that the simple presence of multiple information sources is distracting.

Second, there is the possibility of interference between the two types of information. Within the fingerprint matching context, individuals without expertise have been shown to make false match decisions based on the gender or ethnicity of the 'accused' (Smalarz, Madon, Yang, Guyll, & Buck, 2016). Stereotypical beliefs about crimes and the types of people who commit them leak through into the fingerprint matching judgement despite having no logical bearing on the decision. Research carried out with fingerprint experts has shown that they are also vulnerable to contextual biases such as prior match or mismatch decisions (Dror, Charlton & Péron, 2006). Termed 'forensic confirmation bias' Kassin, Dror and Kukucka (2013) provide a thorough review of the various sources of this bias in image comparison or memory, across a range of forensic contexts which can affect lay viewers and experts alike. Given the potential sources of bias, might a false piece of information on a document (for example one which indicates the wrong sex or age), lead to a bias to reject two face images as being the same person?

These issues have their source in theoretical understanding of human information processing, which consistently show that tasks with multiple demands are harder than single-demand processes (for a forensic example, see Menneer, Stroud, Cave, Li, Godwin, Liversedge & Donnelly 2012). Furthermore, it is well established that the presence of faces can interfere with other perceptual tasks, even when the faces are task irrelevant (e.g. Jenkins, Lavie & Driver, 2003; Lavie, Ro & Russell, 2003).

Thirdly, there is also the possibility that the addition of the passport frame alone may improve accuracy in the face matching task. Galli, Feurra and Viggiano (2006) have shown that unfamiliar faces embedded in newspaper articles are recognised more frequently than unfamiliar faces shown in isolation. The article headlines specified an action committed by the person depicted. Half of the headlines were emotionally positive (positive-context faces: "Taxi-driver saves a child") and half were emotionally negative (negative-context faces: "Hooligan rapes a girl"). Not only were the faces shown in a newspaper context better remembered, the study also found that negative-context faces were also remembered more accurately than the positive context faces. Viewers remembered faces better when supported by the newspaper context, a perceptual context, and the emotional, or semantic, context. The display of face pairs proposed here will have both a perceptual context (the passport frame) and a semantic context (the validity of the data). Galli et al., (2006) suggest that both of these factors may influence accuracy achieved in the face matching task. The forensic confirmation bias supports this hypothesis since forensic confirmation bias is largely based in the semantic context of the evidence - innocent or guilty.

In addition to its theoretical interest, this is also a practical problem across a variety of settings. For example, border control is not the only situation in which face matching within a document takes place. In many countries, passport renewal authorities attempt to guard against fraud in multiple ways, including comparison of application photos to previous applications or other official documentation (e.g. driving licenses). The most recent available data for our own jurisdiction, the UK, show that in the financial year 2013-14, 5.7 million passport applications were processed, and of these 0.15% (over nine thousand) were detected to be fraudulent (Her Majesty's Passport Office, 2014). Therefore, it is important to establish whether the typical lab-based face matching study, in which faces are shown in isolation, generalises well to a situation in which faces are seen within documents.

This experimental chapter aims to investigate face matching in the ID context, presenting viewers with pairs of faces, sometimes embedding one of these images in a passport frame. Participants are asked to judge whether the two photos show the same or different people, and to decide whether the personal data is accurate. This allows the potential for interference and bias both for face matching and data checking to be investigated.

2.2 Experiment 1

Introduction

In this first experiment, I investigate matching a face to an ID document (a passport) containing biographical data. I am concerned with two issues. First, does embedding a face image in an ID document make it any easier or harder to match? Second, what are the effects of adding a biographical data check? To answer these questions, we used stimuli from a standardized test of face matching, the Glasgow Face Matching Test (GFMT, Long Version; Burton et al., 2010). This comprises 168 face image pairs, for each of which viewers decide 'same person' or 'different person'. Using these stimuli, we constructed items in which one of the face pair was embedded into a stylized UK Passport (see Figure 2.1). Biographical data in the passport frames was divided into two categories: valid and invalid data. Participants were then asked to make a face match (same/different person) and to check the data on the passport for accuracy. The main interest of this work is the detection of potentially fraudulent identity documents. I therefore aim to establish conditions in which it is easy or difficult to spot somebody using the 'wrong' ID – i.e. someone whose document does not contain an image of themselves.

Method

Participants

Participants were 45 students (10 male) from the University of Aberdeen, who all reported normal or corrected to normal vision (mean age = 22.06, range = 18-40). To ensure their data checking was not compromised by a lack of familiarity with place names participants were also required to be British Citizens who had lived in the UK for at least the last 10 years. Participants were given course credit, or reimbursed a small fee for their time.

Materials

The 168 face pairs from the GFMT long form were used as stimuli. In order to construct ID-document items, each of the test pairs was recreated in a version with the left-face embedded in a passport frame, and another with the right face embedded in a passport frame. Across all items, distance between the pair was kept constant. Biographical information (all fictitious) was designed to be valid or invalid with the associated face. A full set of the biographical information used can be found in Appendix 1. Key personal data could be rendered invalid as follows:

- wrong-gender forename (see bottom item in Figure 2.1) - 11 items;
- nouns as forename (e.g. 'Fork') – 9 items;
- unlikely male ethnicity surname (e.g. the Sri Lankan name 'Selvaratnam' for a male Caucasian face) – 6 items; and

- wrong birth date (either an impossible date, e.g. 30 Feb; or a birth year more than 20 years discrepant from the age of the target faces - the GFMT face database includes the age of each person when the photo was taken, and so this information is available) – 16 items.

Errors such as these are commonly found in forged documentation, due to inexpert transcription; and are routinely looked for by passport granting authorities. Invalid passport frames contained only one of these errors.

Condition	Presentation
Faces only	
Faces with valid data in a passport frame	
Faces with invalid data in a passport frame	
Conditions were counterbalanced and interleaved across 168 pairs of faces.	

Figure 2.1. Items from GFMT in normal presentation and with one face embedded in a passport frame.

Procedure

This experiment was conducted in Matlab on a MacMini using a 17" Dell screen. Each participant saw 42 face-only pairs and 26 pairs embedded in passport frames (half left, half right, see Figure 2.1), with order of presentation randomized throughout (i.e. not blocked by condition). Blocking was not use to reflect workplace practice where passport officers may have additional historical images for comparison but in the case of new applications may not. Half the items showed same-person pairs, and half different person pairs. For those pairs with passport frames, one third of the items showed invalid data (i.e. 44 items per participant). Face pairs were counterbalanced across the experiment such that each pair occurred equally often in each condition.

For face-only pairs, participants were asked, on-screen, *'Are the images of the same individual?*' and selected responses '*Same'* or *'Different'* with a mouse. For pairs including a passport frame, participants were asked *'Is the data correct?'* and selected either '*Yes'* or *'No'* and then *'Are the images of the same individual?*' and selected *'Same'* or *'Different'*. The order of the questions was counterbalanced and all stimuli remained on screen until responses had been made. The task was self-paced, and typically took about 45 minutes to complete.

Results

Face Matching Accuracy

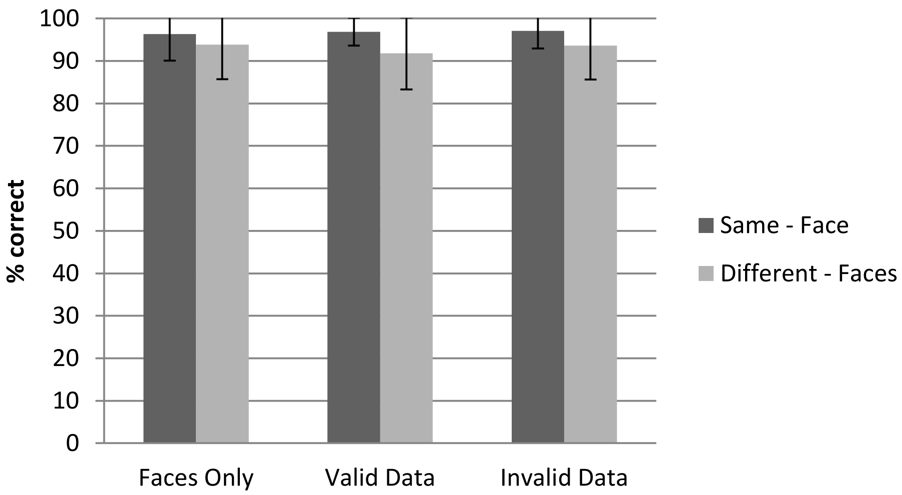

Figure 2.2. Mean accuracy for the face matching task for same and different identity trials when face pairs were shown alone and with valid and invalid data. Note error bars represent standard deviation.

A 2 (match type: same, different) x 3 (presentation type: faces only, face with valid data, face with invalid data) within subjects ANOVA revealed a significant main effect of match type $F(1,43) = 9.71$, $p < .005$, $\eta_\rho^2 = .18$ (see figure 2.2). There was no significant main effect of presentation type $F(2, 86) = 1.49$, $p > .05$, $\eta_\rho^2 = .03$ or interaction $F(2,86) = 2.44$, $p > .05$, $\eta_\rho^2 = .05$. There was no significant difference in face matching accuracy whether participants viewed face pairs alone or with valid or invalid data. Participants had a bias to respond 'same' to face pairs across all presentation types – a bias that will lead to increased errors in accepting fraudulent passports.

Data Checking Accuracy

A 2 (match type; same, different) x 2 (data type; valid, invalid) within subjects ANOVA revealed a significant main effect of match type $F(1,43) = 35.11$, $p < .001$, $\eta_\rho^2 = .45$ (see figure 2.3). There was also a significant main effect of data type $F(1,43) = 41.42$, $p < .001$, $\eta_\rho^2 = .49$ which was qualified by a significant interaction $F(1, 43) = 49.67$, $p < .001$, $\eta_\rho^2 = .54$.

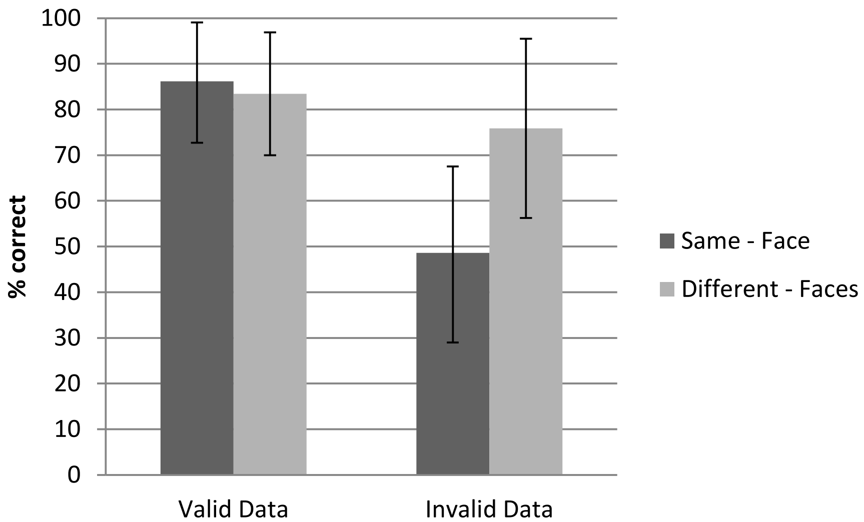

Figure 2.3. Mean accuracy for the data checking task for valid data and invalid data when shown with same and different face pairs. Note error bars represent standard deviation

Simple main effects analyses on match type reveal a significant difference between data accuracy achieved for data presented with same face pairs or different face pairs when shown with invalid data $F(1, 86) = 84.32$, $p < .001$, $\eta_\rho^2 = .50$ but not valid data $F(1, 86) = 0.83$, $p > .05$, $\eta_\rho^2 = .01$. In summary, participants were poor at spotting invalid information – detecting it on only 61% of occasions. There is also an influence of the (task-irrelevant) face match on performance with data checking. Participants are much more likely to miss invalid biographical data when the data are shown together with a same face pair. However, these invalid data could flag fraudulent passports in which the photo of the holder has been replaced with a photo of the person carrying the passport.

Discussion

This experiment shows a number of interesting effects. Across the two rather different tasks (face matching and data verification), participants were poorer at spotting 'fraudulent' passport use. For face matching, participants were very accurate at matching the faces, though performance showed viewers were poorer at detecting a mismatch regardless of whether the face was shown in a passport frame or what type of data the passport frame contained. This suggests that the presence of biographical data within a document context does not make any difference to people's face matching ability; the data itself neither enhances nor harms face matching accuracy. However, since face matching performance showed ceiling effects it is unclear whether this might be the case if the face matching task was more difficult.

In the data checking task participants demonstrated a strong tendency to miss invalid ('fraudulent') data in the passport frames. Participants knew that they would be seeing some invalid data, and yet they failed to spot this on almost 40% of the occasions it was present. This was unexpectedly high and suggests participants may have been distracted from the data checking task by the face pairs. This theory may be supported by the modifying effect of the faces on data-checking accuracy. When the faces matched, viewers were less likely to identify an error on an invalid passport. This pattern of data seems to suggest that same face pairs have an effect on data checking, even though the faces are task-irrelevant – detecting that the faces are the same seems to generalise (falsely) to biographical data when it is in fact invalid. This should not be the case since although the faces and biographical data remain on screen during both the face matching and data checking task, each task is presented independently. Overall, this demonstrates that the nature of the face pair does affect a viewer's ability to check the biographical data. It should be noted that the GFMT long consists of 168 pairs of faces and the passport checking task took 45 minutes to complete. It would be interesting to see if these effects could be replicated if the short version of the GFMT is used. This version contains only 40 face pairs and would be a faster test to administrate in an applied setting.

It would also be useful to see whether the effects generalise to on-line testing which would allow testing of the general public in a more interesting way. Running psychology experiments online has historically been associated with a number of concerns regarding the validity of the data collected. These concerns have included queries regarding whether participants truly meet inclusion criteria or behave in ways that might meet their own economic aims (e.g. speeding through answers without attention) rather than those of the study. However, as a number of experiments have demonstrated, the behaviour of online participants may not be vastly different to those participating in the lab i.e. there are poor participants in both environments and screening measures are used in both; and a number of strategies can be put in place to mediate these concerns (for review see Göritz, 2007; Lowry, D'Arcy, Hammer & Moody, 2016).

In the following experiment I use the short version of GFMT in an online version of the task to test if these effects can be replicated using a shorter number of trials in an online environment.

2.3 Experiment 2

Introduction

The previous experiment demonstrates that performance for face matching is unaffected by the context of a passport frame. Viewers have a bias to respond same when matching faces that occurs regardless of the presence of a passport frame or the type of biographical data contained. Contrary to my hypothesis, the presence of invalid biographical data did not influence the viewer's decision in the face matching task. Viewers were not more likely to identify a same face pair as a different face pair because the data suggested the passport was fraudulent. The opposite outcome may also have occurred. Viewers may have been more likely to identify a different face pair as a same face pair because the data matched the faces shown. This was also not the case.

In contrast, data checking was affected by the presence of faces. Accuracy for invalid data was poor overall; however, for the most part this was driven by accuracy for invalid data presented with same face pairs. At 48%, accuracy for invalid data presented with same face pairs was significantly lower than accuracy for invalid data when presented with different face pairs. The 'valid' nature of the face pairs was incorrectly carried over to the data. Whilst this was unexpected, this was not entirely contrary to my hypothesis. Forensic confirmation bias would suggest that accuracy for valid data presented with different face pairs would drop as the 'invalid' nature of the face pairs is incorrectly carried over to the data. Confirmation bias was seen in the earlier experiment; it was the nature or direction of the confirmation that was unexpected.

It would also be useful to see whether the effects generalise to on-line testing which would allow testing of the general public in a more interesting way. Running psychology experiments online has historically been associated with a number of concerns regarding the validity of the data collected. These concerns have included queries regarding whether participants truly meet inclusion criteria or behave in ways that might meet their own economic (e.g. speeding through answers without attention) rather than those of the study (Göritz, 2007; Lowry, D'Arcy, Hammer & Moody, 2016).

The current experiment seeks to replicate and extend these findings using a shorter and more difficult (in terms of face matching) version of the task completed in an online environment.

Method

Participants

Participants were 35 students (8 males) from the University of Aberdeen, who all reported normal or corrected to normal vision (mean age = 20.29, range = 18-35). All participants were British citizens and had lived in the UK for the previous 10 years. Participants were reimbursed a small fee for their time.

Materials

The 40 face pairs from the GFMT short form were used as stimuli and a subset of the passport frames from Experiment 1 were also used. The short form includes the most difficult trials found in the long form of the test which should reduce ceiling effects found in Experiment 1. In order to construct ID-document items, each of the test pairs was recreated in a version with the left-face embedded in a passport frame, and another with the right face embedded in a passport frame. Across all items, distance between the pair was kept constant. Biographical information (all fictitious) was designed to be valid or invalid with the associated face. A full set of the biographical information used can be found in Appendix 2. Key personal data could be rendered invalid as follows:

- wrong-gender forename (see bottom item in Figure 2.1) – 2 items;
- nouns as forename (e.g. 'Fork') – 2 items;

- unlikely male ethnicity surname (e.g. the Sri Lankan name 'Selvaratnam' for a male Caucasian face) – 1 item;
- and wrong birth date (either an impossible date, e.g. 30 Feb; or a birth year more than 20 years discrepant from the age of the target faces) – 5 items.

Invalid passport frames contained only one of these errors.

Procedure

This experiment was conducted on-line using the Qualtrics survey system. Each participant saw 10 face-only pairs and 30 pairs embedded in passport frames (half left, half right), with order of presentation randomized throughout (i.e. not blocked by condition). Half the items showed same-person pairs, and half different person pairs. For those pairs with passport frames, one third of the items showed invalid data (i.e. ten items per participant). Face pairs were counterbalanced across the experiment such that each pair occurred equally often in each condition. For face-only pairs, participants were asked, on-screen, '*Are the images of the same individual?*' and selected responses *'Same'* or *'Different'* with a mouse. For pairs including a passport frame, participants were asked two questions: '*Is the data correct?*' and '*Are the images of the same individual*?', and made their responses with a mouse. Order of questions was counter-balanced, and all stimuli remained on the screen until responses had been made – i.e. until after both responses for the pairs including a passport frame.

Prior to the experiment participants were given two practice trials, a face-only pair, and a passport frame pair. They were asked to practise same/different face responses by selecting the appropriate button on screen. For the passport item, they were asked to judge whether the personal information was correct with respect to the person shown in the passport. Instructions for this decision were as follows: '*If the data matches the image on the passport, e.g. correct gender first name and year of birth click 'Yes'. If the data is factually incorrect or appears to contradict the image, click 'No'*. No feedback was provided and participants proceeded directly to the experimental phase of the study. The order of questions was counterbalanced across participants. The task was self-paced, and typically took about 20 minutes to complete.

Results

Face Matching Accuracy

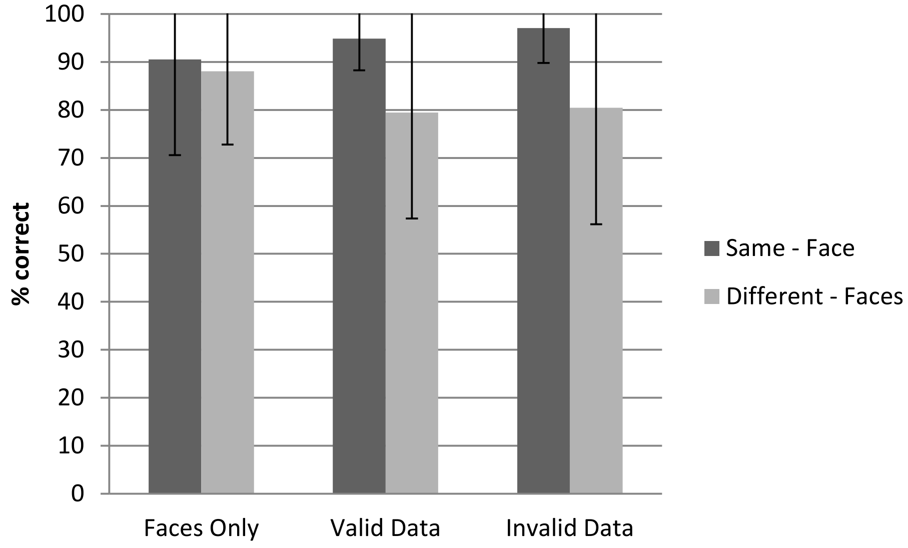

Figure 2.4. Mean accuracy for the face matching task for same and different identity trials when face pairs were shown alone and with valid and invalid data. Note error bars represent standard deviation.

A 2 (match type: same, different) x 3 (presentation type: faces only, face with valid data, face with invalid data) within subjects ANOVA revealed a significant main effect of match type $F(1,34) = 10.80$, $p <.005$, $\eta_\rho^2 = .24$ but no significant main effect of presentation type $F(2, 68) = 0.43$, $p > .05$, $\eta_\rho^2 = .01$ (see figure 2.4). This was qualified by a significant interaction $F(2,68) = 8.34$, $p <.001$, $\eta_\rho^2 = .20$.

Simple main effects analyses on match type reveal a significant difference between face matching accuracy for different face pairs across the presentation types $F(2, 136) = 4.66$, $p <.01$, $\eta_\rho^2 = .06$ but not for same face pairs $F(2, 136) = 2.34$, $p >.05$, $\eta_\rho^2 = .03$. Simple main effects analyses on presentation type reveal that there was a significant difference between accuracy for same and different face matching for face pairs shown with valid data $F(1, 102) = 13.86$, $p <.001$, $\eta_\rho^2 = .12$ and invalid data $F(1, 102) = 16.17$, $p <.001$, $\eta_\rho^2 = .14$ but not for face pairs shown alone $F(1, 102) = 0.36$, $p >.05$, $\eta_\rho^2 = .00$. Overall, face matching accuracy is not affected but the presence of a passport frame leads to significantly poorer performance in the face matching task for ‘different’ items. This pattern of data suggests that the presence of a passport frame introduces a bias to respond ‘same’ to face pairs – a bias that will lead to increased errors in accepting fraudulent passports.

Data Checking Accuracy

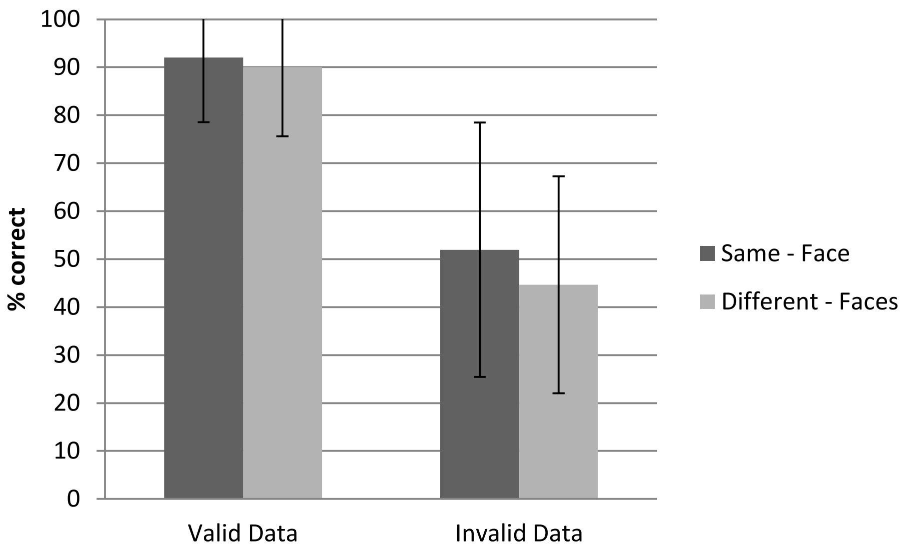

Figure 2.5. Mean accuracy for the data checking task for valid and invalid data when shown with same and different face pairs. Note error bars represent standard deviation.

A 2 (match type; same, different) x 2 (data type; valid, invalid) within subjects ANOVA revealed a significant main effect of data type $F(1,34) = 92.87$, $p < .001$, $\eta_\rho^2 = .70$ (see figure 2.5). There was no significant main effect of match type $F(1,34) = 2.49$, $p > .05$, $\eta_\rho^2 = .07$ or significant interaction $F(1,24) = 1.08$, $p > .05$, $\eta_\rho^2 = .03$. Participants were very poor at spotting invalid information regardless of whether that information is shown with a same face pair or a different face pair.

Discussion

The purpose of the current experiment was twofold; to identify if the effects found in experiment 1 could be replicated using a subset of the face pairs and passport frames and whether these effects generalised to online testing.

This shorter form of the passport checking task showed a number of interesting similarities and differences when compared to the longer form. Participants were again poor at spotting 'fraudulent' passport use. For face matching, overall scores were reduced given the more difficult nature of the shorter form of the GFMT, removing the ceiling effects found in Experiment 1. Viewers were also poorer at detecting a mismatch when one of the faces was embedded in a passport frame. This suggests that face matching in context is different to matching faces shown alone and the effect is detrimental. As in the earlier experiment participants had a bias to respond same but only when the passport frame was present, a trend that was present in the earlier experiment but was not significant. The more difficult nature of the GFMT short may have allowed this effect to be seen more clearly since the ceiling effects were reduced. It is possible that the additional visual context makes the matching process more difficult, and particularly in ways which lead to a ‘same face’ response bias. The problem also arises regardless of the data type, which again suggests that valid data provides no benefit to face matching and the invalid data no detriment. It does not seem that the data match itself leads to a bias in face matching performance. Instead, the mere presence of a passport frame appears to make the match harder.

Second, there is an even stronger tendency to miss invalid (‘fraudulent’) data on the passport itself. Again, participants knew that they would be seeing some invalid data, and yet they failed to spot this on fewer than half the occasions it was present. This is even poorer performance than in the longer form of the task suggesting there may be some benefit in repeated data checking. Unlike the longer form of the task, there was no difference in accuracy whether the data was shown with same or different face pairs and this was across all data types.

In general, these results show that matching a face to an ID-document is difficult. It is well known that unfamiliar face matching is a hard task: embedding photos in an ID document makes it even harder. Furthermore, this seems to be a systematic effect - any passport frame context (whether bearing correct or incorrect biographical data) tends to bias viewers to respond ‘same’ more often than is the case for simple face-only matching. These results also confirm that the effects found in the longer form of the experiment are both replicated and extended in the short form and online environment. A small number of participants were dropped (3) because their data suggested that they were biased towards fast completion i.e. identifying all face pairs as ‘same’ face pairs or all data as ‘Correct’, but equally this behaviour was easily identifiable e.g. 100% accuracy for ‘Same’ face pairs and 0% accuracy for ‘Different’ face pairs. The vast majority of participants showed a variability in their accuracy similar to that shown by participants in the lab and similarly variable decision times suggested they were not either unusually distracted or rushing. The use of online panels also added greater confidence in the identity of participants since the company used to recruit them require verification of identity and monitor responses to studies over time to ensure consistencies of response as noted by Lowry et al., (2016). This suggests practitioners could choose to use either the long or short form in either a test centre or an online environment with confidence in the validity and replicability of their results.

To provide further clarification of the nature of this effect it would be useful to gather further information regarding whether it is the data or the context of the frame that is interfering with the face matching task. In the current experiment and the longer version of the task, participants were asked to check data before matching the faces and always on separate screens. By changing the order and presentation of the tasks, it may be possible to identify whether it is attention to the data or indirect attention to the frame that is creating such an influence. This manipulation may also shed some light on the potential interference of the face pairs on the data checking task. If the faces are matched first this may then allow greater attention to be directed to the potentially less attention grabbing task. In the following experiment, I vary the way the tasks are presented on screen and the order in which they are presented. This manipulation has two aims. First, to identify whether it is the biographical data or the passport frame that influences the face matching task. Second, to identify whether completing the face matching task first leads to an increase in accuracy in the data matching task.

2.4 Experiment 3

Introduction

The previous experiment demonstrates that performance for face matching is affected by the context of a passport frame and that this is regardless of the type of biographical data in the frame. The very poor accuracy for invalid data in the data checking task also suggests that the presence of a face pair may be distracting from the data task. In the current experiment I examine whether these effects will be present if participants match faces before carrying out the data checking task. I also examine whether these effects might be enhanced if the both tasks are presented on the same screen, one under the other. I also invite members of the Qualtrics panel (UK individuals with a wide range of demographics who have signed up to participate in experiments and surveys) to participate in the experiment as a stronger representation of the diverse range of individuals checking ID in the UK.

Method

Participants

Participants were 80 members (31 male) of an online experimental volunteer community (UK Qualtrics panel). All were British citizens who had lived in the UK for the last 10 years and reported normal or corrected to normal vision (mean age = 39.86, range = 18-75). Participants were reimbursed a small fee for their time.

Materials

The 40 face pairs from the GFMT short form and passport frames from experiment 2 were used. Face pairs were counterbalanced across the experiment such that each pair occurred equally often in each condition. The order of questions and whether those questions were placed on the same screen or different screens was varied between participants as shown in figure 2.6.

Procedure

This experiment was conducted on-line using the Qualtrics survey system. Each participant saw 10 face-only pairs and 30 pairs embedded in passport frames (half left, half right), with order of presentation randomized throughout (i.e. not blocked by condition). Half the items showed same-person pairs, and half different person pairs. For those pairs with passport frames, one third of the items showed invalid data (i.e. ten items per participant). Face pairs were counterbalanced across the experiment such that each pair occurred equally often in each condition. For face-only pairs, participants were asked, on-screen, '*Are the images of the same individual?*' and selected responses *'Same'* or *'Different'* with a mouse. For pairs including a passport frame, participants were asked two questions: '*Is the data correct?*' and '*Are the images of the same individual*?', and made their responses with a mouse. All stimuli remained on the screen until responses had been made – i.e. until after both responses for the pairs including a passport frame.

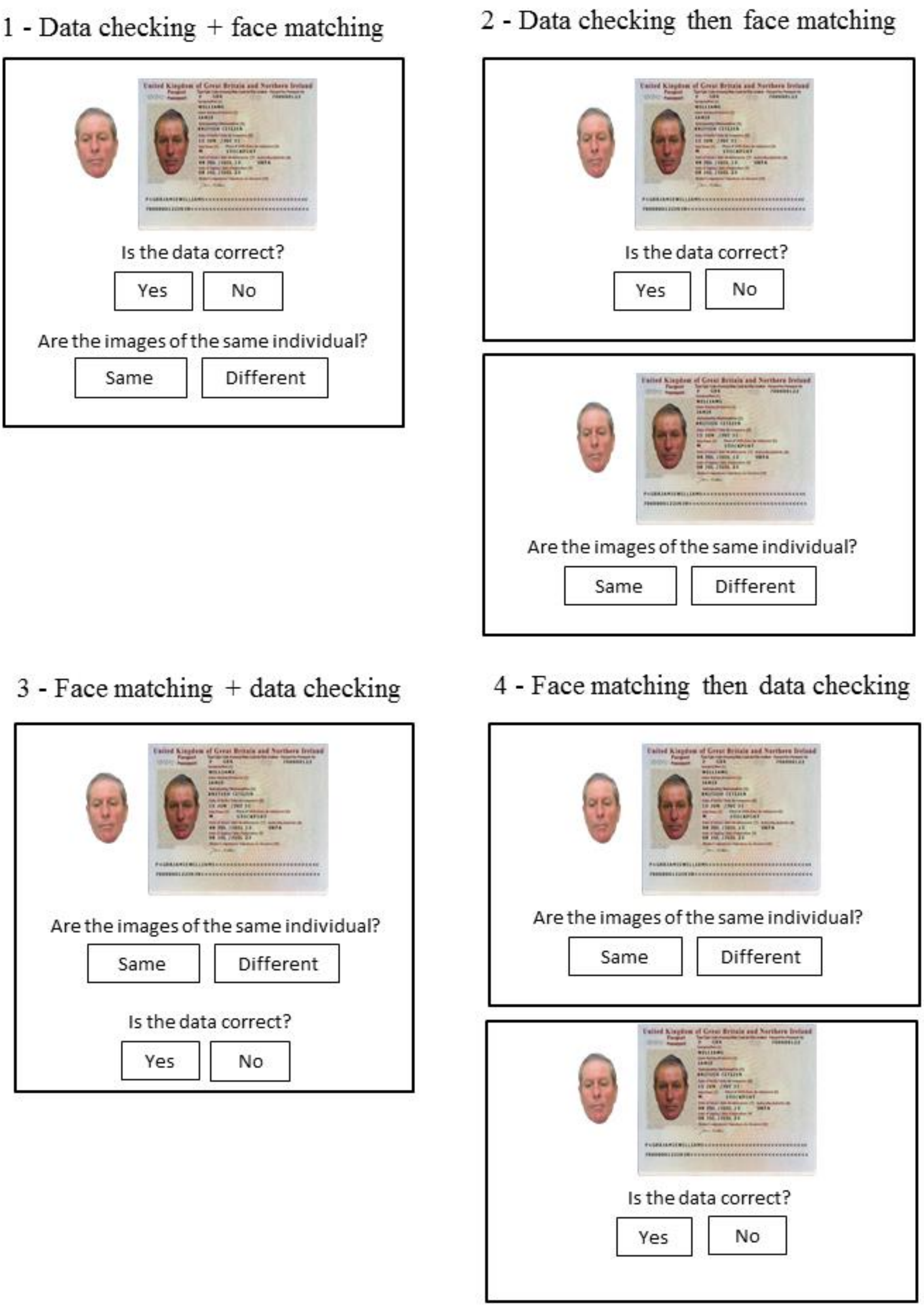

Figure 2.6. Presentation and order of questions by condition.

Prior to the experiment participants were given two practice trials, a face-only pair, and a passport frame pair. They were asked to practise same/different face responses by selecting the appropriate button on screen. For the passport item, they were asked to judge whether the personal information was correct with respect to the person shown in the passport. Instructions for this decision were as follows: ‘If the data matches the image on the passport, e.g. correct gender first name and year of birth click ‘Yes’. If the data is factually incorrect or appears to contradict the image, click ‘No’. No feedback was provided and participants proceeded directly to the experimental phase of the study. The tasks was self-paced, and typically took about 45 minutes to complete.

Results

Face Matching Accuracy

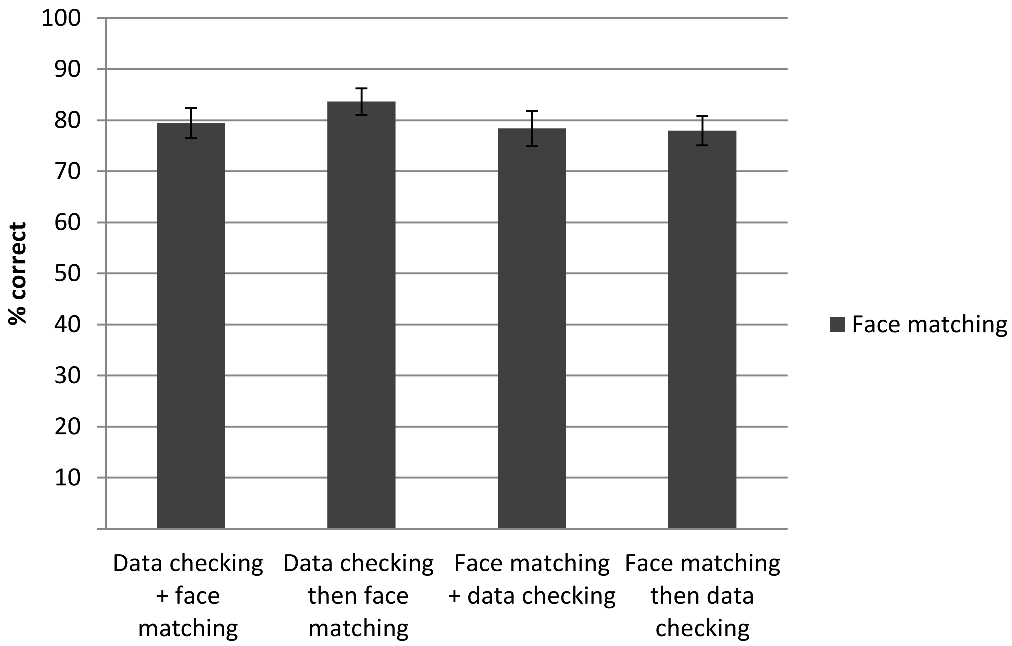

Figure 2.7. Mean accuracy for the face matching task by question order and placement. Note error bars represent standard error.

A 4 (question order and placement type: data checking + face matching, data checking then face matching, face matching + data checking, face matching then data checking) x 2 (match type: same, different) x 3 (presentation type: faces only, face with valid data, face with invalid data) mixed factorial ANOVA for face matching accuracy showed there was no significant main effect of question order and placement type $F(3,76) = 0.90$, $p > .05$, $\eta_{\rho}^2 = .03$ (see figures 2.7 and 2.8). There was no significant interaction between the question order and placement type and the match type $F(3,76) = 0.93$, $p > .05$, $\eta_{\rho}^2 = .04$ or the presentation type $F(6,152) = 0.42$, $p > .05$, $\eta_{\rho}^2 = .02$ or between the question order and placement type, the match type and presentation type $F(6, 152) = 0.67$, $p > .05$, $\eta_\rho^2 = .03$. There was no significant main effect of match type $F(1,76) = 3.21$, $p > .05$, $\eta_\rho^2 = .04$ or presentation type $F(2, 152) = 0.79$, $p > .05$, $\eta_\rho^2 = .01$, however there was a significant interaction between match type and presentation type $F(2,152) = 14.25$, $p < .001$, $\eta_\rho^2 = .16$.

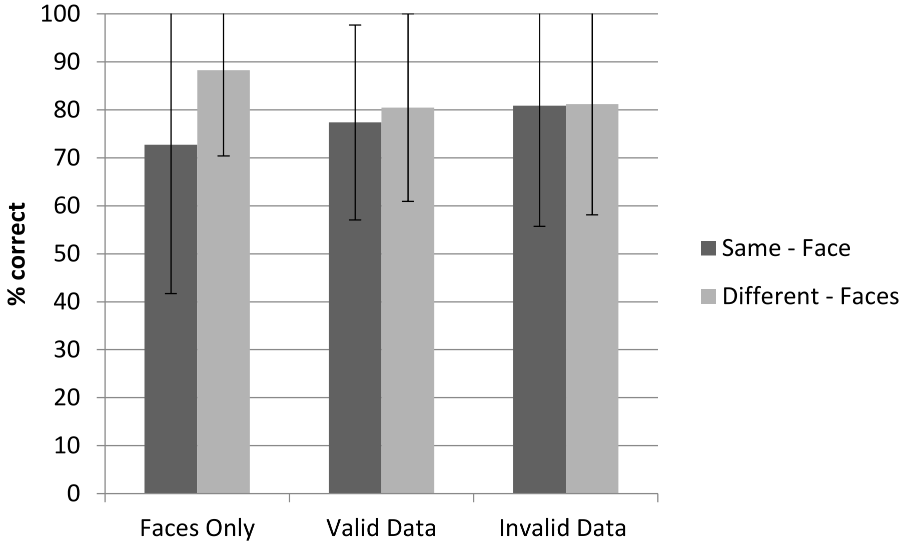

Figure 2.8. Mean accuracy for the face matching task for same and different identity trials when face pairs were shown alone and with valid data and invalid data. Note error bars represent standard deviation.

Simple main effects analyses showed that presentation type affected both same and different match items $F(2, 316) = 6.33$, $p < .005$, $\eta_\rho^2 = .04$; and $F(2, 316) = 6.98$, $p < .001$, $\eta_\rho^2 = .04$, respectively. The presence of a passport frame leads to significantly poorer performance in the face matching task for ‘different’ items but improved performance for ‘same’ items. Like the previous experiment, this pattern of data suggests that the presence of a passport frame introduces a bias to respond ‘same’ to face pairs, regardless of the type of biographical data in the frame. Here the effect is even stronger with both same and different face matching accuracy significantly affected. The modifying effect of the frame is supported by there being no difference in face matching accuracy whether participants completed the data checking task or the face matching task first and whether the questions were shown on same or different screens. If the data had been important we might also have expected a decrease in the bias to respond ‘same’ to face pairs when the data checking task was later than the face matching task or not presented on the same screen - this was not the case.

Data Checking Accuracy

A 4 (question order and placement type: data checking + face matching, data checking then face matching, face matching + data checking, face matching then data checking) x 2 (match type: same, different) x 2 (data type: valid, invalid) mixed factorial ANOVA for face matching accuracy showed there was no significant main effect of question order and placement type $F(3,76) = 0.95$, $p > .05$, $\eta_\rho^2 = .01$ (see figures 2.9 and 2.10). There was no significant interaction between the question order and placement type and the match type $F(3,76) = 0.84$, $p > .05$, $\eta_\rho^2 = .03$ or the data type $F(3,76) = 1.50$, $p > .05$, $\eta_\rho^2 = .06$ or between the question order and placement type, the match type and data type $F(3,76) = 0.49$, $p > .05$, $\eta_\rho^2 = .02$.

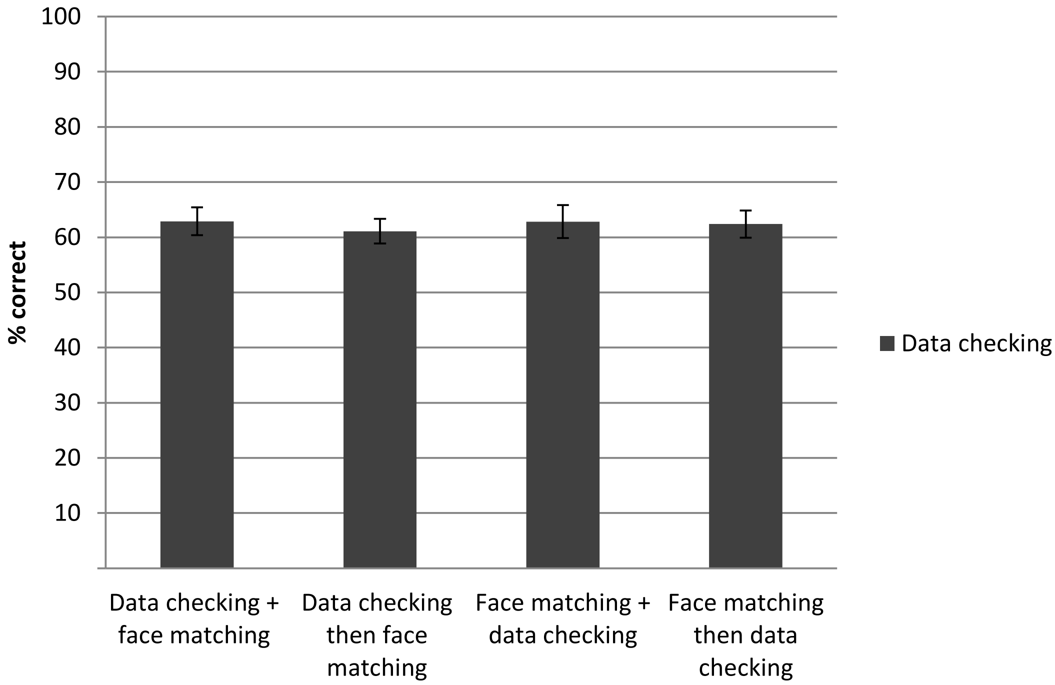

Figure 2.9. Mean accuracy for the data checking task by question order and placement. Note error bars represent standard error.

There was a significant main effect of match type $F(1,76) = 4.45$, $p < .05$, $\eta_\rho^2 = .06$ and data type $F(1, 76) = 97.55$, $p < .001$, $\eta_\rho^2 = .56$. This was qualified by a significant interaction between match type and data type $F(1,76) = 5.28$, $p < .05$, $\eta_\rho^2 = .07$ which arises because performance on valid data is better when faces match than when they do not – a pattern which is absent for invalid data, (simple main effects: $F(1, 158) = 10.34$, $p < .005$, $\eta_\rho^2 = .06$; and $F(1, 158) = 0.55$, $p > .05$, $\eta_\rho^2 = .00$, respectively). Participants were consistently poor at data checking whether the data checking task was presented prior to or after the face matching task and whether the data checking task was presented on the same screen. Any interference created by the face pair remains whether the faces are checked first or second.

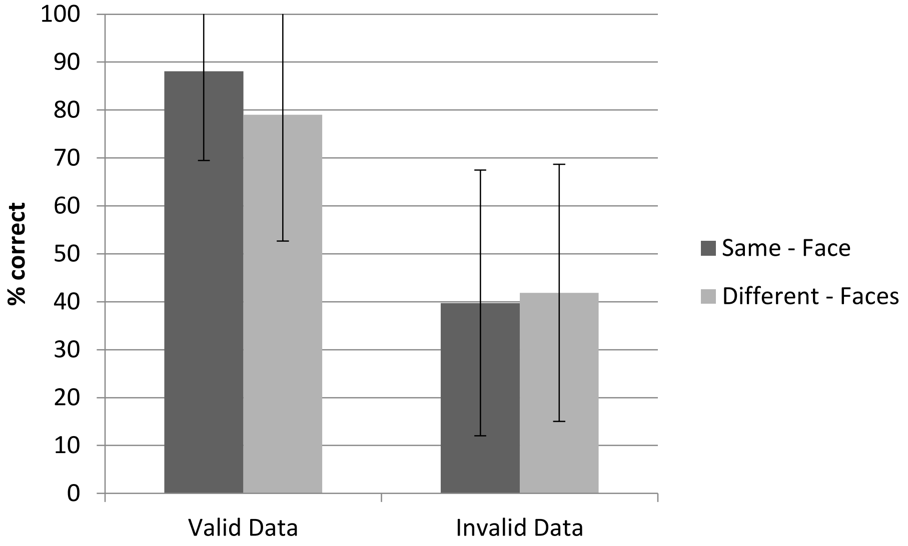

Figure 2.10. Mean accuracy for the data checking task for valid data and invalid data when shown with same and different face pairs. Note error bars represent standard deviation.

Discussion

This experiment shows a number of interesting effects. As in experiment 1 and 2 participants were poor at spotting 'fraudulent' passport use. For face matching, viewers were poorer at detecting a mismatch when one of the faces was embedded in a passport frame. The reduction in accuracy for different face pairs was extended to an increase in accuracy for same face pairs in the current experiment suggesting the effect may be stronger in the wider population who also showed a lower level of face matching accuracy overall.

The effect of the passport frame remained whether the face matching task was completed before or following the data checking task. This supported the findings of experiment 2 and the current experiment where face matching accuracy was affected by the passport frame regardless of the data type. If the data itself had been important we might also have expected a decrease in the bias to respond 'same' to face pairs when the data checking task followed the face matching task or not presented on the same screen - this was not the case.

As in experiments 1 and 2 there was a strong tendency to miss fraudulent data. Indeed, participants in the current experiment achieved lower data checking accuracy scores overall. Participants in the current experiment were less accurate in their identification of valid data when it was viewed with a different face pair. This pattern of data seems to suggest that different face pairs have an effect on data checking, even though the faces are task-irrelevant – detecting that the face pairs are different (or invalid) seems to generalise (falsely) to biographical data when it is in fact valid. This should not be the case since the face matching and data checking tasks are entirely separate and presented on different screens. In addition, as my results show this does not change whether the face matching task is completed first or last or shown on the same screen as the data checking task or different screens.

Extending participation to a wider demographic appears to result in a drop in performance for both face matching and data checking. These reductions in accuracy appear to be due to greater levels of interference. For face matching it is clear that the modifying affect is generated by the passport frame, regardless of the nature of biographical data. However, it is a little more difficult to be clear about the effects on data matching in this experiment. Although there are clearly poor levels of fraud detection within these documents, the erroneous data was inserted in an unsystematic way. There were a number of different types of information that could be wrong in the 'invalid' passports, and it is quite possible that some of these are easier to spot than others. Therefore, in the following experiment, I increase the level of invalid data and make a systematic examination of the different types of valid and invalid data available. I also explore other layouts to establish whether some methods of presentation lead to higher accuracy than others. Photographs and passports can be displayed side by side or one above the other. Does varying these positions lead to greater interference or reduce interference between the stimuli?

2.5 Experiment 4

Introduction

This experiment examines viewers' ability to match faces and check three types of information on a photo-ID document: gender, year of birth and place of birth. Research into the perception of age and gender has shown that individuals are highly skilled at both. Gender classification from face images is typically near ceiling (e.g. Bruce, Ellis, Gibling, & Young, 1987; O'Toole, Deffenbacher, Valentin, McKee, Huff, & Abdi, 1998) and age estimation is highly robust over a range of manipulations (George & Hole, 1995, 2000). Viewers are typically able to judge an adult's age from a photo to within five years (Moyse, 2014; Voelkle, Ebner, Lindenberger & Riediger, 2012) and able to sort photographs into age ranges (18–25 years, 35–45 years, 55–75 years) with a correct response rate of 83.1% (Anastasi & Rhodes, 2006). These findings suggest that any mismatch between the perceived age or gender of a face and the accompanying biographical data could be identified relatively easily.

Unlike gender and age, place of birth is not visually derivable from an unfamiliar face so error checking is entirely based on transcription accuracy. Errors in the spelling of place of birth can indicate a forged passport, particularly for unsophisticated copies by non-native English speakers (Fender, 2008; Leslie & Thimke, 1986; my experience with the UK Passport Office reveal that such transcription errors are used as indicators of fraudulent documents). In the following experiment, I manipulate errors in each of these three types of information, aiming to establish their detectability and the relative levels of effort required. I also look at whether these effects are modified by the positioning of the passport frame since passports and photos are easy to place side by side, whereas in the border control context a passport is usually held below the face of the holder for comparison purposes. Research has also shown that nodding the head up and down while reviewing stimuli can result in more agreement or greater preferences for the stimuli whereas shaking the head from side to side results in less agreement or preference (Wells & Petty, 1980; Tom, Pettersen, Lau, Burton, and Cook, 1991). This suggests that moving the head up and down or left to right to match face pairs may result in different levels of accuracy if face pairs are generally agreed to be, or preferred to show, the same individual.

Method

Participants

Participants were 96 members of the Qualtrics panel (48 male), who were over 18, reported normal or corrected to normal vision and did not have dyslexia (mean age = 40.8, range = 18-73). Participants were also required to be British Citizens who had lived in the UK for at least the last 10 years. Participants were reimbursed a small fee for their time.

Materials

96 face pairs from the GFMT long form, half male and half female pairs, with half matching and half mismatching were used. Participants saw 48 face-only pairs and 48 pairs

in which one of the faces was embedded in a passport frame. 24 of the passport frames contained valid data. Two of the error classes used in the earlier experiments were dropped; noun as first name and unlikely ethnicity surname, in both cases because of their predicted low incident rate in the applied setting. A full set of the biographical information used can be found in Appendix 3. Key personal data could be rendered invalid as follows:

- wrong gender (forename and M/F label) – 8 items;
- invalid year of birth (YOB) a year of birth that would indicate an age 20 years older than the face on the passport – 8 items:
- invalid place of birth (POB) for which a misspelled UK town was used, e.g. ‘Luuton’ rather than ‘Luton’ – 8 items.

Spelling errors were orthographic, creating non-words that would not usually occur in the English language to ensure they were most easily identified by native speakers (Fender, 2008; Leslie & Thimke, 1986). In all cases, there was only one data error in each frame. Faces were counterbalanced across the experiment such that each face appeared equally often in each condition. The position of the passport frame was varied between participants as shown in figure 2.11 with passports only presented on the right of the screen and the image on the left, in contrast to experiments 1, 2 and 3, since this follows the specifications for passports followed by the UK (International Civil Aviation Organisation, 2015).

Procedure

This experiment was conducted on-line using the Qualtrics survey system. Each participant was randomly assigned to one of the four conditions shown in figure 2.11 and saw 48 face-only pairs and 48 pairs embedded in passport frames with order of presentation randomized throughout (i.e. not blocked by condition). Half the items showed same-person pairs, and half different person pairs. For those pairs with passport frames, 50% of the items showed invalid data). For those pairs with passport frames, one third of the items showed invalid data (i.e. ten items per participant). Face pairs were counterbalanced across the experiment such that each pair occurred equally often in each condition. For face-only pairs, participants were asked, on-screen, ‘*Are the images of the same individual?*’ and selected responses *‘Same’* or *‘Different’* with a mouse. For pairs including a passport frame, participants were asked two questions: ‘*Is the data correct?*’ and ‘*Are the images of the same individual*?’ and made their responses with a mouse. Order of questions was counter-balanced, and all stimuli remained on the screen until responses had been made – i.e. until after both responses for the pairs including a passport frame.

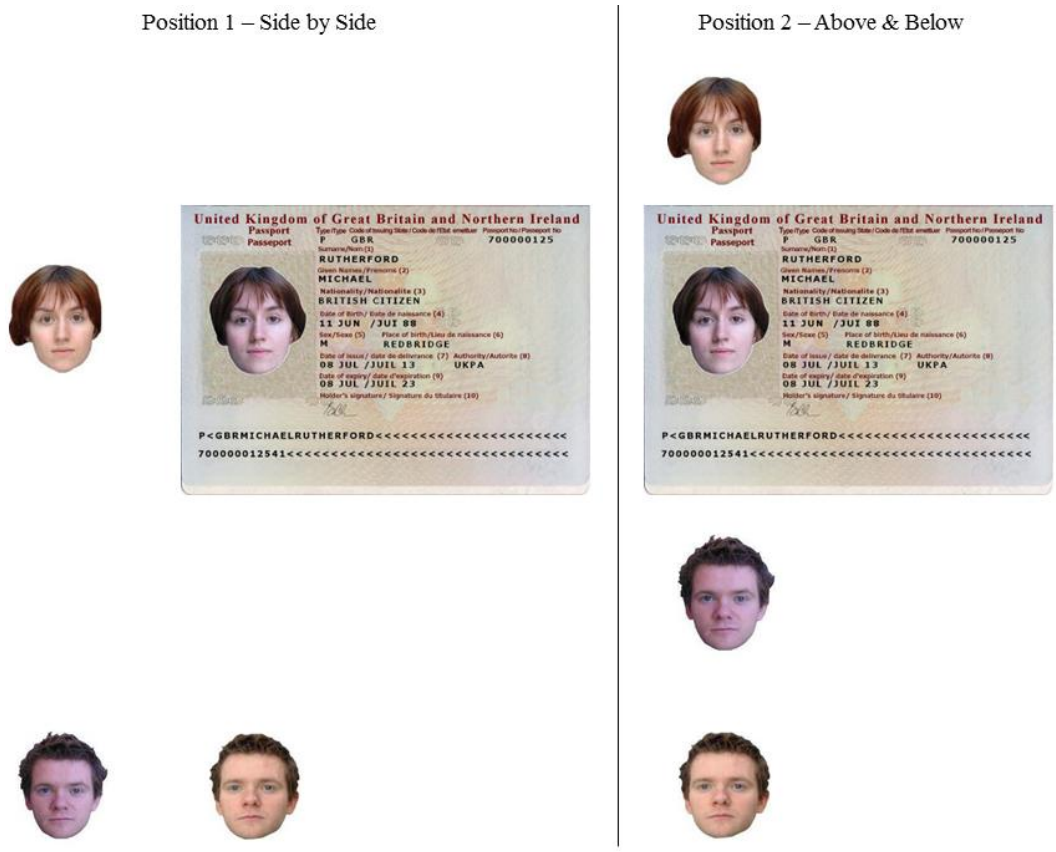

Figure 2.11. Presentation of stimuli by condition.

Prior to the experiment participants were given two practice trials, a face-only pair, and a passport frame pair. They were asked to practise same/different face responses by selecting the appropriate button on screen. For the passport item, they were asked to judge whether the personal information was correct with respect to the person shown in the passport. Instructions for this decision were as follows: '*If the data matches the image on the passport, e.g. correct gender first name and year of birth click 'Correct'. If the data is factually incorrect or appears to contradict the image, click 'Incorrect'*. The task was self-paced, and typically took about 30 minutes to complete.

Results

Face Matching Accuracy

A 2 (position type: side by side, above & below) x 2 (match type: same, different) x 3 (presentation type: faces only, face with valid data, face with invalid data) mixed factorial ANOVA with a Huyn Felt correction for face matching accuracy showed there was no significant main effect of position type $F(1,94) = 0.17$, $p > .05$, $\eta_\rho^2 = .01$. There was no significant difference for face matching overall whether the face pairs were shown side by side or above and below (means 85.67% and 84.64% respectively).

There was no significant interaction between the position type and the match type $F(1,94) = 1.10$, $p > .05$, $\eta_\rho^2 = .01$ or the presentation type $F(1.8,170.88) = 0.86$, $p > .05$, $\eta_\rho^2 = .01$ or between the position type, the match type and presentation type $F(1.96, 184.04) = 0.25$, $p > .05$, $\eta_\rho^2 = .00$. There was no significant main effect of presentation type $F(1.8, 170.88) = 0.62$, $p > .05$, $\eta_\rho^2 = .01$, (see figure 2.12) However there was a significant main effect of match type $F(1,94) = 9.11$, $p < .005$, $\eta_\rho^2 = .09$ and a significant interaction between match type and presentation type $F(1.96,184.04) = 15.32$, $p < .001$, $\eta_\rho^2 = .14$.

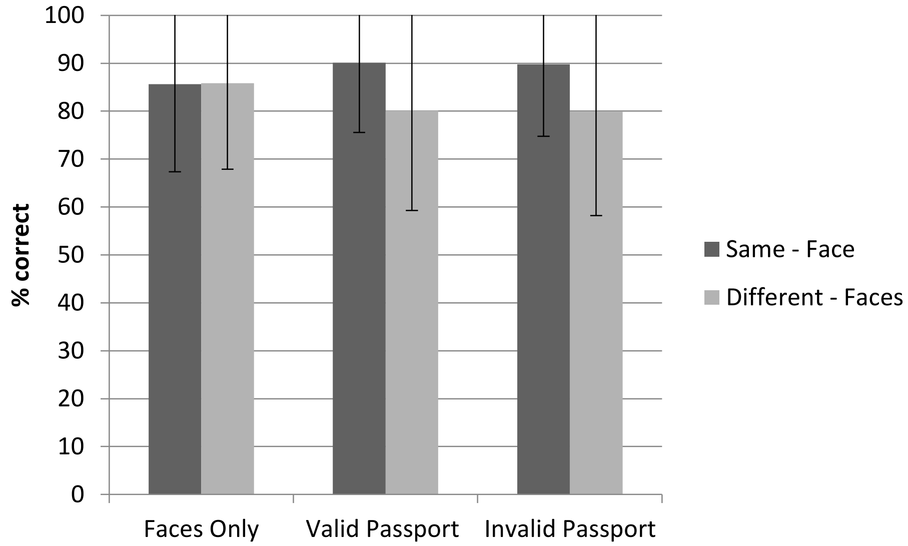

Figure 2.12. Mean accuracy for the face matching task for same and different identity trials when face pairs were shown alone and with valid data and invalid data. Note error bars represent standard deviation.

Simple main effects analyses showed that presentation type affected same and different match items $F(2, 380) = 6.13$, $p < .005$, $\eta_\rho^2 = .03$, and $F(2, 380) = 11.36$, $p < .001$, $\eta_\rho^2 = .06$, respectively. As in Experiment 3, the presence of a passport frame leads to better performance for valid passports and worse performance for invalid passports. Again, this pattern of data suggests that the presence of a passport frame biases viewers to respond 'same' to face pairs.

Figure 2.13 shows mean face matching accuracy for the different types of invalid passport information. A 2 (match type: same/different) x 3 (invalid data type:

gender/POB/YOB) ANOVA showed no main effect of data type $F(2,190) = .68$, $p > .05$, $\eta_\rho^2 = .01$ or interaction $F(2,190) = .45$, $p > .05$, $\eta_\rho^2 = .00$. Only the main effect of match type was significant $F(1, 95) = 13.28$ $p < .001$, $\eta_\rho^2 = .12$, confirming higher overall performance for same-face trials – i.e. the response bias for pairs with a passport frame, which is also clear in Figure 2.13

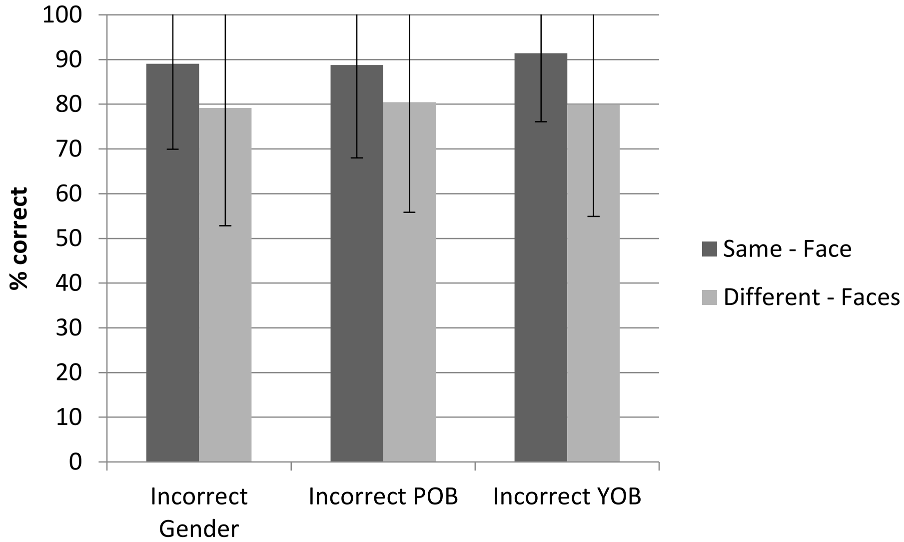

Figure 2.13. Mean face matching accuracy across types of invalid data. Note error bars represent standard deviation.

Since this experiment was designed to reflect work based practice, participants were given unlimited time to make their responses. However, latency measures have been identified as a useful indicator of effort and qualitative changes in behaviour (Marotta, McKeeff & Behrmann 2002; Davidoff & Landis, 1990; Bruce, 1982). While not directly comparable to studies where participants have taken part in speeded tests, the time taken by participants here may provide some insight into the effort and qualitative nature of the tasks. Decision times are reported here across all response types - correct and incorrect, and indicated that viewers spent longer looking at the face-only pairs than the pairs embedded in valid or invalid passport conditions - means 7.5 sec ($SD = 3.5$), 5.8 sec ($SD = 2.8$) and 6.6 sec ($SD = 4.8$) respectively; $F(2,190) = 12.00$, $p < .001$, $\eta_\rho^2 = .11$. For faces-only and for valid-passport stimuli there was no difference in decision time for same and different face pairs. However, when viewers were inspecting pairs embedded in invalid passport photos, they spend longer on different pairs than same pairs - means 7.6 sec ($SD = 7.6$) and 5.6 sec ($SD = 3.0$) respectively, $F(1,285) = 17.72$, $p < .001$, $\eta_\rho^2 = .06$.

Data Checking Accuracy

A 2 (position type: side by side, above & below) x 2 (match type: same, different) x 2 (data type: valid, invalid) mixed factorial ANOVA for data checking accuracy showed there was no significant main effect of position type $F(1,94) = 1.35$, $p > .05$, $\eta_\rho^2 = .01$. There was no significant difference for data checking overall whether the data was presented with face pairs shown side by side or above and below (means 62.42% and 64.59% respectively).

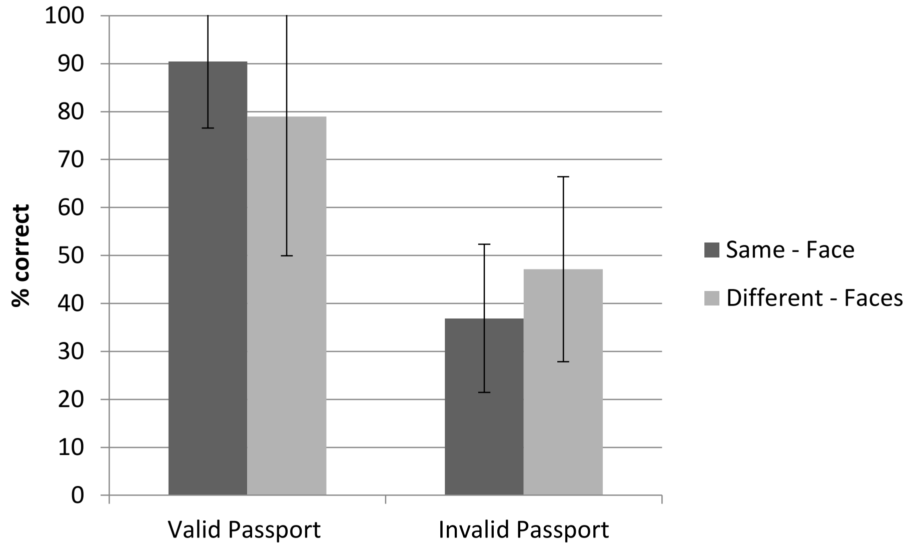

Figure 2.14. Mean accuracy for the data checking task for valid data and invalid data when shown with same and different face pairs. Note error bars represent standard deviation.

There was no significant interaction between the position type and the match type $F(1,94) = 0.54$, $p > .05$, $\eta_\rho^2 = .01$ or the data type $F(1,94) = 0.12$, $p > .05$, $\eta_\rho^2 = .00$ or between the position type, the match type and data type $F(1,94) = 0.99$, $p > .05$, $\eta_\rho^2 = .01$. There was no significant main effect of match type $F(1,94) = 0.54$, $p > .05$, $\eta_\rho^2 = .01$, (see figure 2.14). However, there was a significant main effect of data type $F(1, 94) = 217.10$, $p < .001$, $\eta_\rho^2 = .70$ which was qualified by a significant interaction between match type and data type $F(1,94) = 22.08$, $p < .001$, $\eta_\rho^2 = .19$. Simple main effects analyses showed that data type affected both same and different match items $F(1, 190) = 216.91$, $p < .001$, $\eta_\rho^2 = .53$, and $F(1, 190) = 76.74$, $p < .001$, $\eta_\rho^2 = .29$, respectively.

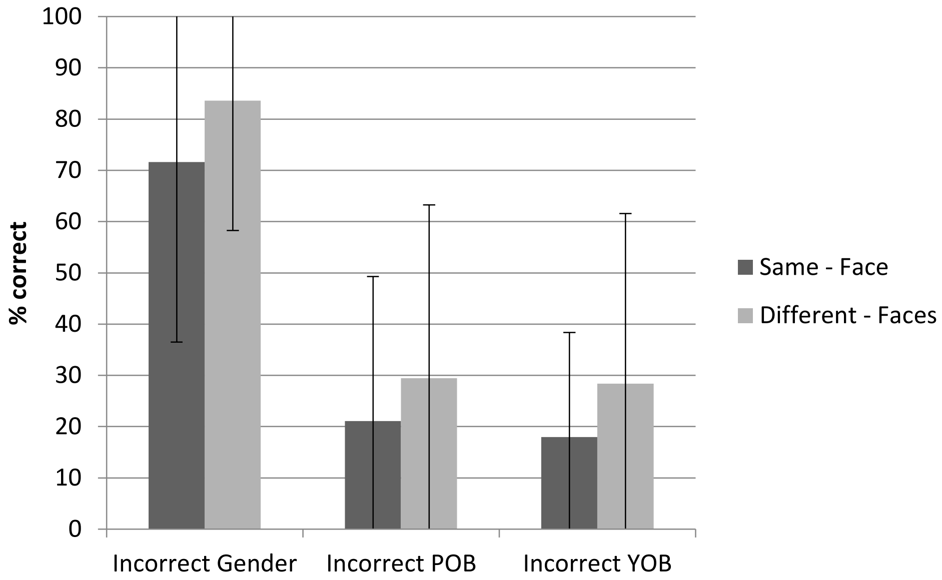

Figure 2.15. Mean data checking accuracy for invalid data types as shown with same and different face pairs. Note error bars represent standard deviation.

Most strikingly, these results show that detection of invalid passport information is very poor – consistent with the earlier experiments. To explore this further, Figure 2.15 shows performance across the three different types of invalid data. A 2 (match type: same/different) x 3 (invalid data type: gender/POB/YOB) ANOVA showed a main effect of data type $F(2,190) = 124.57$, $p < .001$, $\eta_\rho^2 = .57$ and a main effect of match type $F(1, 95) = 18.95$ p $<.001$, $\eta_\rho^2 = .17$ but no interaction $F(2,190) = .57$, $p > .05$, $\eta_\rho^2 = .01$. Gender was the easiest false information to detect, but POB and YOB were equally poor. Most interestingly, the presence of the matching face introduced no difference in effects for the three data matching tasks. In each case, there was a consistent bias to respond 'different'.

Data Checking Decision Times

Decision times are reported here across all response types - correct and incorrect. Decision times indicated that time to verify a valid passport was marginally longer than time to spot an invalid one, though this difference did not reach significance - means 9.7 sec (*SD* = 6.0) and 9.3 sec (*SD* = 5.1) respectively $F(1,95) = 0.84$, $p > .05$, $\eta_\rho^2 = .01$. Detection of incorrect sex was faster than incorrect POB, which was in turn faster than incorrect YOB - means 8.3 sec (*SD* = 5.7), 9.1 sec (*SD* = 5.3) and 10.6 sec (*SD* = 6.7) respectively, $F(2, 190) = 9.85$ p $<.001$, $\eta_\rho^2 = .09$.

Discussion

This experiment shows clear effects, consistent with the earlier studies. As in Experiment 2 and 3 participants were compromised in performing a face match when a frame was added. Viewers were significantly more likely to identify same-face pairs as same, but less likely to identify mismatch trials, i.e. they are less likely to spot a fraudulent use. This is consistent across valid and all different types of invalid passport – so the simple fact of having a frame present induces this shift in face matching performance.

Data matching performance is largely consistent with the earlier experiments, despite the proportion of invalid data being increased. However, the strict counterbalancing of same and different face pairs over the valid and invalid data allows the effects and nature of the interference between face pairs and data checking to become clearer. In particular, viewers are very poor at spotting invalid passports. Incorrect gender was spotted relatively accurately overall (about 78% of the time) but incorrect place of birth and year of birth was detected very infrequently (roughly 25% of the time) despite the fact that viewers had been alerted to look out for these problems. Interestingly, accuracy of data checking was affected by whether two faces showed same or different people – despite the fact that this was task irrelevant. Performance on confirming valid data was better when faces matched than when they did not. However, performance in spotting invalid data was better when the faces did not match. This suggests that the information from the faces 'leaks into' the data decision – which tends to be pulled in the direction of the faces. This is possibly because the faces are the more salient cue, even though their matching/non-matching status is independent of the data checking task. Interestingly, this effect of faces on data checking is consistent across the different types of invalid passport – i.e. it is the same in the easy gender check task as in the harder POB and YOB tasks. This suggests a strong effect of the secondary face. That there is no difference in data checking accuracy or face matching accuracy between the position conditions demonstrates that this interference is not affected by positioning. This gives confidence that images may be placed either above and below each other or side by side on a computer screen and suggests these findings may generalise to applied practice. However, the interference effect of matching a photograph with a live individual may differ and is an area of investigation in its own right.

2.6 General Discussion

The results presented here show, across multiple experiments, that face matching is affected by the presence of an ID context. Like Galli et al., (2006) found in their face recognition task, additional perceptual context changes performance for unfamiliar face matching. However, unlike Galli et al.,'s findings, perceptual context does not enhance performance overall, but reduces performance in a very specific way. The growing experimental literature on facial matching very often uses isolated face stimuli – for the good reason that researchers wish to study this process in the absence of potentially interfering material. However, in real world identity checks, faces are very often compared to documents carrying other biographical information – as in passports, driving licences or workplace ID. The evidence presented here suggests that this significantly alters the patterns of accuracy obtained.

Galli et al., (2006) also found that the semantic context, the emotional valence of the headline, affected face recognition accuracy. Here, the semantic context or validity of the data does not affect face matching accuracy. When a face is embedded in a document, this seems to bias the viewer to make a 'same person' decision. This is not strictly confirmation bias since the face matching is not affected by the nature of the biographical data. In some ways this may be more analogous to the 'halo effect' (Thorndike, 1920) whereby a global evaluation of a person is made regardless of individual attributes of a person. In this case, the presence of the passport frame provides a global attribution of validity, leading the viewer to be more likely to say the face pair is valid or a match.

The second important observation here is that the presence of a face match (two images) affects checking of biographical data on photo ID. Overall, the data checking accuracy across all of the experiments was low. It is particularly interesting to note that the validity of the face match affects data-checking even when viewers were explicitly instructed to ignore the second face, as in Experiment 4. It seems that same-face pairs (i.e. those which are valid for the face) influence the likelihood that viewers will detect fraudulent biographical data. In short, if the faces match, viewers are more likely to say that the data is correct too – even when it is not. Interestingly, this effect is completely consistent across different types of data-invalidity, and is independent of whether the biographical-error is generally easy to spot (gender) or hard to spot (place and year of birth). For data checking then, there is a clear forensic confirmation bias driven by the semantic context of the face pairs presented with the passport frame. The low levels of data checking accuracy also suggest that the perceptual context of the face pair reduces performance.

Could the effects of faces on data checking be due to dual-task interference (e.g. Pashler, 1994)? At first glance, this looks rather unlikely. Although the face matching and data-checking tasks shared stimuli they were always presented separately in these experiments, and so there seems no opportunity for one task to be affected by load due to the other. Even when the order and presentation of the tasks was varied, there were no differences in either face matching or data checking accuracy. However, it is possible that the presence of a face leads to automatic processing (Farah, Wilson, Drain & Tanaka, 1995a). If so, there could then be competition for resources to process face and biographical-data stimuli. The traditional view is that the shallower of two tasks is affected by dual-task interference (Jones, Miles & Page, 1990), and so an attention-demanding face task (albeit an incidental one) may affect performance on other tasks. Some research suggests that faces are processed mandatorily at a semantic level, and not just for superficial visual characteristics (Boehm, Klostermann, Sommer & Paller, 2006; Burton, Kelly & Bruce, 1998). This may provide a route to understanding how an apparently superficial face-matching process could interfere with apparently complex tasks such as calculating someone's age and making a judgement about its veracity. This may also explain why there is no difference in either face matching or data checking accuracy when the positioning of the stimuli is altered. It is not the positioning that is important; it is the presence of a face *pair* that makes the difference.

These results have some implications for practitioners. Of course, these studies are performed on non-expert viewers, and it will be important to establish in future whether these effects generalise to people who conduct ID checks professionally (White et al., 2015). Within the fingerprint matching context individuals without expertise have been shown to make false decisions based on stereotypical beliefs about the 'accused' and experts are vulnerable to contextual biases such as prior match or mismatch decisions (Smalarz et al., 2016; Dror et al., 2006). This would suggest that it is likely that experts in face matching and ID checking may well be vulnerable to the biases found here. For such people, it may be possible to separate data checks and face checks in the workflow. The results also have implications for the relationship between laboratory experiments and real world settings providing additional support to the concerns raised by researchers in forensic science. My data emphasise that if researchers intend to generalise their results outside the lab, then it is important to incorporate all relevant task demands. Apparently unrelated components of the task can have significant effects on each other, and I have demonstrated here that previous research on face matching may have under-estimated the problem for those performing the task professionally. However, the nature and extent of these effects is still somewhat unclear. That the nature of the face pair interferes with the face matching task has been clearly demonstrated, however it is only the unexpectedly low performance on the data checking task that leads me to assume the presence of the face pairs is reducing performance overall. Without base rate accuracies for checking data in isolation and checking data in a passport frame with only one face present, it is difficult to know whether this is the case. In Chapter 3, I investigate data checking accuracy in both of these conditions to quantify the reduction in accuracy being viewed with face pairs creates.

Chapter 3 – Categorising faces and data

3.1 Introduction

In the previous chapter, participants were asked to carry out a number of experiments involving face matching tasks and data checking tasks in an ID context. The aim of these experiments was to test whether the document context influenced the ability to match faces. The results of these experiments demonstrated that the document context did affect face matching ability. Perhaps more surprisingly, the data checking task was affected by the nature of the face pair shown with the passport frame. This occurred despite the face pair being task irrelevant, and even when participants were asked to refer only to the face in the passport frame when checking the data. Experiment 4 demonstrated that performance on confirming valid data was better when faces matched than when they did not. However, performance in spotting invalid data was better when the faces did not match. Had the overall performance in data checking accuracy remained close to the norms previous research might lead us to expect, it might have been enough to conclude that it is only the nature of the face pairs that influences data checking accuracy. However, Experiment 4 also provided accuracy rates for the invalid data types that were also at odds with expectations.

Earlier research has shown that categorising faces by gender and age is an ability that participants generally excel at (Bruce et al., 1987; O'Toole, Deffenbacher, Valentin, McKee, Huff & Abdi, 1998; George & Hole, 1995, 2000). O'Toole et al., (1998) found that participants were able to categorise male and female faces with 95.9% accuracy. Experiment 4 found that participants only identified incorrect gender 77% of the time, despite the error being marked with both an incorrect gender first name and incorrect sex. This 20% reduction in accuracy was surprising, but even more so were the identification rates for incorrect year of birth. Voelkle et al., (2012) found that young adults were able to estimate the age of faces between 19 and 80 with an error rate of 5.91 years. The individuals photographed for the GFMT were all within this age range and provided their age at the time, allowing correct age for each image to be known. To ensure the best rates of accuracy I ensured all incorrect dates of birth recorded on the passport frames were 20 years too old. Anastasi and Rhodes (2006) showed that photographs of faces could be sorted into three age ranges (18–25 years, 35–45 years, 55–75 years) with a correct response rate of 83.1%. With the greatest range here being 20 years it might be reasonable to expect that participants could identify an incorrect year of birth date within 20 years at a similar accuracy level. This was not the case. Participants identified an incorrect year of birth only 23% of the time, less than a quarter of the accuracy achieved in Anastasia and Rhode’s study.

Identification rates for incorrect spellings of place of birth were similarly affected. Detection of orthographic or non-word errors embedded in text by native speakers has been shown to be more error prone than error rates for categorising faces, with accuracy rates of approximately 85% (Levy, Newell, Snyder & Timmins, 1986; Levy, Di Persio & Hollingshead, 1992). In experiment 4, participants only identified incorrect spellings of the place of birth 29% of the time.

Whether taken independently or together these accuracy rates suggest that something different is happening in these studies when compared to earlier research. Overall, across all four experiments, the accuracy for invalid data detection was 48%. Given the accuracy rates provided by earlier research, we might expect an overall accuracy for invalid data detection of 78%. In this context participants either appear unable to categorise the stimuli presented or are having difficulties carrying out the task in the passport checking context. This chapter presents a number of experiments that aim to investigate data checking in the ID context. To verify the validity of the face stimuli participants will be presented images of the faces in isolation to establish how accurately the age and gender of the faces can be estimated when shown without the ID context. To verify the validity of the orthographic errors, participants will be presented with the incorrectly spelled places of birth to establish how accurately the spelling errors can be identified when shown without the ID context. Finally, the passport frames used in Experiment 4 will be used in a data checking task to establish how well the data can be checked when only one face, not a face pair, is present.

3.2 Experiment 5

Introduction

The detection rates for invalid age and gender across all experiments in Chapter 2 were lower than expected. These large falls in accuracy suggest that identifying gender and age is more difficult within the context of checking biographical data within an ID document. However, without baseline performance levels for identifying the gender and estimating the age of the faces used in the passport checking experiments it is not clear whether the data checking is compromised by the stimuli, the ID context or the ID context when shown with face pairs. The aim of this experiment is to identify baseline performance for the gender categorisation and age estimation of the faces used in Experiment 4 when shown in isolation.

Method

Participants

Participants were 25 members (12 male) of the online community Reddit who all reported normal or corrected to normal vision (mean age = 27.42, range = 18-50), were British citizens and did not have dyslexia. Participants completed the experiment in good will.

Materials

The 96 faces shown in passport frames in Experiment 4 were shown here in isolation. Each participant estimated the age of 50% of the faces and categorised the gender of the remaining 50%. Faces were counterbalanced across the experiment such that each face appeared equally often in each condition.

Procedure

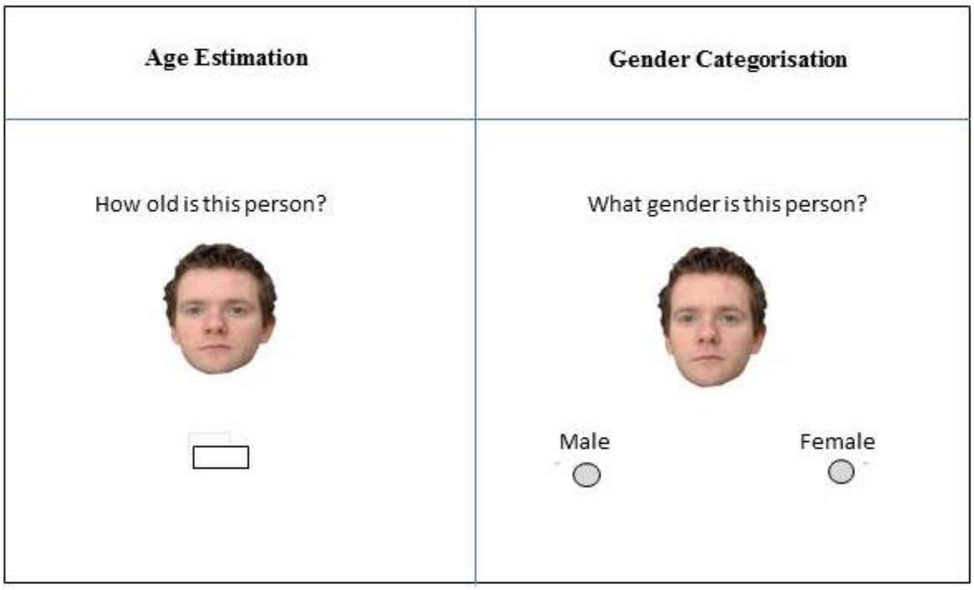

Figure 3.1. Example stimuli from the age estimation and gender categorisation tasks.

This experiment was conducted on line using the Qualtrics survey system. Participants were shown faces one at a time in the centre of the screen (see figure 3.1). For half of the faces they were asked, on-screen, ‘How old is this person?’ and were required to enter their estimation to the nearest year using the keyboard. For the other half of the faces they were asked, on-screen, 'What gender is this person?' and selected responses 'Male' or 'Female' with a mouse. Order of presentation of the faces was randomized within each block, and blocks were presented separately. The task was self-paced, with no time limit for responses, and typically took about 10 minutes to complete

Results

Age & Gender Accuracy

Overall participants had a mean absolute error of 5.10 years (SD = 3.18) when estimating age. Accuracy for age estimation was based on the +/- 20 years allowance used in Experiment 4 where incorrect years of birth were always 20 years older than the age of the person in the image. Here participants were required to estimate an age that was either less than twenty years too young or twenty years too old for the age estimation to be identified as correct. Mean accuracy for age estimation was 97.08 % (*SD* = 3.40) and mean accuracy for gender estimation was 94.08 % (*SD* = 9.77). There was no significant difference between the two, $t(24) = 1.68$ $p > .05$.

Age & Gender Decision Times

As in Experiment 4 participants were given unlimited time to make their responses in this experiment. While not directly comparable to studies where participants have taken part in speeded tests, the time taken by participants here may provide some insight into the effort and qualitative nature of the tasks. Decision times are reported here across all response types - correct and incorrect. There was a significant difference between decision times for age estimation (within +/- 20 years), ($M = 6.75$, $SD = 1.70$) and gender categorisation ($M = 3.43$, $SD = 1.50$); $t(24) = 7.47$, $p < .001$. This may support earlier research that suggests age is more difficult to estimate than gender and therefore requires a longer processing time. However, it should also be taken into account that typing in an age estimation would also take longer than selecting a binary response.

Discussion

This experiment indicates that individuals perform very well when estimating the age and categorising individual faces from the GFMT shown in isolation. Accuracy for both age and gender estimation was close to ceiling, reflecting the findings of earlier research (Bruce et al., 1987, O'Toole et al., 1998; George & Hole, 1995, 2000). The error rate for estimation of age at 5.10 years was lower than the error rate of 5.91 years found by

Voelke et al., (2012) suggesting their face stimuli was more difficult to categorise than the face set from the GFMT. Mean accuracy for age estimation at 97% was greater than found in Anastasi and Rhodes (2006) where mean accuracy overall was 83%. Anastasi and Rhodes used age ranges that ranged from seven years apart to twenty years apart – a more difficult task than estimating age within +/- twenty years. This experiment achieves the aim of identifying baseline performance for age estimation and gender categorisation of the faces shown in the passport frames in Experiment 4. These results provide confidence in the stimuli and demonstrate that when shown in isolation it is easy to estimate the age of the faces within +/- twenty years and categorise them by gender. These results also suggest that age estimation and gender categorisation for faces is compromised in Experiment 4. However, whether this is due to the ID context or when the ID context is shown with face pairs is not clear and will be investigated later in this chapter.

3.3 Experiment 6

Introduction

Unlike gender categorisation and age estimation, place of birth is not visually derivable from an unfamiliar face so error checking would be entirely data based. Errors in the spelling of a place of birth can flag a fraudulent passport. Since there is strong evidence suggesting shared relationships between orthographic knowledge, word recognition and spelling production (Fender, 2008) non-native speakers of English may inaccurately copy handwritten place names using illegal structures e.g. the letter ‘c’ would never follow ‘th’ (Leslie & Thimke, 1986). This type of fraud may be less easy to identify overall since detection of orthographic errors in text by native speakers has been shown to be more error prone with accuracy rates of approximately 85% (Levy et al., 1986, 1992).

In experiment 4, participants only identified incorrect spellings of the place of birth 29% of the time. This suggests that participants either are unable to identify the stimuli presented or are having difficulties in ID context. Without baseline performance levels for checking the spelling of place names used in the passport checking experiments it is not clear whether the data checking is compromised by the stimuli, the ID context or the ID context when shown with face pairs. The aim of this experiment is to identify baseline performance for the identification of the orthographic errors used in Experiment 4 when the words are shown in isolation. This should provide confidence in the stimuli and further evidence regarding whether the data checking was compromised in Experiment 4.

Method

Participants

Participants were 24 members (11 male) of the general public who all reported normal or corrected to normal vision (mean age = 34.13, range = 19-65), were British citizens and did not have dyslexia. Participants completed the experiment in good will.

Materials

32 place names were used from Experiment 4. All place names used were selected on the basis of population (Office for National Statistics, 2013). Highly populated cities and towns were chosen to increase the chances of participants being familiar with their names and spellings and the orthographic rules used (Bannard & Matthews, 2008). Half (16) of the place names were presented spelled correctly, the other half (16) were spelled incorrectly and in such a way that the letter combinations would not be usual in the English place names e.g. 'WESTMINTSER '. Capital letters were used to replicate the format required in British passports (International Civil Aviation Organisation, 2015).

Procedure

This experiment was conducted on line using the Qualtrics survey system. Participants were shown places names one at a time in a size 16 font in the centre of the screen and were required to use the mouse to select either 'Correct' or 'Incorrect' as shown in figure 3.2. Order of presentation was randomized throughout (i.e. not blocked by condition). The task was self-paced, with no time limit for responses, and typically took about 5 minutes to complete.

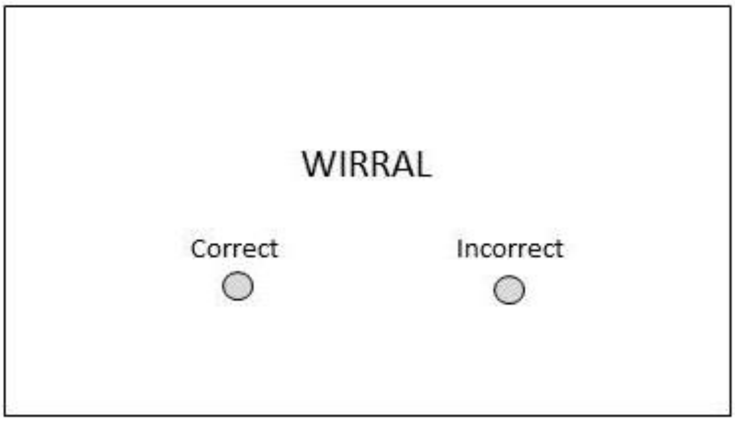

Figure 3.2. Stimuli from the place name checking task.

Results

Accuracy for correct place names (*M* = 97%, *SD* = 4.85) and incorrect place names (*M* =97%, *SD* = 4.87) was not significantly different $t(23) = .592$, $p > .05$. A full breakdown of accuracy by item is shown in table 3.1.

Table 3.1. Mean accuracy for each place name.

Correct Place Names	**Accuracy (%)**	**Incorrect Place Names**	**Accuracy (%)**
WIGAN	100%	EDINBORUG	100%
COVENTRY	100%	MILNOT KEENES	100%
DURHAM	100%	WARKICW	100%
KINGSTON UPON THAMES	100%	STAFDROF	100%
NORWICH	100%	SWASNEA	100%
BROMLEY	100%	LINLCON	100%
YORK	100%	SOLHULLI	100%
BRIGHTON	100%	ETERX	100%
LEICESTER	100%	LUUTON	100%
DUDLEY	96%	BOURNMEOUHT	100%
KENSINGTON AND CHELSEA	96%	CNATERBURY	100%
HARROGATE	96%	PLLYMOUTH	96%
TOWER HAMLETS	96%	OXFROD	96%
FALKIRK	96%	KINGSTON UOPN HULL	96%
WIRRAL	92%	SOUTHAMTPON	79%
GREENWICH	88%	WESTMINTSER	79%
Total	**97%**	**Total**	**97%**

Discussion

This experiment indicates that individuals perform very well when checking data in isolation. Correct place names may be misidentified as incorrect if the individual is not familiar with the name and / or the spelling. Incorrect names misidentified as correct were accounted for by 5 place names overall and SOUTHAMTPON and WESTMINTSER in particular. This would suggest that transposition errors were more difficult to identify than swapping the position of letters generally found further apart e.g. WARKICW or doubling a letter e.g. PLLYMOUTH.   This is supported by considerable evidence demonstrating that similarity to English words affects the ease with which they are rejected and remembered (for a review see Humphreys & Evett, 1985). The way in which non-words and English words differ is an entirely independent area of investigation that will not be addressed here. However, it is important to note that if these place names were to be used in future experiments particularly in terms of face matching, place names in both conditions should be counterbalanced across match type so that difficulty is distributed equally across same and different face pairs.

Participants exceeded the 85% accuracy documented by (Levy et al., 1986, 1992). This is not surprising since in their task participants were required to identify 12 non-word errors and 12 word errors in a text that was 350 words long – a much more onerous task than the task presented here. Other studies that have investigated identification rates for non-words shown in isolation have shown an accuracy rate of 95- 98% (Frost, Katz & Bentin, 1987; Forster & Chambers, 1973). This provides confidence that the non-words used in this study are sufficiently identifiable.

The aim of this experiment was to identify baseline performance for the identification of the orthographic errors used in Experiment 4 when the words are shown in isolation. The 29% accuracy achieved in Experiment 4 has been exceeded in even the least identified incorrect place name of ‘Westmintser’ that was identified on 79% of occasions. It should be noted that the prevalence of place name spelling errors in these two experiments was different. Experiment 4 had a prevalence of place name spelling errors of just over 16% whereas this experiment had a prevalence of 50%. Reducing the prevalence of targets (in this case incorrect biographical data) in visual search tasks, has been shown to reduce the accuracy of identifying targets (Wolfe, Horwitz & Kenner 2005; Wolfe, Horowitz, Van Wert, Kenner, Place & Kibbi, 2007). However, Wolfe et al., (2007) demonstrated that reducing the prevalence from 50% to 10% reduced accuracy for identifying errors from 93% to 84%. This would suggest that the drop in accuracy for place name spelling errors from 97% in Experiment 6 to 29% in Experiment 4 was not only due to a reduction in prevalence but the influence of additional factors such as multiple errors to be identified or the ID context. However, whether the data checking is compromised by the ID context or when the ID context is shown with face pairs is not clear and will be investigated further in the next experiment.

3.4 Experiment 7

Introduction

Experiments 5 and 6 provided baseline data checking accuracy for stimuli shown in isolation. Data checking accuracy for place name spelling errors when shown in isolation was 97%, more than twice the accuracy achieved for place name spelling errors when checking biographical data shown with a face pair. Age estimation for faces shown in isolation within +/- 20 years was 97.08 % and mean accuracy for gender estimation was 94.08 % both significantly higher than the accuracy achieved when checking biographical data shown with a face pair (23% and 77% respectively). These findings confirm that the data checking was compromised in Experiment 4 and that this was not caused by unusual or difficult stimuli. However, it is not clear whether the data checking was compromised by the ID context or by the ID context *and* the presence of a face pair. In this experiment, I aim to establish baseline data checking accuracy for identifying incorrect gender, incorrect year of birth (YOB) and incorrect spelling of place of birth (POB) when placed in a passport frame. This will allow performance in biographical data checking to be compared when shown with the ID context only and when shown with the ID context and the face pair as carried out in Experiment 4.

Method

Participants

Participants were 36 members (16 male) of an experimental volunteer community at the University of York consisting of students, staff and residents. All were over 18 and reported normal or corrected to normal vision (mean age = 24.83, range = 18-69) and did not have dyslexia. To ensure their data checking was not compromised by a lack of familiarity with place names, participants were also required to be British Citizens who spoke English as their native language. Participants were reimbursed a small fee for their time.

Materials

The 96 passport frames shown in Experiment 4 were used here. Each passport frame contained one of the 96 faces used in Experiment 4 and 5. The biographical data (all fictitious) from Experiment 4 and 6 was also used and in the combinations to be required to be valid or invalid with the associated face. A full set of the biographical information used can be found in Appendix 3. Key personal data could be rendered invalid as follows:

- wrong-gender forename and sex identifier – incorrect gender – 16 items;
- misspelled place of birth using orthographical errors e.g. Luuton – incorrect place of birth (POB) – 16 items;
- and a birth year more than 20 years older than the age of the target faces – incorrect year of birth (YOB) – 16 items.

In all cases, there was only one data error in each frame.

Procedure

This experiment was conducted on-line. Viewers saw 96 passport frames, one at a time and were asked for each ‘Is the data correct?’ and selected either ‘Correct’ or ‘Incorrect’ as in figure 3.3. Half the frames showed valid, and half invalid data. Each invalid data type was displayed in 16 passport frames. Faces were counterbalanced across the experiment such that each face occurred equally often in each condition. The task was self-paced, and typically took about 20 minutes to complete.

Figure 3.3. Stimuli from the data checking task.

Results

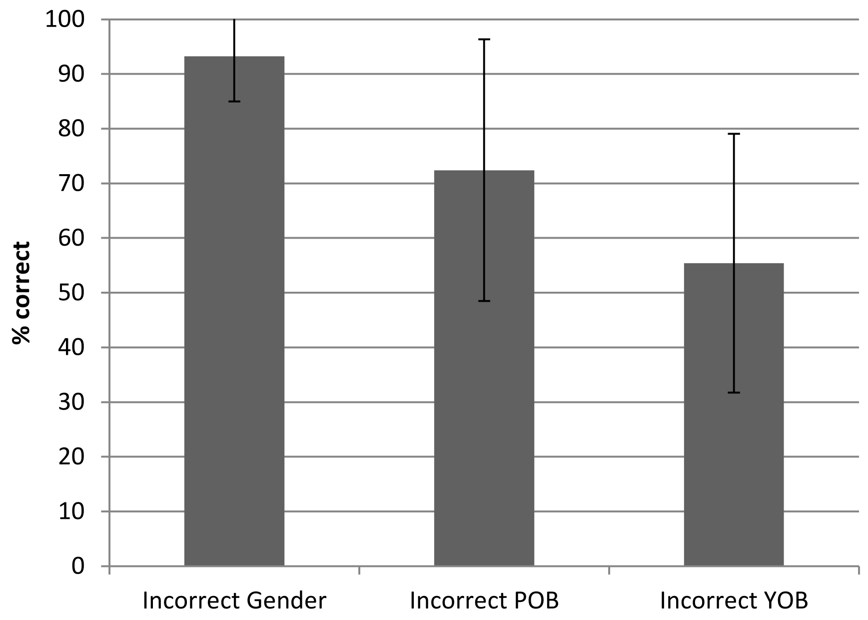

Figure 3.4. Mean data checking accuracy for invalid data types. Note error bars represent standard deviation.

Participants correctly classified 86% of the valid passports and 74% of the invalid passports, a significant effect of validity $t(35) = 4.22$, $p < .001$. Figure 3.4 shows mean accuracy for each of the three invalid data types. A single factor ANOVA (invalid data type: gender, POB, YOB) showed a main effect of data type $F(2,70) = .38.19$, $p < .001$, $\eta_\rho^2 = .52$. Tukey HSD revealed significant differences between all pairs of conditions. In sum, viewers were best at spotting incorrect gender, intermediate at spotting incorrect place of birth, and worst at spotting incorrect year of birth.

As in earlier experiments although this was not a speeded judgement task, I nevertheless recorded time taken to make these decisions. Time to verify a valid passport was significantly longer than time to spot an invalid one (means 9.7 sec vs 7.1 sec; *SD*s 5 and 2.6 respectively; $t(35) = 5.08$, $p < .001$. This is consistent with the fact that valid passports required exhaustive search of all three possible errors, while invalid passports could be classified correctly after any single dimension was identified as fraudulent. For the different types of fraudulent information, RTs followed accuracy data, with no sign of a speed/accuracy trade-off. Mean times to spot errors in gender, POB and YOB were 4.6 sec (*SD* = 2.2), 6.6 sec (*SD* = 2.6) and 10.2 sec (*SD* = 4.8) respectively $F(2,70) = 43.30$, $p < .001$, $\eta_\rho^2 = .55$. Tukey HSD tests once again showed reliable differences between all conditions.

Discussion

Overall, viewers' accuracy on checking these passport documents was low – our participants spotted only 73% of fraudulent documents. Although performance was low in this check of passport internal consistency, it was nevertheless higher than in Experiment 4, where a second face was present – suggesting once again that identity checks using cards with photos *and* other data may be even harder in the field than laboratory studies suggest.

The accuracy of data checking for different types of information was straightforward. Gender errors were spotted very accurately, place of birth errors less so, and year of birth errors were the least easily detected. The high levels of accuracy with gender judgements is consistent with previous research (e.g. Bruce et al., 1987) and the results achieved in Experiment 5 where participants categorised faces shown in isolation.

Detection of year of birth errors is perhaps poorer than one might expect from research and the high levels of accuracy achieved in Experiment 5 where participants estimated the age of faces shown in isolation. However, it should be noted that this is not an age matching task. Passports show year of birth and so computation of the bearer's age requires a calculation. The viewer must subtract the year of birth from the year of the passport's issue (in this case 2013) to identify the age provided for the face in the photo. This given age must then be compared with the estimated age derived by the viewer from the face in the photo. This makes checking the appropriate age of the photo a several-step process - and may account for the poor accuracy rate. Finally, the place of birth errors are interesting, because they do not require any comparison with the photo. Instead, simple checking for correct orthography of UK place names is shown to be rather error-prone and certainly compromised in the ID context.

The differences in accuracy for place of birth errors and age estimation may be due to the methodological differences between this study and those referenced, but also because of deleterious effects of carrying out multiple tasks simultaneously (Pashler, 1994; Jones et al., 1990). Gender categorisation of faces has been shown to be relatively robust to effects of dual task interference (Reddy, Wilken & Koch, 2004). However, it would appear that the same is not the case for checking place name errors. Accuracy for place of birth errors was significantly lower when the place names were checked in the ID context (72%) than when presented in isolation (97%) – this suggests that the combination of multiple data checking tasks may have resulted in reduced accuracy. This may be supported by the levels of accuracy (85%) achieved in Levy et al.,'s studies (1986, 1992) which suggest that error rates for detecting non-words are reduced when looking for multiple errors and multiple types of errors in a large piece of text. In contrast, the increased complexity of the YOB task in combination with carrying out multiple tasks may also account for the reduced accuracy for identification of incorrect years of birth. The significantly lower accuracy rate and slowest decision time indicates that this is a significantly different and more difficult task than simply estimating date of birth from an image.

This study has provided a useful baseline for data checking using passport frames without the presence of an additional same or different face. The levels of accuracy achieved can now be compared with participants carrying out the same data checks while viewing the compound stimuli required for the post data check face matching task. Experiment 4 provided data checking accuracy for a subset of the stimuli used here when the passport frames were presented with face pairs. At all levels data checking accuracy was affected by the nature of the face pairs with valid data more likely to be identified as incorrect if presented with a different face pair and invalid data more likely to be identified as correct if presented with a same face pair. However, it was unclear whether the overall levels of data checking accuracy were affected by the presence of the face pairs. A comparison of the data checking accuracy found in Experiment 4 and this experiment indicates this was the case. Here participants correctly classified 86% of the valid passports and 74% of the invalid passports. When shown with face pairs, these levels fell to 84% and 42% respectively. The dramatic drop in invalid data checking accuracy was due to accuracy being compromised across all error types. Here, participants correctly identified 93% of gender errors, 72% of POB errors and 55% of YOB errors. When shown with face pairs, these levels fell to 78%, 25% and 23% respectively. The changes in decision times across experiments also reflect a change in behaviour. Here, time taken to verify a valid passport (9.7 seconds) was significantly longer than time to spot an invalid passport (7.1 seconds). This is consistent with carrying out an exhaustive search of the valid passports to eliminate all three possible errors. However, in experiment 4 this strategy appeared to change with invalid passports taking as long as valid passports to check but with no accuracy trade off. This supports the increased level of difficulty of checking invalid data when face pairs are present demonstrated by the fall in accuracy levels.

Taken together these findings demonstrate that data checking is not only compromised in the ID context but is also compromised, and to an even greater extent, by the presence of a face pair.

3.5 General Discussion

The results presented here and in chapter 2 show that data checking is affected by both the ID context *and* the presence of a face pair.

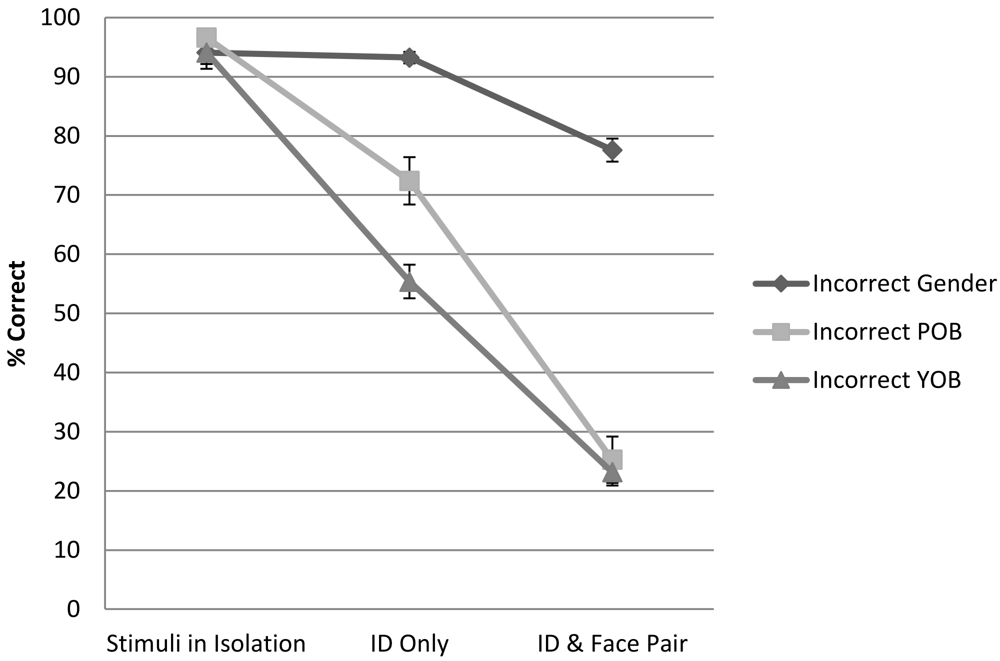

Figure 3.5. Mean data checking accuracy across error types when biographical data is presented in isolation, in the ID context and in the ID context with a face pair. Note error bars represent standard error.

It is important to note that this occurred despite the face pair being task irrelevant, and even when participants were asked to refer only to the face in the passport frame when checking the data. Figure 3.5 illustrates that levels of accuracy for the different types of biographical data are compromised by the ID context and the presence of a face pair. Three potentially influencing factors were identified: the deleterious effects of carrying out multiple tasks; the increasing complexity of the age checking task; and the mandatory processing of faces distracting attention from the data checking task.

By its very nature, the ID context requires multiple data checks to be carried out concurrently. The biographical data page of a passport contains six biographical data elements. To establish that the biographical data is correct all biographical data elements must be checked resulting in competition for resources between the elements. However, this would not account for the much more significant drop in accuracy for YOB and POB when shown in the ID context. The increasing complexity of the age checking task would account for the further reduction in accuracy for YOB errors. However, this would not account for the drop in accuracy for POB errors.

Levy et al., (1986, 1992) demonstrated that checking for a larger number of non-word errors in a larger piece of text saw accuracy levels of 85%. Accuracy for POB errors dropped in the ID context to 72%. Like the gender categorisation task, the POB task was relatively similar to the task carried out in isolation. However, accuracy for the gender categorisation task remained highly accurate in the ID context. The only notable difference between the two tasks was that the gender check required referencing the face in the passport frame and the POB error check did not. Given that the traditional view that the shallower and less meaningful of two tasks is affected by dual-task interference (Jones et al.,1990) and that the presence of a face may lead to automatic processing (Farah et al., 1995a) then even in the context of the ID there could be competition for resources to process face and biographical-data stimuli. This would lead to a greater proportion of resources being available for the gender checking task and less so for a text specific task like proofreading places of birth.

That faces retain attention during perceptual tasks, even when the faces are task irrelevant (e.g. Jenkins et al., 2003; Lavie et al., 2003), may also explain why word level errors appear particularly susceptible to interference in this task. Jones et al., (1990) found that detection of spelling errors could be disrupted by irrelevant speech but only when the speech was meaningful. If faces are processed automatically (Farah et al.,1995a) at both a visual and semantic level (Boehm et al., 2006; Burton et al., 1998) this might explain why there was the largest drop in POB errors. The mechanisms underlying word recognition are the subject of much debate (for a full review see Carreiras, Armstrong, Perea & Frost, 2014). However, the importance of visual processing, if only at the very earliest level of processing, remains constant across many models. If the faces absorb attentional resources, then there is little attentional and / or visual resource left to process non-words and particularly those that do not relate to the faces.

The reduction of resources available for the biographical data checking task when face pairs are present was proposed in Chapter 2. The findings in this chapter provide further evidence and extend this proposal. Here the presence of a single face reduces accuracy in biographical data checking tasks. This makes it even more likely that it is the presence of the face pair, as in Experiment 4, that results in even fewer attentional and visual resources available for the data checking task and a further reduction in accuracy overall.

This chapter not only provides the baseline measurements required to identify the interactions between faces and biographical data detailed here and in Chapter 2, it also demonstrates the need to identify and breakdown tasks into their underlying parts. It is the complex nature of these tasks and their interactions, that makes checking photo ID a much more difficult task than previously understood. Chapter 2 demonstrated that unfamiliar face matching is undoubtedly a more difficult task when carried out within the ID context. This chapter demonstrates that checking biographical data is also much more difficult when carried out in the ID context - not only in the presence of a face pair but when one face is shown in the passport frame alone.

Chapter 4 – Predictors of Face Matching

4.1 Introduction

Chapters 2 and 3 of this thesis focused particularly on unfamiliar face matching in the applied context. Examining the unfamiliar face matching task and the biographical data checking task carried out by individuals in a passport checking role allowed the nature of those tasks to become clear. Unfamiliar face matching was more difficult in the context of a passport frame. This makes the need to have individuals in these roles whose unfamiliar face matching skills are very good even greater. However, as discussed in Chapter 3, those individuals we might expect to be better at the task, including passport officers, are typically no more accurate than the general population (White et al., 2014a; Burton et al., 1999a; Kemp et al., 1997).

The range of performance on the Glasgow Face Matching Task (GFMT) is large, from around 55 – 100% (Burton et al., 2010). This is amongst experts and general population alike (ranges 58 – 96% and 51 – 100% respectively; White et al., 2014a; Burton et al., 2010). This indicates that some individuals are better than others at unfamiliar face matching. Research investigating the potential for training to improve ability has shown that improving performance across unfamiliar faces is difficult. Feedback and working in pairs creates improvement but generally around only 10% (White et al., 2014c; Dowsett & Burton, 2015). It therefore seems that improvements in applied settings could most easily be made by recruiting high performers directly into those roles that demand face processing skills. This may well be why there is also a focus on investigating high performers in face processing tasks, or 'super-recognisers'.

Members of the Metropolitan police who have been identified as 'super recognisers' using in house tasks, perform much better than controls in the GFMT and familiar face matching tasks (Davis et al., (in press); Robertson et al., 2016). Individuals believing they had superior face processing skills and tested by Russell et al., (2009) were shown to perform better than controls in both face recognition and face perception tasks. More recent research has also shown that other self-described super recognisers have superior face recognition and face matching skills across a range of tasks and media (Bobak et al., 2016a; Bobak et al., 2016b).

These findings demonstrate that high performers can be found. However, most investigations of individual differences in face perception are tightly focussed on specific tasks. On particular tasks, commonly those involving unfamiliar face memory, there appear to be large individual differences. When examining the individual performance of super recognisers the associations between different face processing tasks are not as reliable as those found at the group level. Davis et al., (in press) note that one super recogniser (1 out of 10) performed very poorly on the GFMT and Bobak et al., (2016c) found that less than half of the super recognisers (3 out of 6) in their study were significantly better than controls on the GFMT. Bobak et al., (2016c) also investigated to what extent super recognisers' skills were specific to faces through a range of tasks, including hand matching, global processing and memory for cars. Half of the super recognisers showed superior skills only when processing faces. Of the other three, each showed enhanced skills in the hand matching task, the memory for cars task and the global processing tasks. The distributed nature of these wider skills across separate individuals does not provide a great deal of confidence in these findings. The small numbers of individuals involved in each case may indicate relationships that are observed by chance rather than an association that might be observed across the general population. More convincing are the general findings that face processing skills are specific to faces and that face processing tasks are associated.

Examining face processing ability in high performers may also prove to be as complex as findings in low performers. Much of what we know about face processing has previously been identified from the ways in which individuals with an inability to process faces well (prosopagnosia) complete face processing tasks. Studies with prosopagnosic participants have demonstrated that performance in familiar and unfamiliar face tasks is dissociated, that configural processing is associated with face recognition and perhaps most fundamentally that face recognition is dissociated from object recognition (Malone, Morris, Kay & Levin, 1982; Behrmann, Avidan, Marotta & Kimchi, 2005; Farah et al., 1995b). These findings in themselves are difficult to interpret. Farah et al., (1995b) suggest that face processing is a specific process unrelated to object processing, however, Behrmann et al., (2005) suggest that face processing is related to more general processes for example, global processing. Malone et al., (1982) suggest that face processes themselves are unrelated. This lack of consensus has formed the basis for much research including the development of many models of particular types of face recognition e.g. Bruce and Young (1986) and investigations into the specificity of face processing (for a review see Kanwisher, 2000).

Further review of studies with individuals with prosopagnosia may provide insight into these apparent contradictions. In the case of configural processing, Behrmann et al., (2005) found that global processing was associated with performance in both a face recognition and face matching task for individuals with congenital or developmental prosopagnosia. Yovel and Duchaine (2006) however, found different results when testing developmental prosopagnosics who could identify configural and featural changes in houses (the door is further from the window, the door is different) but were impaired in identifying configural and featural changes in faces.

The labels "global processing", or the tendency to decompose a scene rather than build it up from its constituents (Navon, 1977), and "configural processing", or the perception of relations between features or parts of a scene (Maurer et al., 2002), are terms that have often been used interchangeably in research (Sergent, 1984; Young, Hellawell & Hay, 1987). If global processing and configural processing are the same process, then the results above are contradictory. However, global processing and configural processing are not interchangeable. Maurer et al., (2002) in their review of configural processing, noted that configural processing is used to describe a range of processes, which in themselves are separable. Configural processing has been used to refer to any process that uses the spatial relations within the face. This is in contrast to using single features e.g. eyes or nose.

Featural processing uses the structural features of the face e.g. eyes, nose or mouth rather than the pictorial features e.g. shadows or marks on the image (Bruce, 1982). Leder & Bruce (2000) draw on earlier work carried out with object recognition to bring further clarity to the definition of facial features. Each feature is described as an independent constituent element differing from others in dimensions such as texture or shape. Bruce & Young (2012) also point out that like objects, the features of faces are generally those areas that can be named, usually as a result of performing a function e.g. teeth. In addition, there may also be other features that are used by viewers to recognise individuals that do not serve a purpose but have names e.g. moles. In contrast configural processing is concerned with the spatial relationships between these features. These relationships typically do not have names and may be more related to global processing than features which O'Donnell & Bruce (2001) term 'local'.

However, Maurer et al., suggest there are three types of configural processes: 1) First order relations whereby we identify faces because of their consistent arrangement, the eyes above the nose and the nose above the mouth; 2) Second order relations, or the distances between the features; and 3) Holistic processing, or perceiving the face as a whole or gestalt. If these processes are separable this might explain how different manifestations of prosopagnosia affect a particular type of configural processing, for example second order relations, but not another, e.g. holistic or gestalt processing, a type of processing perhaps more associated with global processing.

The apparently contradictory results may also be explained by research with individuals with prosopagnosia. Barton (2009) found that individuals with acquired prosopagnosia had less difficulty with global processing but were impaired in their ability to judge spatial configurations. These findings support Maurer et al's (2002) separation of the different processes underlying configural processing, but also suggest that individuals with acquired prosopagnosia may behave differently than individuals with developmental prosopagnosia. The studies cited earlier suggest individuals with developmental prosopagnosia have difficulties with global processing, whereas individuals with acquired prosopagnosia have difficulties with second order relations (Behrmann et al., 2005; Yovel & Duchaine 2006). Damasio, Tranel and Damasio (1990) note in their review of prosopagnosia, face processing is complex and as a result, failures can occur at many different stages. Prosopagnosia then is not likely to be a single disorder but a family of related but separable disorders. Even within one form of prosopagnosia, there may be many different ways in which face processing may fail. Duchaine and Nakayama (2006b) advise caution in testing groups of individuals with developmental prosopagnosia as their work using FMRI and MEG studies has shown that they are a very heterogeneous group. Therefore, findings from individuals with prosopagnosia may be very particular to that individual and not easily or beneficially translated as general findings.

These studies highlight the importance of task validity, the risks of assuming constructs such as face processing and configural processing as singular, and the risks of using participants from perhaps either end of the range of performance. Assuming homogeneity within either super recognisers or individuals with prosopagnosia could result in quite specific findings being inaccurately applied to the much larger group of individuals in the middle of the range. Not only this, but they could also lead to assumptions about not only the task but the mechanisms being investigated via the task. It may be far easier to assume that performance in a face recognition task will relate to performance in a face matching task and therefore face processing skills are highly related if this has been found in five high achieving individuals. Nevertheless, if this performance is not shown to covary across high, low and middle ranging performers we cannot have confidence in the relationship.

There is face processing research that has taken an individual differences approach, with wider ranges of performance. This research provides findings that can be tested and built upon in order to investigate the questions that have already been raised. Firstly, to what extent are face processing tasks related? Secondly, to what extent is face processing specific - using mechanisms that are uniquely developed or specialised? These questions are the focus of this chapter and Experiment 8, a battery designed to measure performance across a range of face processing and wider perceptual, cognitive and personality tasks. The results confirm that generally face processing tasks are well related, and do not relate to wider perceptual, cognitive or personality measures. However, unfamiliar face matching is associated with object matching, and to a lesser degree, local processing and space perception. The battery also demonstrates that accuracy for same and different face matching is dissociated and leads to differences in the way matching same and different face pairs relates to other face processing and more general tasks.

4.2 Experiment 8

Introduction

Face matching research began as an investigation of matching identities across varying viewing conditions e.g. viewpoint or expression without the burden of memory load (Bruce et al., 1999). The expectation was that participants would have very few issues identifying identities in arrays of images shown simultaneously. However, even when all images were presented in a neutral, full face pose, participants only identified the correct individual in an array 80% of the time. Unfamiliar face matching was identified as difficult in its own right and potentially offered an opportunity to investigate the difficulties of unfamiliar face perception. These findings were not isolated. Earlier research had shown that even when matching ID with live individuals one to one, checkout operators failed to identify customers carrying images of someone else 36% of the time (Kemp et al.,1997). More recent research has demonstrated that participants asked to match the internal features (eyes, nose and mouth) of unfamiliar faces only achieved a mean accuracy of 80% (Clutterbuck & Johnston, 2002). However, when Clutterbuck & Johnston (2002) asked participants to match the internal features of highly familiar individuals, participants achieved a mean accuracy of 91.5%. Their study showed that, like face recognition, face matching has a familiarity advantage and, like face recognition, this advantage is lost when participants view external features (hair, face shape, chin) only (Ellis, Shepherd & Davies, 1979). These findings suggest that there is a great deal of commonality between face recognition and face matching processes. If this is the case, i.e. if face matching and face recognition share underlying face processing mechanisms, can we assume that individuals who excel in face recognition will also excel in face matching? Equally, if abilities that are more general are found to be associated with face recognition for example, can we assume they will also be associated with face matching?

Studies that demonstrate accuracy decreases for both face recognition and face matching tasks when faces are less familiar or external features are viewed do not necessarily provide any information regarding the relationships between the tasks. Yovel, Wilmer and Duchaine's (2014) recent review of individual differences research in face processing indicates that an individual differences approach provides a highly useful way to investigate relationships in face processing further. They demonstrate that the capacity to review association and dissociations across individuals' performance clarifies the nature of relationships, particularly in terms of the degree to which mechanisms underlying face processing may be specific to faces or used for a range of processes or stimuli. A useful example provided is Konar, Bennett and Sekuler (2010) who showed that holistic processing as a general mechanism, or processing stimuli as a whole, was not correlated with an unfamiliar face matching task. These results provide a direct challenge to the widely held, but largely untested, assumption that holistic processing is a primary mechanism underlying face identification. These findings were also later extended by Wang, Li, Fang, Tian and Liu (2012) who showed that neither holistic nor global processing (prioritising the overview rather than its parts) was associated with unfamiliar face recognition. This suggests that unfamiliar face matching and unfamiliar face recognition share face processing mechanisms that are specific to face processing.

Using an individual differences approach and a battery of face processing and more general tasks, Megreya and Burton (2006) demonstrated that face matching was potentially a very important area of investigation in its own right. They found that upright unfamiliar face matching accuracy was positively associated with inverted unfamiliar face matching. This was surprising since performance on upright face recognition tasks is generally found to have either no association or a negative association with performance on inverted face recognition tasks (see Valentine, 1988 for a review). This negative or lack of association between upright and inverted face recognition has been taken as a marker of the highly specific nature of face processing. Any contradiction of this relationship suggests that unfamiliar face matching is not using the same set of processes used in other face processing tasks. Megreya and Burton (2006) also found that performance on unfamiliar face matching tasks was positively associated with an unfamiliar face recognition task when the images were upright. However, performance on unfamiliar face matching tasks was only positively associated with performance on familiarised face matching tasks and familiar face recognition tasks when the faces in the familiar face matching tasks were inverted. These findings were so compelling that Megreya and Burton questioned whether unfamiliar faces were faces at all but rather objects. The unique nature of unfamiliar face matching was extended even further in work that identified that even performance in same and different unfamiliar face matching was dissociated (Megreya and Burton, 2007), suggesting again that face matching was very different from other face processing tasks. This research suggests that face processing tasks do not share the same level of commonality that the findings of group means based research propose (Ellis et al., 1979; Clutterbuck & Johnston, 2002) and that individuals who might excel at unfamiliar face matching cannot be predicted from their performance in other face processing tasks.

When comparing unfamiliar face matching performance with ability on more general measures, Megreya and Burton (2006) found more positive relationships. Megreya and Burton found a positive association between unfamiliar face matching and perceptual speed. They also found positive associations between unfamiliar face matching accuracy with a visual short-term memory task and the Matching Familiar Figures Task (MFFT) (Kagan, 1965). Though designed to measure impulsivity and accuracy the MFFT required participants to match a target image of an object with one of six variants of the image in an array. Since the visual short-term memory task also required memory for objects, the association between these tasks and the unfamiliar face matching task suggests that unfamiliar face matching makes use of wider cognitive processes. These findings also suggest that performance on more general tasks might be a more useful indicator of face matching ability than face processing tasks.

Megreya and Burton’s (2006) findings suggest that unfamiliar face matching is not associated with other face processing tasks and makes use of more general processes. However, their findings regarding the dissociation between familiar and unfamiliar face processing have not been replicated. Research with high performers in face identity processing tasks, or ‘super recognisers,’ found several positive associations between face processing tasks (Russell et al., 2009; Davis et al., in press, Robertson, et al., 2016; Bobak et al., 2016a; Bobak et al., 2016b). In the first study to examine super recognisers' performance, (Russell et al., 2009) participants completed a familiar face recognition task – the Before They Were Famous (BTWF) Test. In this test participants are shown the faces of celebrities as children and are asked to identify them by either name or another unique identifier e.g. film/context or character played. Since the participants never knew or saw the celebrities as children they may have to mentally transform the face seen in order to compare it with the face as known or stored as a face recognition unit. This not only makes the task more difficult, reducing the ceiling effects often found in familiar face recognition tasks, but may also test participants' ability to mentally transform images.

Participants then completed an unfamiliar face recognition task, the Cambridge Face Memory Test (CFMT: Duchaine & Nakayama, 2006a) in which participants were familiarized with 6 identities and asked to identify them from arrays of three, initially with the same pose and lighting, then with altered pose and / or lighting and finally when partially obscured by noise. Performance on the CFMT and BTWF was highly positively correlated. In the second test participants were asked to complete the CFMT and also the Cambridge Face Perception Test (CFPT) (Duchaine, Germine & Nakayama, 2007) in which participants are asked to sort 6 morphs of faces from most like a target image to least like a target image for upright and inverted faces. Performance on the CFMT was significantly and positively associated with the upright version of the CFPT but not with the inverted CFPT. This supports Yin's (1969) findings that faces are processed significantly differently when inverted, but also that processes underlying familiar and unfamiliar face recognition and perception tasks are related.

Robertson et al., (2016) tested members of the Metropolitan Police Super recogniser pool on a range of face matching tasks: the GFMT; a more difficult unfamiliar one to one face matching task using male models; and a familiar one to one face matching task where faces of celebrities and look a likes were pixelated to reduce ceiling effects. The police super recognisers performed well above controls across the full range of tasks suggesting that performance across familiar and unfamiliar face tasks and face matching and face recognition is associated. However, it is important to note that at an individual level, work with super recognisers has also queried the relationship between unfamiliar face matching and unfamiliar face recognition (Davis et al., in press; Bobak et al., 2016c) which seems relatively secure in Megreya and Burton's (2006) work.

Together these papers suggest that face processing tasks are not consistently associated when more general populations and highly specific populations are compared. We cannot assume that an individual who excels at familiar face recognition would necessarily excel at unfamiliar face matching. Given that unfamiliar face matching for same and different pairs is dissociated, and performance on same and different face matching has not been compared with ability on other face processing tasks, we cannot assume that individuals who excel at familiar face recognition for example would also excel on either same or different unfamiliar face matching tasks.

The relationships identified between unfamiliar face matching and more general tasks may provide further insight into the relationships between face processing tasks. Megreya and Burton's (2006) findings that unfamiliar face matching was associated with perceptual speed and dissociated from field dependence, or the reliance on an object's external context (Witkin & Goodenough, 1977), built on relationships also found with face recognition tasks (Rose & Feldman, 1995; Rose, Feldman & Jankowski, 2003, Schretlen, Pearlson, Anthony & Yates, 2001). As discussed earlier, Konar et al., (2010) and Wang et al., (2012) showed that holistic processing was not correlated with either an unfamiliar or familiar face matching task. In contrast, individual differences research into unfamiliar face matching has also shown that, like unfamiliar face recognition, it is negatively associated with anxiety and neuroticism (Megreya & Bindemann, 2013, Mueller, Bailis & Goldstein, 1979, Nowicki, Winograd and Millard, 1979). These findings suggest that face processing tasks may share associations and dissociations with more general measures. This not only raises the possibility that face processing tasks may be more related than Megreya and Burton (2006) identified but also that more general measures that have been associated with face recognition could form the basis of a battery testing relationships between these measures and a range of processing tasks including unfamiliar face matching.

There is a large body of research comparing performance on a range of cognitive tasks, measures of personality and measures of perception with unfamiliar face recognition ability. Initially the focus was memory, identifying that individuals who were good at remembering unfamiliar faces, were also good at remembering paintings and words (Woodhead & Baddeley, 1981). Personality traits in addition to neuroticism and anxiety have also been associated with unfamiliar face recognition skills. In a new/old task, individuals high in extraversion were found to have an advantage in recognizing unfamiliar faces when compared to recognizing unfamiliar flowers (Li, Tian, Fang, Xu, Li & Liu, 2010). Empathy has also been shown to be related to unfamiliar face recognition accuracy. When a group of participants identified as being low and high in empathy completed an unfamiliar face recognition task, individuals high in empathy were significantly superior in their unfamiliar face recognition accuracy (Bate, Parris, Haslam & Kay, 2010). However, research investigating the relationship between perceptual style, in particular field dependence, and face recognition is less clear. Field dependence has been found to be associated with higher performance on unfamiliar face recognition tasks (Messick & Damarin, 1977) lower performance (Hoffman & Kagan, 1977) and to have no relationship with unfamiliar face recognition ability (Courtois & Mueller, 1982).

Not all behavioural research has found that face recognition is associated with more general measures. Twin studies using an individual differences approach suggest that the mechanisms underlying face recognition are very specific to face processing alone. Wilmer, Germine, Chabris, Chatterjee, Williams, et al., (2010) using the CFMT, demonstrated that unfamiliar face recognition was both highly heritable and dissociated from more general measures of *g* and object recognition. Also using the CFMT, Shakeshaft & Plomin (2015) identified that the heritability of unfamiliar face recognition accounted for over half (61%) of the observed individual variability and that most of this influence was unique and not shared with other cognitive abilities. Behavioural research with random samples of participants has also indicated that face processing is highly specific. Hildebrandt, Wilhelm, Schmiedek, Herzmann & Sommer, (2011) investigated a range of face memory and face perception tasks and found a dissociation between face cognition and general cognition e.g. working memory and mental speed, that remains even into old age

The lack of consensus found in behavioural research is also found in cognitive neuroscience. Face recognition is proposed to use processing regions within the brain such as the fusiform face area (Kanwisher et al., 1997) that are not used in object recognition. As Yovel et al., (2014) detail, an individual differences approach has shown that face selectivity in the fusiform face area (FFA) is positively associated with higher unfamiliar face recognition ability (Huang, Song, Li, Zhen, Yang & Liu, 2014). Greater activation of the right fusiform face area has also been shown to predict the behavioural advantage for the perception of familiar faces (Weibert & Andrews, 2015). In contrast, Gauthier, Curby, Skudlarski & Epstein (2005) found that an individual differences approach confirmed earlier research showing that activity in the FFA was positively associated with expertise with non-face objects. Event-related brain potential (ERP) studies have shown that by controlling for established and more general cognitive abilities, access to structural representations in memory or the early repetition effect (ERE) amplitude may not be an indicator of face-specific cognition as previously thought. This study shows that access to these structural memories may be an indicator of more general abilities (Kaltwasser, Hildebrandt, Recio, Wilhelm & Sommer, 2014).

The lack of clarity in both behavioural research and cognitive neuroscience confirms that there is much more work to be done in investigating the nature of face processing. Reviewing the approach traditionally taken by face researchers may provide further insight into why findings can appear so contradictory. Face research has typically sought to answer whether face processing uses specific mechanisms. This has resulted in tasks traditionally used to measure face processing being applied to object processing. Far fewer researchers have taken mechanisms used in broader capacities and investigated their application to face processing – seeking to answer whether other tasks use mechanisms underlying face processing. This has allowed greater opportunity for face specific mechanisms to be identified that do not translate well generally, such as the fine discrimination for distances that appears to apply only to faces (Yovel & Kanwisher, 2008). However, this has also led to fewer general skills being investigated and identified as associated with, or predictors of, high performance in face processing.

Though fewer in number, there have been some useful developments in other fields of research that use face recognition to inform factorial structures or validity of tests. A review of tests of executive function revealed that the unfamiliar face recognition section of the Recognition Memory Test (Warrington, 1984) and the Visual Object and Space Perception (VOSP; Warrington & James, 1991) position discrimination task both loaded onto an Inhibition factor (Burgess, Alderman, Evans, Emslie & Wilson, 1998). These results suggest that face recognition performance may be associated with both inhibition and space perception. A study assessing the factorial structure of perceptual style (Milne & Szczerbinski, 2009) may provide further insight into the lack of consistency in associations between field dependence and unfamiliar face recognition. Milne and Szczerbinski (2009) identified that a range of tasks were commonly used to identify a local rather than global processing style – the tendency to build up a scene rather than decompose a scene (Navon, 1977). Relations had been assumed between field independence, weak central coherence (the inability to experience the whole without attention to every detail (Kanner, 1943)), closure flexibility (detecting and dis-embedding a known stimulus array from a more complex array (Carroll, 1993)) and local processing with little attempt to identify the relationships between these constructs. Factor and task analysis identified that weak central coherence or field independence did not predict reduced global perception and that perceptual style was not a unitary construct. Both weak central coherence and field independence were strongly associated with the ability to detect and dis-embed a target from a more complex display in a Gestalt completion task such as the Mooney Face Test (Mooney, 1957) and measures of perceptual speed and local bias. Dis-embedding was also associated with another factor – Cognitive Flexibility, or the ability to allocate attention resources optimally, which was measured through performance on multiple unrelated tasks: Navon Local Accuracy, Visual Object and Space Perception (VOSP Battery) Silhouettes task (Warrington & James, 1991) and a Gestalt Completion Task. These findings suggest that perceptual style and field (in)dependence are separate constructs and require separate tasks for measuring their association with face processing.

Overall, the individual differences approach to face recognition research provides a mixed picture, identifying face recognition as both highly heritable and specific but also associated with memory, personality and wider cognitive tasks. Often, these findings have been thought to demonstrate relationships between a general face processing mechanism and another more general mechanism e.g. memory. However, as Megreya and Burton’s (2006) findings point out, and studies with prosopagnosic participants have demonstrated (Malone et al., 1982), we cannot assume relationships exist between face processing tasks. Individuals who perform well in face recognition tasks may or may not perform well in face matching tasks. More general tasks that predict performance in face recognition may or may not predict performance in face matching tasks, particularly when performance in both same and different face pair matching must be considered. The consequences of poor performance in both types of face pair matching are important, particularly in an applied context. Failing to identify an individual is the same as the person shown in their passport can result in their being detained at border control without grounds and delays for all passengers. Failing to identify that an individual is different from the person shown in their passport can result in individuals with criminal and terrorist histories entering countries illegally.

The research reviewed here has demonstrated that relationships between face processing and memory, perception and personality may be particularly fruitful areas of investigation particularly when efforts are made to separate constructs into their separable parts. However, it is important that no assumptions be made about the associations between face processing tasks, as the evidence supports both associations and dissociations. Different types of face processing should be investigated separately, even to the level of processes that have been shown to be dissociated e.g. same and different unfamiliar face matching. So too should the tasks measuring wider abilities e.g. perceptual tasks split down to perceptual style (global/local), space perception and disembedding.

The following study uses a battery of face processing and wider perceptual, cognitive and personality measures based on those used previously in the literature and adopting the approach described above. Face processing is split into four separate tasks; unfamiliar face matching, unfamiliar face recognition, familiar face recognition and face detection. The more general tasks include measures of memory, perceptual style, space perception, disembedding, inhibition, object matching, letter detection, personality and empathy. To ensure a wide range of face processing abilities, a volunteer sample of participants was recruited from an online community of individuals who are paid to take part in surveys. The purpose of this battery is twofold; first, to investigate the relationships between different face processing tasks and second, to investigate the relationships between face processing and wider perceptual, cognitive and personality measures.

Method

Participants

Participants were 107 members of the Qualtrics panel (51 male), who all reported normal or corrected to normal vision (mean age = 53.45, range = 18 - 85). Participants were reimbursed a small fee for their time.

Materials

All participants completed four face processing tasks. Example stimuli from each of these tasks are shown in figure 4.1. Each task is described in detail below.

The Before They Were Famous Test (BTWF) - task – adapted from (Russell et al., 2009)

The BTWF Test is a familiar face recognition task. The task consists of 40 photos of celebrities as children or adolescents (recognition test) and 40 corresponding photos of the celebrities as adults (familiarity test). Each photo was cropped to capture the celebrities' face, all of the images were standardised to a height of 400 pixels and were shown in grayscale and colour, depending on the age of the photograph. Participants were presented with each of the before-fame celebrity photos and, in line with the instructions used in Russell et al.,, were required to indicate the celebrity's name (e.g. Daniel Craig) or, alternatively, provide a unique identifying description (e.g. the actor who plays the current

James Bond). In contrast to the BTWF Test used in Russell et al., the present task included a familiarity test in which participants were also presented with adult (after-fame) photos of each of the celebrities. In the Russell et al., version of the test, a high score could simply indicate that the participant knew a wider range of celebrities, rather than providing an indication of familiar face recognition ability. Therefore, in the present task, face recognition performance is calculated as the number of celebrities recognised before fame, divided by the total number of celebrities known (x 100 to provide percent accuracy) (see Bindemann, Attard & Johnston, 2014). The task was self-paced and typically took about 15 minutes to complete.

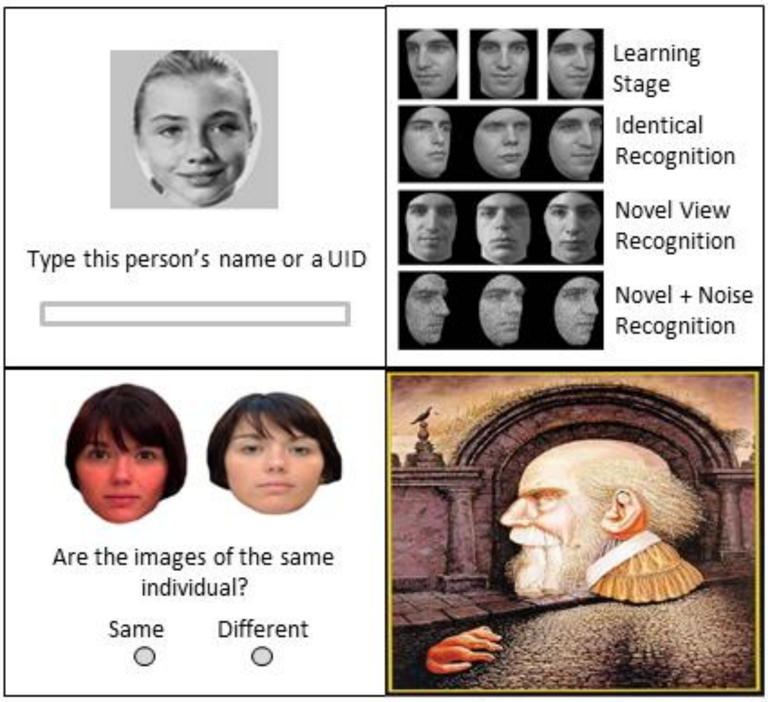

Figure 4.1. Example stimuli from (clockwise from top left) the BTWF test, the CFMT, the GFMT and the Face Detection task.

Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006a)

The CFMT (Duchaine & Nakayama, 2006a) is a 72-item face recognition-memory task that is split into three sections. In section one, participants are told to learn a target face shown in three orientations (left facing, forward facing, right facing). Participants are then presented with a three-alternative forced choice task in which they have to pick out the identical face image. This process is repeated for each of six target faces. In section two, the three-AFC test is retained, with participants now having to identify novel instances of each target face. Section three is identical to section two, with the exception that the test images have had visual noise added to them in order to make the task more challenging. Presentation of the stimuli for memorising was timed; presentation of the target stimuli was not. There was no time limit for responses and the task typically took about 15 minutes to complete.

Face Detection

This was an in house task developed to test participants' ability to detect faces in a scene. Participants were given 1 minute to find as many faces in possible in one image. Each image contained faces embedded into the scenery as well as belonging to individuals in the scenes. Participants were given one practice image and two test images. Answers were banded rather than using the raw scores. The optimum score of 4 was given for the correct answer – 15 faces in each image. Answers of 1-5 faces were given a score of 1, 6-10 faces were given 2, and 6-10 or 11-14 faces were given a score of 3. Accurately identifying the number of faces was a very difficult task and banding allowed participant's ability to be measured more accurately and sensitively since a score of 12 was much closer to the correct answer than 3. Scoring guesses above the target number was only carried out for a small number of numbers, the aim being to reduce the ability of individuals randomly entering high numbers to achieve high scores. The task was timed, participants were given 1 minute to view each image and complete the task.

Glasgow Face Matching Test (GFMT - short form) (Burton et al., 2010).

The GFMT is an unfamiliar face matching task. The GFMT (short version) consists of 40 pairs of unfamiliar faces, half of which are same face pairs and half of which are different face pairs. Each face image is front facing in pose, neutral in expression, shown in colour and standardised to a width of 151 pixels. In order to ensure that the GFMT would provide a non-trivial matching task, the photos within each pair were taken a few minutes apart using different cameras (for more details see Burton et al., 2010). Participants were asked 'Are the images of the same individual?' and selected 'Same' or 'Different'. The task was self-paced, and typically took about 5 minutes to complete.

All participants also completed 11 measures of cognition, perception and personality. Examples of each of these trials are shown in figure 4.2.

1) *Behavioural Assessment of the Dysexecutive Syndrome (BADS) Card Sorting Task (Wilson, Evans, Alderman, Burgess & Emslie, 1997).*

This is a rule shifting task measuring cognitive flexibility and inhibition. Participants view 20 images of playing cards individually and in part one have to answer

‘Yes’ or ‘No’ to question ‘Is the card red?’ In part two the rule is changed and participants see the cards again but have to adapt their responses inhibiting their original response to answer ‘Yes’ or ‘No’ to the question ‘Is the card the same colour as the previous card?’. The responses are timed, but this is a self-paced task and typically took about 5 minutes to complete.

2) Big Five Inventory (BFI) (John, Donahue & Kentle, 1991)

This is a 44-item questionnaire measuring five independent personality factors – Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. The questionnaire is self-paced and each question consists of a short phrase such as ‘I am always prepared’. Participants must state how accurately this reflects their experience by selecting the most appropriate response from a 5-point Likert scale ranging from “disagree strongly” to “agree strongly”. The task was self-paced and typically took about 5 minutes to complete.

3) Birmingham Object Recognition Battery (BORB Battery) Position of Gap Match (Warrington & James, 1991)

This is a visual space perception task measuring location discrimination. This is a subtask of the BORB, which was designed to measure low-level visual deficits in patients with neuropsychological disorders of visual object recognition. A self-paced task, participants view one of forty pairs of two circles each with a small gap. Participants are required to identify whether the gaps in the circles are in the ‘Same’ position or ‘Different’ positions. The task was self-paced and typically took about 5 minutes to complete.

4) Interpersonal Reactivity Index (Davis, 1980)

This is a 28-item questionnaire measuring four independent empathy factors – Empathic Concern, Personal Distress, Perspective Taking and Fantasy. The questionnaire is self-paced and each question consists of a short phrase such as ‘After seeing a play or a movie, I have felt as though I am one of the characters’. Participants must state how accurately this reflects their experience, by selecting the most appropriate response from a 5-point Likert scale ranging from “does not describe me well” to “describes me very well”. The task was self-paced and typically took about 5 minutes to complete.

5) Letter Detection Task

This was an in house developed visual search task in which participants were required to read a passage about ‘France’ and count the number of ‘F’s in the text. This task was included to provide an object detection task with which performance on the face detection task could be compared. The text was 300 words long and contained 50 'F's. Answers were banded rather than using the raw scores. The optimum score of 5 was given for the correct answer. Answers of 1-19 'F's were given a score of 1, 20–32 'F's were given 2, 33–39 'F's were given 3, and 40–49 or 51–60 'F's were given a score of 4. Accurately identifying the number of 'F's was a very difficult task and banding allowed participant's ability to be measured more sensitively and accurately since a score of 45 was much closer to the correct answer than 12. Scoring guesses above the target number was only carried out for a small number of numbers, the aim being to reduce the ability of individuals randomly entering high numbers to achieve high scores. The task was timed, participants were given 1 minute to read and complete the task.

6) Matching Familiar Figures Test (Kagan, 1965)

This task measures the cognitive style task impulsivity versus reflexivity. The task consists of 40 line drawings of common objects in which a target is depicted and an array of six variants. Participants are required to identify the variant that matches the target. The task was self-paced and typically took about 10 minutes to complete.

7) Mooney Faces Task (Mooney, 1957)

This is a task measuring perceptual closure – the perception of a face that is not immediately or completely represented. The test comprises of forty black and white images in which a face can be perceived. 20 of the images are male faces and 20 female male. Participants are required to identify whether the face shown depicts a male or female. The task was self-paced and typically took about 5 minutes to complete.

8) Navon Local Task (based on Navon, 1977)

This is a local processing task and was timed with each letter appearing on screen for 5 seconds. The Navon Local Task consisted of 24 presentations of either the letter H and S made up of either small letter Hs or Ss. Participants were required to identify the identity of the small letters by using the mouse to click on 'H' or 'S'. Participants' responses were timed and the task typically took about 5 minutes.

9) Navon Global Task (based on Navon, 1977)

This is a global processing task and was timed with each letter appearing on screen for 5 seconds. The Navon Global Task consisted of 24 presentations of either the letter H and S made up of either small letter Hs or Ss. Participants were required to identify the identity of the large letters by using the mouse to click on ‘H’ or ‘S’. Participants’ responses were timed and the task typically took about 5 minutes.

10) Visual Object and Space Perception (VOSP Battery) Position Discrimination Task (Warrington & James, 1991)

This is a visual space perception task from the VOSP battery for measuring fine detail space perception. The VOSP Battery was designed to measure mid-level visual skills that show deficits in patients with right-hemisphere injuries. A self-paced task, participants view one of twenty pairs of two squares, each containing a single black dot. Participants are required to identify which of the squares contains the centred spot. The task was self-paced and typically took about 5 minutes to complete.

11) (VOSP Battery) Silhouettes Task (Warrington & James, 1991)

This is an object perception task measuring central coherence and is a subtask of the VOSP Battery. Participants are required to identify common objects in silhouette from unusual perspectives consisting of 15 animals and 15 objects. The task was self-paced and typically took about 10 minutes to complete.

Procedure

This experiment was conducted on-line using the Qualtrics survey system. Subjects were tested individually and completed all of the measures. The face identity processing tasks were presented first in randomized order to reduce potential confounds from wider cognitive tasks e.g. Navon tasks. Participants then completed the remaining cognitive and social measures in randomized order. The task was self-paced, and typically took about 1hr 30 minutes to complete.

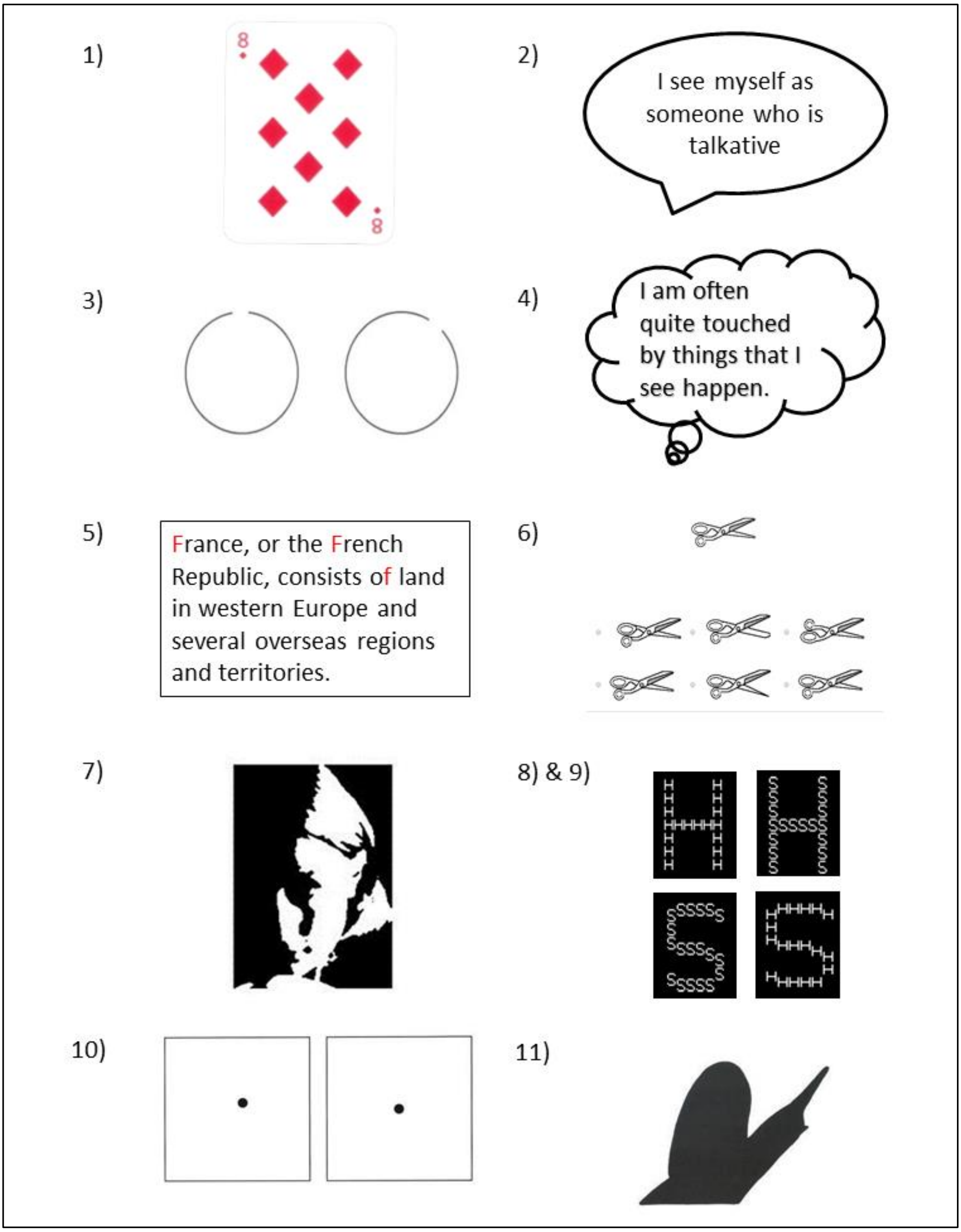

Figure 4.2. Example stimuli from each of the battery measures. Clockwise from the top left these are: 1) the BADS Card Sorting Task; 2) the Big 5 Inventory; 3) the BORB Position of Gap Match Task; 4) the IRI Empathic Concern Scale; 5) the Letter Detection Task ('f's highlighted here for demonstration purposes only); 6) the Matching Familiar Figures Task; 7) the Mooney Faces task; 8) & 9) Navon Local & Global Tasks; 10) the VOSP Position Discrimination Task; and 11) the VOSP Silhouettes task.

Results

Table 4.1. Summary data for all measures in the battery.

Measure	Mean	Std. Deviation	Range	Minimum	Maximum
GFMT	80.75	10.87	52.50	47.50	100.00
GFMT : Same	79.58	18.29	75.00	25.00	100.00
GFMT : Diff	81.92	14.77	65.00	35.00	100.00
BTWF	15.58	12.40	55.26	0.00	55.26
CFMT	69.85	15.06	66.67	31.94	98.61
CFMT : Identical	96.00	8.00	44.44	55.56	100.00
CFMT : Novel	65.11	18.59	76.67	23.33	100.00
CFMT : Novel + Noise	56.15	21.32	87.50	12.50	100.00
Face Detection	4.03	1.04	7.00	0.00	7.00
BADS	97.71	4.87	30.00	70.00	100.00
BADS Time	96.40	35.66	263.84	47.82	311.67
Big 5 Extraversion	24.82	6.31	29.00	9.00	38.00
Big 5 Agreeableness	32.85	5.76	27.00	18.00	45.00
Big 5 Conscientiousness	34.06	5.37	22.00	23.00	45.00
Big 5 Neuroticism	22.94	6.19	31.00	8.00	39.00
Big 5 Openness	35.09	6.11	33.00	17.00	50.00
BORB	85.70	13.57	60.00	40.00	100.00
BORB : Same	90.56	14.23	60	40	100
BORB : Diff	80.84	18.80	100	0	100
Letter Detection	1.91	0.84	4.00	0.00	4.00
IRI Perspective Taking	16.92	4.26	21.00	7.00	28.00
IRI Fantasy Scale	13.64	4.57	21.00	5.00	26.00
IRI Empathic Concern	18.51	4.41	20.00	8.00	28.00
IRI Personal Distress	11.32	4.20	22.00	0.00	22.00
MFFT	77.31	12.24	57.50	40.00	97.50
MFFT Time	527.98	268.68	1337.98	149.42	1487.40
Mooney Faces	84.38	12.68	100.00	0.00	100.00
Navon Global	92.48	16.50	50.00	50.00	100.00
Navon Global Time	47.19	10.08	46.56	30.02	76.58
Navon Local	95.44	13.18	50.00	50.00	100.00
Navon Local Time	45.40	9.49	49.48	26.20	75.69
Silhouettes	73.02	17.44	100.00	0.00	100.00
VOSP PD	97.38	4.42	20.00	80.00	100.00

Correlation Analysis

Table 4.1 shows the mean performance on each measure. Table 4.2 shows the relationships between performance (accuracy) on the unfamiliar face matching task (GFMT) the face detection task, the unfamiliar face recognition task (CFMT) and the familiar face recognition task (BTWF). Overall accuracy has been provided for the GFMT together with accuracy for same and different face pairs. Accuracy for the CFMT has been similarly broken down for the Identical task, the Novel task and the Novel + Noise task.

Table 4.2. Pearson's Correlation between performance on the face processing tasks.

	GFMT	GFMT : Same	GFMT : Diff	BTWF	CFMT	CFMT : Identical	CFMT : Novel	CFMT : Novel + Noise	Face Detection
GFMT	-	.740**	.554**	.200*	.504**	.433**	.424**	.485**	.079
GFMT : Same		-	-.149	.228*	.359**	.386**	.314**	.309**	.102
GFMT : Diff			-	.011	.298**	.159	.235*	.331**	-.010
BTWF				-	.333**	.143	.311**	.327**	.206*
CFMT					-	.553**	.943**	.936**	.159
CFMT : Identical						-	.422**	.431**	.051
CFMT : Novel							-	.791**	.145
CFMT : Novel + Noise								-	.164
Face Detection									-

Note - *p<.05 **p<.01

As Burton et al., (2010) found, there were significant individual differences in performance on the GFMT marked by a large standard deviation (M = 80.75, SD = 10.86). There was a wide range of accuracy (48% - 100%) with only 1 participant performing at ceiling and just 4 participants (3.7%) achieving a score in the top 2% (the cut-off point used by Davis et al., (in press) for identifying super-recognisers when completing the extended CFMT). The overall accuracy score was very close to the norms (M= 81.3%, SD = 9.7; Burton et al., 2010) and was both significantly and positively correlated with all face processing tasks with the exception of face detection. The correlation between unfamiliar face matching and unfamiliar face recognition was also found across all blocks; CFMT Identical, where no novel views of faces are shown; the CFMT Novel, where novel view of faces are shown; and the CFMT Novel + Noise, where views of faces are degraded. Power analysis in G Power 3.1 indicated that with a sample size of 107 participants, only *r* values greater than or equal to 0.34 have the required power for the null hypothesis to be confidently rejected. On this basis, the positive and significant relationship found between unfamiliar face recognition and unfamiliar face matching stands, however, the positive and significant relationship between unfamiliar face matching and familiar face recognition is underpowered.

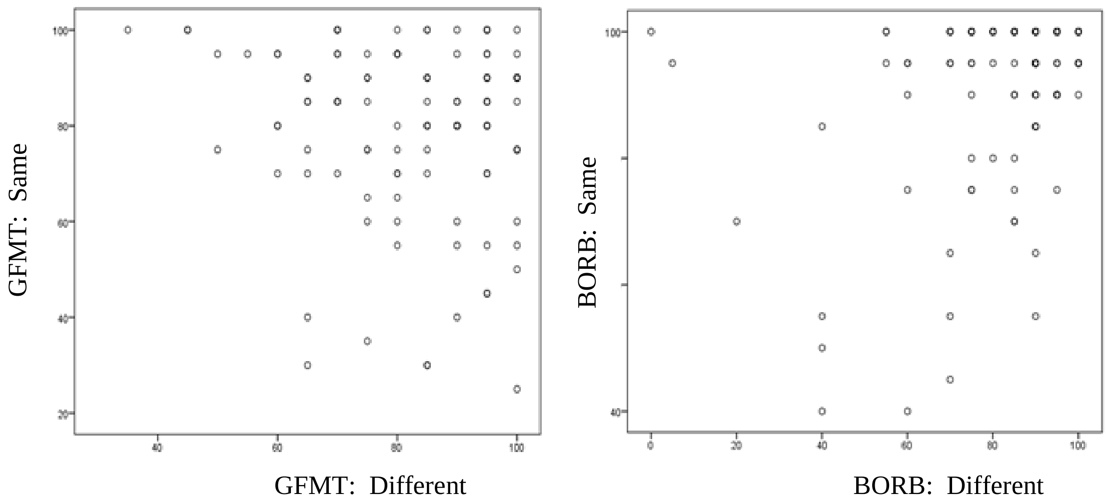

Figure 4.3. Scatter plots of GFMT Same & Different and BORB Same and Different accuracy by participant.

The scores for different face pair matching and same face pair matching were also investigated separately, since earlier research has indicated that these tasks are dissociated (Megreya & Burton, 2007). Whilst there was no significant difference between mean scores for GFMT same face pairs ($M = 79.58$, $SD = 18.29$) and GFMT different face pairs ($M = 81.92$, $SD = 14.77$), $t(106) = -9.60$, $p > .05$, there was no significant relationship between these scores, $r(107) = -.149$, $p > .05$, as shown in figure 4.3. This is in contrast to mean scores for the BORB same circles ($M = 90.56$, $SD = 14.23$) and BORB different ($M = 80.84$, $SD = 18.80$) circles where there was a significant difference in accuracy $r(106) = 5.189$, $p < .001$, and where there was a significant relationship between these scores, $r(107) = .338$, $p < .001$, as shown in figure 4.3. Accuracy for both GFMT same face pairs and GFMT different face pairs was positively associated with accuracy for the GFMT overall. However, there was significant variation in the relationships between these scores and other face processing measures. Accuracy on the GFMT same face pairs tasks correlated strongly with the CFMT and all three sub tasks; whereas accuracy for GFMT different face pairs correlated with the CFMT overall and the CFMT novel and novel + noise, but not with the CFMT Identical. Accuracy on the GFMT same face pair task was also significantly, but weakly, correlated with the BTWF task but this was not the case for the GFMT different face pairs task. This suggests that GFMT same face processing is qualitatively different from GFMT different face processing. GFMT same face processing is associated with matching identical images of faces (picture matching) and novel views of faces, whereas GFMT different face processing is associated to a larger extent with novel views of faces where noise is applied, a measure thought to test configural processing.

Unlike the dissociation found between GFMT Same and GFMT Different, the CFMT Overall correlates positively with all of the CFMT subtasks and all of the subtasks are positively associated. Like the GFMT neither the CFMT overall or any of the sub tasks are associated with the face detection task. The CFMT is also positively associated with the BTWF and with an *r* value of .33 is much closer to the required power to have confidence in the association. This suggests a stronger relationship between face recognition tasks than between face matching and face recognition tasks. However, not all of the CFMT subtasks are positively associated with the BTWF; the CFMT Identical task is not associated with the BTWF task. The BTWF task requires participants to recognise known adults as children, a condition in which they have not generally been seen. Since faces go through significant change between childhood to adulthood (Enlow, 1982) participants must carry out significant transformations to recognise the child version of a face from the known adult face. This may explain why the BTWF task is associated with the CFMT novel and CFMT novel + noise tasks since both require participants to carry out transformations to recognise a known face shown in unseen conditions e.g. different viewpoints/lighting and quality of images. This may also explain why the BTWF is the only face processing task to be associated, though weakly, with the face detection task. The scenes in which the faces were embedded may have required significant transformations for the face to be detected – an image that at first glance may have been seen as an ear, could also be seen as an image of a woman holding a baby.

The correlational analysis demonstrates that the face processing tasks, with the exception of the face detection task, are associated. The strengths and nature of these associations however differs both between and within face processing tasks. Table 4.3 shows the relationship between performance on the face processing tasks and the perceptual, cognitive and personality measures. Differences and similarities found here may provide even greater insight into the relationships between face processing tasks.

Table 4.3. Pearson's Correlation between performance on the face processing tasks and the perceptual, cognitive and personality measures.

Measure	GFMT	GFMT: Same	GFMT: Diff	BTWF	CFMT	CFMT: Identical	CFMT: Novel	CFMT Novel + Noise	Face Detection
BADS	.204*	.201*	.052	.136	.116	.109	.103	.103	-.117
BADS Time	-.045	.037	-.112	.042	-.137	-.017	-.147	-.125	-.153
Big 5 Extraversion	.032	-.059	.121	.036	.002	-.020	.029	-.023	-.048
Big 5 Agreeableness	-.054	.027	-.112	.017	-.065	.059	-.123	-.020	-.208*
Big 5 Conscientiousness	.171	.113	.111	-.030	.077	.149	.029	.090	-.144
Big 5 Neuroticism	.023	.061	-.041	-.036	-.046	-.030	-.061	-.022	.032
Big 5 Openness	.231*	.159	.142	.096	.114	.110	.094	.108	.004
BORB	.134	-.025	.229*	.060	-.048	-.026	-.057	-.032	-.088
IRI Perspective Taking	.170	.093	.135	.131	.075	.181	.002	.106	-.068
IRI Fantasy Scale	.161	.119	.090	.149	.071	.012	.030	.113	.040
IRI Empathic Concern	.203*	.131	.135	.058	.056	.160	-.017	.092	-.118
IRI Personal Distress	-.101	-.060	-.074	-.056	.002	-.024	.006	.004	.076
Letter Detection	.046	-.027	.102	.070	.011	-.040	.011	.024	.229*
MFFT	.367**	.238*	.245*	-.004	.026	.078	.075	-.049	-.055
MFFT Time	.042	-.081	.162	-.128	-.071	-.135	-.059	-.049	-.186
Mooney Faces	.097	.148	-.040	.175	.222*	.169	.201*	.203*	.197*
Navon Global	-.005	.042	-.059	.091	-.094	-.043	-.056	-.127	.031
Navon Global Time	.042	-.003	.066	-.074	-.117	-.090	-.121	-.090	-.119
Navon Local	.274**	.263**	.077	.029	.033	.020	-.020	.086	.101
Navon Local Time	.033	-.020	.074	-.118	-.136	-.131	-.126	-.115	-.190
VOSP PD	.100	.062	.070	.056	.141	.139	.134	.114	-.117
Silhouettes	.144	.175	-.006	.287**	.291**	.229*	.310**	.215*	.094

Note - *p<.05 **p<.01

Of the perceptual, cognitive and personality measures, there were very few significant associations overall with the face processing tasks. It was not the case that any one general measure was associated with all of the face processing tasks. This lack of association indicates that overall, face processing is a highly specific process and does not share mechanisms with other more general processes. Those associations that were found between face processing and other measures were specific to face processing tasks. Only one task, the MFFT, was associated with the GFMT overall and both same and different face pair matching. Though the relationships with the types of face pair were weaker, the relationship between MFFT and the GFMT overall was larger, $r(107) = .367$, $p < .01$, suggesting a strong relationship between face matching and object matching. The GFMT overall was associated with both the IRI empathic concern and the Big 5 Openness. Though both relationships were relatively weak, this might suggest a benefit in unfamiliar face matching for both emotionally engaging with the face and being able to abstract patterns from data. The GFMT overall also shared associations with GFMT same face pair matching in its positive relationship with Navon Local processing and the BADs card sorting task. This suggests that both the GFMT overall and matching same face pairs may involve a greater reliance on features and being able to switch between tasks than matching different face pairs. In contrast, matching different face pairs was only significantly, though again weakly, associated with the BORB gap match task. This indicates that the more able a participant was at discriminating between distances between features in faces the more able they would be at identifying different face pairs.

Of the other face related tasks, the CFMT and all of its subtasks were significantly, but weakly, associated with the Silhouettes task, a task measuring recognition of animals and objects shown from unusual viewpoints. These associations may have been driven by the requirement to transform and recognise shapes from new or unusual perspectives and memory. This is supported by the significant, but weak, association between the BTWF and the silhouettes task indicating a shared memory and requirement to transform images between both unfamiliar and familiar face recognition with familiar object recognition. The CFMT and the CFMT Novel and CFMT novel + noise tasks were also significantly, but weakly, associated with the Mooney Faces task, a task that measures perceptual closure. This suggests that unfamiliar face recognition is associated with weak central coherence or field independence. However, there was no association between unfamiliar face recognition and either local or global processing. This supports the lack of association between field independence and perceptual style found by Milne and Szczerbinski (2009) and the dissociation between global processing and unfamiliar face recognition found by Wang et al., (2012). Since the CFMT was not associated with local processing or the VOSP position discrimination task, it would not seem that unfamiliar face recognition is associated with cognitive flexibility as defined by Milne and Szerbinksi (2009). It should also be noted that the shared loading onto an Inhibition factor by the unfamiliar face recognition task with the VOSP position discrimination task found by Burgess et al., (1998) was not suggested here in any form of association. However, these results were found in a patient population with brain injury, which suggests that this relationship may not be found in a healthy population.

It is also interesting to note that the CFMT was not significantly associated with any of the personality measures, even those where associations might have been expected. Both extraversion and neuroticism have been found to be associated with unfamiliar face recognition in previous research (Li et al., 2010, Megreya & Bindemann, 2013). However, the relationship Li et al., identified was underpowered (r = 0.10, p = .03) increasing the risk of a type 1 error and explaining why no relationship was found here. Megreya and Bindemann (2013) found a negative relationship between face matching and neuroticism for their female participants only, which may explain why no relationship was found here since a wider population was used.

The CFMT was only associated with two of the 11 perceptual, cognitive and personality measures, the BTWF task was only associated with one. This suggests that face processing is a very specific mechanism showing greater levels of separation from general processes as face processing moves from matching towards recognition and familiarity. The contrast between the lack of association between familiar face recognition and the general processes measure here and the associations between the face detection task and other tasks also supports this separation. The face detection task had the least associations with any of the other face processing tasks, but is significantly associated with the letter detection task, the Mooney Face Task and the Big Five Agreeableness score. Though these associations are weak, this suggests that face detection is more related to processes underlying detection rather than faces and supports research showing that the detection and categorisation of faces are not only separate processes but also dissociable (Bindemann & Lewis, 2013).

Regression Analysis

Multiple hierarchical regression analysis was used with the face processing tasks entered first to identify the amount of variance explained by the wider measures once the variance explained by the face processing tasks was removed. Since performance on the GFMT same face pairs and different face pairs is dissociated, performance on the GFMT overall, GFMT same face pairs and GFMT different face pairs were all predicted separately.

Predicting performance on the GFMT

A multiple hierarchical regression analysis was conducted to predict performance on the GFMT from performance on the CFMT and the BTWF tasks. The regression equation with the face processing measures was significant R^2 = .26, F(2,104) = 17.86, p <

.001. This demonstrates that overall the face processing tasks account for 26% of the variance in GFMT performance. However, once the variance from the CFMT $t(104)$ = 5.49, $p < .001$, has been partialled out, the BTWF task, $t(104)$ = 0.40, $p > .05$, does not account for any significant residual variance. This would suggest that familiar face recognition does not predict unfamiliar face matching performance to any greater level than unfamiliar face recognition.

A second analysis was conducted to evaluate whether perceptual, cognitive and personality measures (MFFT, Navon Local, BADS, Big 5 Openness, and the IRI Empathic Concern) would predict performance on the GFMT over and above the face processing tasks. The perceptual, cognitive and personality measures accounted for a significant proportion of the GFMT variance (19%) after controlling for the effects of the faces processing tasks, R^2 change = .19, $F(5,99) = 11.22$, $p < .001$. Once the variance from the MFFT, $t(98)$ = 3.67, $p < .001$, and the Navon Local task, $t(98)$ = 2.08, $p < .05$, were partialled out, neither the BADS task, $t(98)$ = 0.49, $p > .05$, nor the Big 5 Openness measure, $t(98)$ = 1.19, $p > .05$, nor the IRI Empathic Concern measure, $t(98)$ = 1.51, $p > .05$, significantly account for any residual variance. The standardized beta values for the MFFT and Navon Local task (0.29 and 0.16 respectively) indicate that the MFFT has more impact in the model. Overall, unfamiliar face matching is predicted to the greatest extent by unfamiliar face recognition, followed by object matching and local processing performance.

Predicting performance on GFMT same face pairs

A multiple hierarchical regression analysis was conducted to predict performance on the GFMT same face pairs from performance on the CFMT and the BTWF tasks. The regression equation with the face processing measures was significant $R^2 = .14$, $F(2,104) = 8.60$, $p < .001$. This demonstrates that overall the face processing tasks account for 14% of the variance in GFMT same face pairs performance. However, once the variance from the CFMT, $t(104)$ = 3.30, $p < .001$, has been partialled out, the BTWF task, $t(104)$ = 1.27, $p > .05$, does not account for any significant residual variance. This would suggest that familiar face recognition does not predict unfamiliar face matching performance for same face pairs to any greater level than unfamiliar face recognition.

A second analysis was conducted to evaluate whether perceptual and cognitive measures (MFFT, Navon Local, BADS) would predict performance on the GFMT same face pairs over and above the face processing tasks. The perceptual and cognitive measures accounted for a significant proportion of the GFMT same face pairs variance (10%) after controlling for the effects of the faces processing tasks, R^2 change = .19, $F(3,101) = 6.36$, $p < .001$. Once the variance from the Navon Local task, $t(101) = 2.10$, $p < .05$, was partialled out, neither the MFFT, $t(101) = 1.85$, $p > .05$, nor the BADS task, $t(101) = 0.84$, $p > .05$, significantly account for any residual variance. Overall, unfamiliar face matching same face pairs is predicted to the greatest extent by unfamiliar face recognition, followed by local processing performance.

Predicting performance on GFMT different face pairs

A multiple hierarchical regression analysis was conducted to predict performance on the GFMT different face pairs from performance on the CFMT and the BTWF tasks. The regression equation with the face processing measures was significant $R^2 = .10$, $F(2,104) = 5.63$, $p < .005$. This demonstrates that overall the face processing tasks account for 10% of the variance in GFMT different face pair performance. However, once the variance from the CFMT, $t(104) = 3.35$, $p < .001$, has been partialled out, the BTWF task, $t(104) = 1.00$, $p > .05$, does not account for any significant residual variance. This would suggest that familiar face recognition does not predict unfamiliar face matching performance for different face pairs to any greater level than unfamiliar face recognition.

A second analysis was conducted to evaluate whether perceptual and cognitive measures (MFFT, BORB) would predict performance on the GFMT different face pairs over and above the face processing tasks. The perceptual and cognitive measures accounted for a significant proportion of the GFMT different face pairs variance (9%) after controlling for the effects of the faces processing tasks, R^2 change = .09, $F(2,102) = 6.07$, $p < .001$. Once the variance from the BORB task, $t(102) = 2.21$, $p < .05$, was partialled out, the MFFT task, $t(102) = 1.97$, $p > .05$, did not significantly account for any residual variance. Overall, unfamiliar face matching different face pairs is predicted to almost the same extent by unfamiliar face recognition and space perception.

This analysis suggests that performance in face processing, and in particular unfamiliar face recognition can be used to predict performance in unfamiliar face matching. Of the measures and tasks used in the battery, unfamiliar face recognition predicts the largest amount of variance for unfamiliar face matching overall (26%). Performance in the wider tasks also predicts an additional 19% of performance in the unfamiliar face matching task, with object matching and local processing predicting the largest amounts of residual variance. Together, the CFMT, MFFT and the Navon Local

Task provide a useful set of tools for predicting almost half of the variance in performance on the GFMT. These results also suggest that unfamiliar face matching uses both face specific mechanisms and mechanisms that are more general.

Performance in same face pair matching can also be predicted by performance in face processing but to a lesser degree (14%). More general tasks can also predict performance in same face pair matching but again to a lesser degree than for unfamiliar face matching overall (10%). It is important to note that unlike unfamiliar face matching overall only the local processing task predicted the largest amount of residual variance for the wider tasks. This suggests that object matching and unfamiliar face matching share mechanisms that are not used when matching unfamiliar same face pairs alone.

Performance in different face pair matching can also be predicted by performance in face processing but again, to an even lesser degree (10%). General task can also predict performance in different face pair matching but to again a slightly lesser degree (9%) than for same face pair matching or unfamiliar face matching overall. Most strikingly, performance in space perception alone accounts for largest amount of residual performance in different face pair matching. This is only the case for different face pair matching; neither same face pair matching nor face matching overall is predicted by space perception performance.

Performance in unfamiliar face matching and both same and different unfamiliar face pair matching is predicted by performance in both face specific and more general mechanisms. It may well be that some of the 24% of performance in unfamiliar face matching overall is made up through the summing of the face processing mechanisms used by both same and different face pair matching. Both same and different face pair matching associate with some shared aspects of the CFMT but they do so to different degrees and also associate quite separately with other aspects of the CFMT allowing these different degrees and separate associations to be summed. Perhaps most importantly the regression analysis identifies that the more general mechanisms predicting performance in the different face pair matching do not predict performance in any other types of face matching. Predicting performance across all three measures (overall, same and different) can only be achieved by using the CFMT, the MFFT, local processing *and* the BORB Gap match task.

General Discussion

This aim of this Chapter was twofold; to investigate the relationships between different face processing tasks and, to investigate the relationships between face processing and perceptual, cognitive and personality measures.

An individual differences approach was used specifically to identify relationships between measures. It was hoped that these relationships would provide greater insight into the processes or mechanisms underlying behaviour. Earlier research used single measures that they believed would represent singular constructs or mechanisms e.g. face recognition or holistic processing would tap mechanisms or processes such as face processing or configural processing in their entirety. This ability of single measures to represent particular mechanisms has been shown to be unreliable. Therefore, each measure used in this chapter was selected because of its ability to tap particular aspects of a mechanism or process. It is the combination of these aims and this approach, which has both supported earlier findings and provided novel findings.

The findings of this battery demonstrate that face processing tasks are highly related and that face processing as a general mechanism is highly specific. This is the first time that this has been shown across familiar face recognition, unfamiliar face recognition and unfamiliar face matching in the general population. With the exception of the face detection task, unfamiliar face recognition, familiar face recognition and unfamiliar face matching are associated. However, the relationship between unfamiliar face matching and the familiar face recognition task was weak. The weakness of this relationship may explain why this relationship has not been found consistently in earlier research (Megreya & Burton, 2006). Earlier research provides findings that might explain the weaker nature of this relationship. Familiarity influences face processing, providing an advantage in accuracy for both face recognition and face matching (Clutterbuck & Johnston, 2002; Ellis et al.,1979), and removes the dissociation found between hits and false positives in face matching tasks (Megreya & Burton, 2007). Familiar face recognition models demonstrate that this familiarity provides associations to rich semantic knowledge that support and enhance facer recognition ability (Bruce & Young, 1986; Haxby et al., 2000). The mechanisms underlying unfamiliar face recognition are less understood, but it is clear that unfamiliar faces have less association to semantic knowledge and this may account for the reduction in accuracy. However, the lack of association with semantic knowledge does not appear to result in unfamiliar faces having no association with familiar face processing, unlike Megreya and Burton's (2006) claim. The weak but significant relationship found in this study suggests that there are still shared, if attenuated, mechanisms underlying face processing regardless of their familiarity, such as structural codes that have not been enhanced through exposure to multiple viewpoints or viewing conditions (Bruce & Young, 1986).

This battery has demonstrated that face processing tasks were highly associated and, overall, highly specific. There were no general tasks or measures that associate with all of the face processing tasks. However, there were specific general tasks and measures that associate with specific face processing tasks. Unfamiliar face matching is one of these but it is not unique in this position. Face recognition tasks show an association with object recognition tasks and face matching tasks show an association with object matching tasks and other wider measures. However, these relationships are weak. Unfamiliar face matching, however, is unique in the dissociation between its sub tasks. Neither the subtasks of the unfamiliar face recognition task nor the subtasks of the space perception task were dissociated. It is important to note however that the space perception task was not an object matching task per se but a test of space perception. Given that 2nd order relations, the distances or spaces between the features, has been identified as a type of configural processing, a space perception task might also tap mechanisms used in configural processing of faces and objects. Configural and featural processing for objects have been shown to be dissociated (Yovel & Kanwisher, 2008). This would suggest that face processing tasks usually use shared mechanisms despite differences in stimuli whereas same and different decisions regarding objects do not usually require different mechanisms.

Same and different face matching associates and dissociates with the face processing tasks and the general processing tasks in different ways. These general tasks are not shared but specific to each type of face matching. Same face pair matching is associated with and predicted by local processing. Different face pair matching is associated with and predicted by space perception. The influence of these separate general mechanisms may also account for the different ways the subtasks of the GFMT associated and dissociated with the CFMT and BTWF task. GFMT same matching had stronger associations with the CFMT Identical images and the CFMT novel images. Whereas GFMT different matching was dissociated from CFMT Identical images and more associated with CFMT novel images shown with noise than those shown without noise. The difference in the associations between same and different face matching and the

CFMT support the difference in their associations with more general tasks. Different face matching is most highly associated with recognising novel images of unfamiliar faces shown with noise, a task designed to isolate configural processing - a form of processing which includes perception of space between features.

The results here would suggest that Megreya and Burton (2006) might not have been correct in their assertion that unfamiliar faces are objects. Processing unfamiliar faces when compared to familiar faces uses more general mechanisms - memory for unfamiliar face recognition and local processing and space perception for unfamiliar face matching. This suggests that unfamiliar faces are processed to some extent as objects. However, the correlation analysis clearly shows that familiar face recognition, unfamiliar face recognition and unfamiliar face matching overall are associated; unfamiliar faces are also faces. Together these results suggest a highly specific face processing mechanism underlies face processing tasks, but like other researchers have shown, also suggests that there are independent latent variables of face cognition: such as face perception, face memory and speed of face cognition (Wilhelm, Herzmann, Kunina, Danthiir, Schacht & Sommer, 2010). These independent variables differ in terms of the proportion and type of more general mechanisms being used, and are activated according to the type of faces processed and the processing task being performed.

These differences influence the degree by which performance in any face processing task can be predicted by performance in any other face processing task. Unfamiliar face matching is unique in the type and combination of mechanisms underlying the task. Overall, in terms of associations with other face processing tasks, unfamiliar face matching performance can most reliably be predicted by performance in the unfamiliar face recognition task. Almost equally, unfamiliar face matching can be predicted by performance in the object matching task plus the local processing tasks. However, if predicting performance is to be truly optimised then this study would suggest that performance on same face pairs and different face pairs should not be combined, but predicted independently. Further work is required to confirm these findings and identify how same and different face matching might be affected by different conditions e.g. inversion or with noise applied and this will form the basis of the experiments contained in Chapter 5.

Chapter 5 – Face Matching Mechanisms

5.1 Introduction

Chapter 4 of this thesis identified that unfamiliar face matching was related to performance in other face processing tasks and more general measures. Perhaps most interesting of all, are the relationships found once same face pair and different face matching are examined separately. Since earlier research had shown that performance on same face matching and different face matching is dissociated (Megreya & Burton, 2007) same face pair and different face matching were reviewed separately in Experiment 8. The results replicated the finding that these unfamiliar face matching tasks were dissociated but also demonstrated that relationships between these tasks and other face processing and more general measures were also dissociated. Multiple hierarchical regression identified that the unfamiliar face recognition task and Navon local processing task accounted for 14% and 10% of the variance in the GFMT same face matching task. Whereas, the unfamiliar face recognition and space detection tasks accounted for 10% and 9% of the variance in the GFMT different face pairs task. Same face matching appears to be primarily based on face processing mechanisms whereas different face matching is based on both face processing mechanisms and more general space perception mechanisms. This was also reflected in the different ways same face matching and different face matching are associated with the CFMT novel and CFMT novel + noise tasks. Same face matching's association with CFMT novel is higher than the association with CFMT novel + noise task, whereas for different face matching this pattern is reversed. This distribution of variance and associations indicates that different face matching may have much less to do with face processing mechanisms than previously thought.

This may seem contradictory, as there is a large body of research that suggests configural processing is fundamental to face processing. The inversion effect first demonstrated by Yin (1969) showed that when faces and objects were viewed in an upright position, faces were identified with much greater accuracy than objects. In contrast, when both the faces and objects were inverted, the reduction in accuracy was much greater for faces than objects. Yin (1969) required participants to recognise identical images of individuals and objects, a test of image recognition rather than face recognition. However, the inversion effect was shown to be robust when recognising different views of learned individuals, changed emotional expressions, and unaffected by the familiarity of the faces (Valentine & Bruce, 1986). As discussed in Chapter 4, it is important to note that configural processing refers to three separable types of processing: 1st order relations – the eyes being above the nose and the mouth; 2nd order relations – the distances between the features; and holistic processing – perceiving the face as a whole or gestalt (Maurer et al., 2002). Inversion disrupts the first order relations resulting in a reduction for accuracy. However, inversion has also been found to disrupt holistic and second order relations processing. Farah, Tanaka and Drain (1995c) demonstrated that when participants were required to learn faces in terms of their separate parts (featurally) and as whole faces (holistically), faces learned holistically were recognised much more poorly when inverted than the faces learned featurally. Young et al., (1987) demonstrated that when the top and bottom halves of separate famous faces were combined into a new gestalt face, the original identities were much more difficult to recognise. It was only when the faces were inverted and the gestalt disrupted, that the original famous faces became easier to identify. This effect was also replicated using unfamiliar faces in a face matching task by Hole (1994). Inversion has also been shown to affect performance in both unfamiliar face matching and recognition tasks. When faces are inverted featural changes have been identified much more accurately than changes in distances between the features (Freire et al., 2000; Mondloch et al., 2002). These findings led researchers to conclude that the inversion effect appeared to be due to a greater reliance on configural processing for upright faces that was compromised when the faces were inverted.

Many researchers have argued that viewing stimuli holistically and incorporating the spatial distances rather than using individual features is what makes processing faces different to processing objects (Tanaka & Farah, 1993; Farah, Wilson, Drain & Tanaka, 1998; Maurer et al., 2002). Other researchers have countered this view, suggesting that the reliance on configural processing is due to developing expertise in faces, an expertise that can be developed for other classes of objects (Valentine, 1988). The reliance on configural processing over featural processing has also been contested. Tanaka and Sengko (1997) used a face recognition task to demonstrate that configural and featural information is interdependent, with subjects able to recognise features best when presented in the original face configuration, next best in a new configuration and most poorly when shown in isolation. Yovel and Kanwisher (2004) showed that when the configural and featural changes made to stimuli in a face matching task were controlled for difficulty, the inversion effect was not greater for configural changes. The featural changes made to the stimuli were made using brightness and contrast rather than replacing features in an original face with features from a donor face. This was done in order to reduce the configural changes that also occurred when features were changed in a face (Rakover, 2002). Yovel and Kanwisher (2004) also ensured that the stimuli were not presented in a blocked format i.e. all trials of one type shown before trials of another type, as they were concerned that this might have resulted in the effects found in earlier studies (Freire et al., 2000; Mondloch et al., 2002). Riesenhuber, Jarudi, Gilad and Sinha (2004) also used an unfamiliar face matching task to test the inversion effect on matching accuracy with configurally and featurally altered faces. Unlike Yovel and Kanwisher (2004), they used donor features to make featural changes to the faces, but like Yovel and Kanwisher they were also concerned of the effects of blocking by face change type. Riesenhuber et al., (2004) ran two versions of their matching task, a blocked task and an unblocked task. As predicted, there was a blocking effect. Participants who viewed featurally altered faces first were much poorer in their performance over all configurally-altered faces than all featurally altered faces. The participants who viewed configurally altered faces first, or for whom faces were not blocked, showed no difference in accuracy between all configurally altered faces and all featurally altered faces. These findings suggest that participants are not fixed in their primary use of configural or featural processing, but that they can be strategic in their use of information, adapting to the task at hand. Other researchers have gone further, comparing human and ideal performance on a familiarised face recognition task (Gold, Mundy & Tjan, 2012) argue that the perception of a face is no more than the sum of its parts.

Whilst there continues to be debate around this issue, the majority of researchers, suggest that any differentiation between a configural or featural approach to face processing is not so much in kind, but of degree. Faces, unlike objects, are more reliant on configural processing, however recognising or matching faces uses a combination of both featural and configural mechanisms (for review see Farah et al., 1998). It is this *combination* that Yovel and Kanwisher (2004) argue is specific to face processing. Using an individual differences approach and correlation analysis Yovel and Kanwisher (2008) showed that featural and configural processing were associated when upright faces were being matched but dissociated when the stimuli being matched were inverted faces and upright and inverted houses. This finding was in direct contrast to researchers that have identified configural and featural processing in faces as dissociated (Cabeza & Kato, 2000).

There are many studies that would lend support Cabeza and Kato's (2000) view. Configural and featural face processing strategies evoke different scan patterns when matching unfamiliar faces. Manipulations isolating configural processing showed more interfeatural saccades whereas manipulations that isolated featural processing showed longer gaze durations (Bombari, Mast & Lobmaier, 2009). Studies with children and adults have shown that featural and configural processing differ developmentally. Children predominantly use featural processing to recognise faces, while configural processing develops at a much slower rate only really maturing in adulthood (Carey & Diamond, 1977; Mondloch et al., 2002). The different developmental pathways for configural and featural processing in adults and children were also found in a passive unfamiliar face recognition study recording event related potentials (ERPs). Adults showed that the effect on the right hemisphere N170, the component thought to be specialised for faces or objects of expertise, was significantly greater for configural changes compared to featural changes and that the left hemisphere N170 is significantly greater for featural than for configural changes. 8 month old participants showed the same hemispheric differences in the P400 component (the component thought to be specialised for faces or objects of expertise in infants) whereas there were no differences observed for 4 month old infants (Scott & Nelson, 2006). Maurer et al., (2007) used fMRI imaging to demonstrate that whilst participating in an unfamiliar face matching task, participants showed relatively greater activation in the right fusiform gyrus and the right frontal cortex when matching configurally altered faces. However, when they were matching featurally altered faces there was greater activation in the left prefrontal cortex. Whilst the FFA was shown to be active in processing both types of altered faces these results suggest again that featural and configural processing is both lateralized and distinct. This double dissociation has also been shown using repetitive transcranial magnetic stimulation (TMS) in an unfamiliar face matching task. Stimulation over the left middle frontal gyrus has been shown to selectively disrupt featural processing, whereas stimulation over the right inferior frontal gyrus selectively disrupts configural processing (Renzi, Schiavi, Carbon, Vecchi, Silvanto, & Cattaneo, 2013).

There is then a great deal of evidence suggesting that configural and featural processing of faces is dissociated, however Cabeza and Kato (2000) may have been premature when they explicitly stated they were dissociated. Configural and featural processing may well develop at different rates and be associated with different physiological locations but this does not mean that they are not associated. This dissociation may have been inferred because of the differing effects of inversion on configurally and featurally altered faces in face processing tasks rather than shown. They did not carry out any correlational analysis and those that have, found that unlike for objects, configural and featural processing was associated (Yovel & Kanwisher, 2008). The results from Experiment 8 seem to both support and refute Cabeza and Katov (2000) and Yovel & Kanwisher (2008). In the unfamiliar face recognition task, performance for novel views of faces was correlated with performance for novel views of faces shown with noise, a task thought to isolate configural processing. Since it is widely agreed that face recognition uses aspects of both featural and configural processing, a task recognising novel views of faces would recruit both types of processing. Since both types of processing are recruited, one might expect a task that isolates configural processing to dissociate from a task recruiting both if Cabeza and Koto (2000) are correct. However, this is not the case. If Yovel and Kanwisher (2008) are correct and configural and featural processing in faces are associated then this would explain why performance on novel views of faces and novel views of faces with noise was associated. What then might this mean for same and different face matching?

Only two of the studies referenced above separated accuracy across same and different face matching. Riesenhuber et al., (2004) provided accuracy by trial type for the unblocked condition in their study. They showed that accuracy for identifying upright same face pairs was approximately 87% and upright featurally altered faces was approximately 86%. However, upright configurally altered faces were identified only approximately 78% of the time. When presented inverted, featurally altered faces were identified 68% of the time, configurally altered around 66% of the time and same face pairs 63% of the time. There was no significant difference in matching accuracy between featurally and configurally altered faces when they were inverted. This suggests that whilst inversion affects all types of processing equally, when the faces are presented upright, identifying configural changes is more difficult that identifying featural changes.

(Farah et al., 1998) also broke down results for same and different trials in their study. They ran a series of experiments investigating holistic processing in faces and objects and found that overall participants were more accurate at identifying same face pairs than different face pairs. There were no other significant differences between same face pairs and different face pairs with the exception of a task where participants viewed pairs of faces where one was unaltered and the other altered in either one feature (mouth, nose or eyes) or all features. After viewing the face pair they were asked to compare either a single features or all features to identify whether the faces were the ‘Same’ or ‘Different’. Same face pairs were much more easily identified when all of the features remained the same but this was reduced when the other features were different. However, this was not the case for different face pairs where performance was not significantly different whether one feature or all features had been changed. This might support a featural strategy being used for same face pairs since accuracy dropped when one feature was the same but others were different, leading to confusion about the correct response. The results for the different face pair may also suggest configural processing was being used because there was no difference between accuracy whether all features differed or one feature differed.

As Rakover (2002) noted, it is very difficult to make a featural change without making a configural change and vice versa. By changing one feature, distances between this feature and the other features will be changed. By changing all of the features distances between the feature queried and the other features will be changed. Attending to the distances between features will provide the same level of accuracy regardless of how many features have been changed. This might account for their being no difference in accuracy for 'Different' face pairs whether one feature or all features have been changed. Farah et al., (1998) did not examine the divergence in accuracy for same and different face pairs when shown upright and inverted but the figures were provided. In the experiment described above, accuracy for identifying same face pairs was 80.6% for upright faces and 82.6% for inverted faces, whereas accuracy for identifying different face pairs was 69.1% when upright and 62.8% when inverted. This again suggests that different face matching may be more reliant on configural processing and therefore more affected than inversion than same face pair matching.

The changes in methodology described above and their effects lends weight to those researchers, including Young et al., (1987) and Hole (1994), who stated that the contributions of configural and featural processing to any face processing task may well depend on the nature of the task. The breakdown of accuracy into same and different face pairs shows potential for greater learning about the mechanisms underlying unfamiliar face matching. However, this is not the only way the unfamiliar face matching task in Experiment 8 was different to those cited above. The task used in Experiment 8, was the Glasgow Face Matching Task (GFMT: Burton et al., 2010). The GFMT uses multiple images in both same face pairs and different face pairs. The images shown in a same face pair were taken only seconds apart with different cameras. In doing so, these images capture a small but none the less significant amount of the ways in which images of individuals can vary, or within person variability (Jenkins et al., 2011). The studies referenced above used the same image for same pair matching or old face recognition and configurally or featurally altered versions of those images for different face matching or new face recognition.

The type of image used was also highly variable across these studies ranging from photographs, mac a mug photofits to pencil drawings (Riesenhuber et al., 2004; Farah et al., 1998; Sergent, 1984). Such alterations are not representative of faces in the natural world and certainly do not capture any of the ways an individual face can vary, e.g. lighting, viewpoint, image quality. Given this lack of variation in the images, these studies do not provide any insight into how variation can be overcome to identify that two images show the same individual. The studies mentioned also made changes to the spatial relations between features and features themselves in order to induce either featural or configural processing. The results of experiment 8 suggest that these mechanisms occur naturally and are used independently when matching same and different face pairs from the GFMT.

The results of experiment 8 also suggest that the types of configural processing used when matching different face pairs may be associated with those used in unfamiliar face recognition but also those used in perceiving spaces in objects. The lack of association between same and different face matching may be due to the lack of association found between featural and configural processing in objects rather than any lack of association found between featural and configural processing in faces. This appears to support Yovel and Kanwisher's (2008) findings; however, this is the first time that these relationships have been examined. It is the aim of this chapter to examine the findings of experiment 8 in more detail and discover to what extent same face matching uses featural processing and different face matching uses configural processing. To this end two approaches are taken; first to replicate and extend the findings of experiment 8 with latency measures to ensure the associations found are robust; and second to manipulate the stimuli used in Experiment 8 to investigate causal relationships between configural and featural processing and same and different face matching. Experiment 10 aims to isolate featural processing by requiring participants to match unfamiliar pairs of eyes, Experiments 11, and 12 aim to isolate and disrupt configural processing through the addition of random heavy noise and inverting the faces respectively.

5.2 Experiment 9

Introduction

Accuracy for same and different scores in face matching tasks has been shown to be dissociated in both 1 in 10 array matching tasks and for simple paired items (Megreya & Burton, 2007). This suggests that same and different face pairs may be processed in a qualitatively different way. Experiment 8 demonstrated that unfamiliar face recognition and unfamiliar face matching were associated when the GFMT and CFMT were used as part of the face processing battery. However, same face matching and different face matching were dissociated and associated with the CFMT sub tasks in different ways. Same face matching was associated with all of the CFMT sub tasks and least so with novel faces shown with noise. Different face matching was associated with only two of the CFMT sub tasks and to the largest degree with novel faces shown with noise. Earlier research had shown that unfamiliar face matching and unfamiliar face recognition was not associated (Megreya & Burton, 2006). However, this was the first time that performance across these particular tasks had been compared. These findings suggest that different face matching may have a stronger relationship with configural processing than same face matching which is a novel finding. To have greater confidence in these findings and the ways in which same and different face matching were associated with the CFMT sub tasks this experiment aims to replicate and extend the findings of the battery. These findings will be extended through latency measures that have been identified as a useful indicator of effort and qualitative changes in behaviour (Marotta et al., 2002; Davidoff & Landis, 1990; Bruce, 1982)

Method

Participants

Participants were 84 members of the Call for Participants, Facebook, Gumtree and Reddit online communities (36 male) who were over 18, British citizens, and reported normal or corrected to normal vision (mean age = 34.02 range = 18-76). Participants completed the task as a goodwill gesture.

Materials

All participants completed two face identity processing tasks; the GFMT and the CFMT. Example stimuli from each of these tasks are shown in figure 5.1.

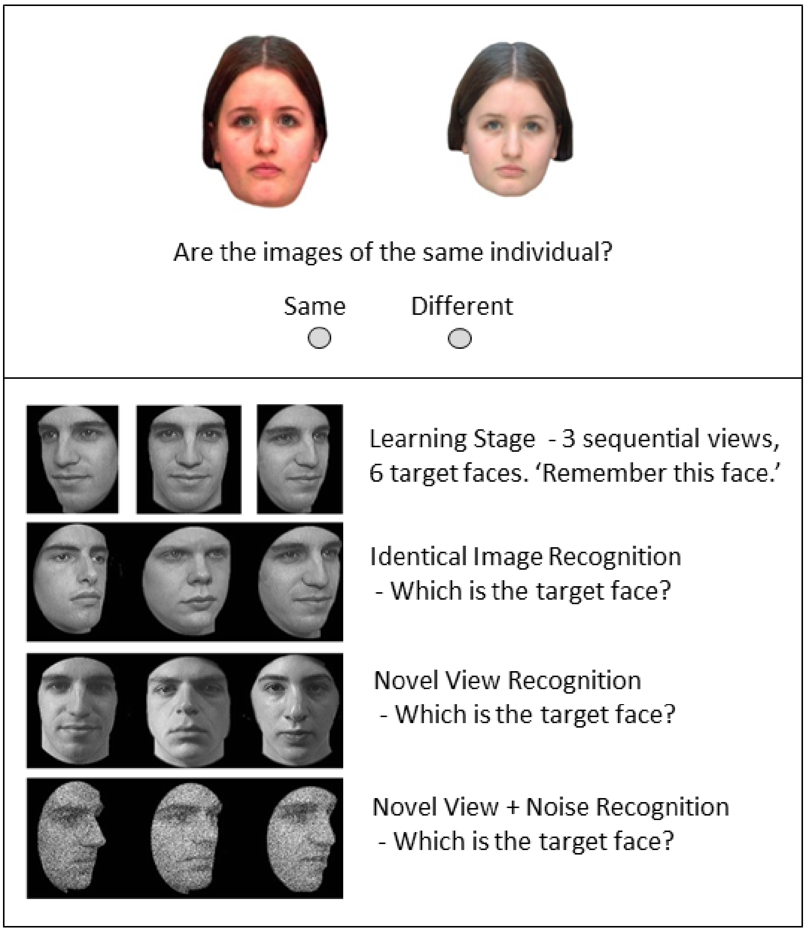

Figure 5.1. Example stimuli from the GFMT and the CFMT.

Glasgow Face Matching Test (GFMT - short form) (Burton et al., 2010).

The GFMT is an unfamiliar face matching task. The GFMT (short version) consists of 40 pairs of unfamiliar faces, half of which are same face pairs and half of which are different face pairs. Each face image is front facing in pose, neutral in expression, shown in colour and standardised to a width of 151 pixels. The photos within each pair were taken a few seconds apart using different cameras (for more details see Burton et al., 2010). Participants were asked 'Are the images of the same individual?' and selected 'Same' or 'Different' using the mouse. The task was self-paced, and typically took about 5 minutes to complete.

Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006a)

The CFMT (Duchaine & Nakayama, 2006a) is a 72-item face recognition-memory task, which is split into three sections. In section one, participants are told to learn a target face shown in three orientations (left facing, forward facing, right facing); they are then presented with a three-alternative forced choice task in which they have to pick out the identical face image. This process is repeated for each of six target faces. In section two, the three-AFC test is retained, with participants now having to identify novel instances of each target face. Section three is identical to section two, with the exception that the test images have had visual noise added to them in order to make the task more challenging. Presentation of the stimuli for memorising was timed; presentation of the target stimuli was not. There was no time limit for responses and the task typically took about 15 minutes to complete.

Procedure

This experiment was conducted online. Each participant completed both the GFMT and CFMT. The presentation order of the tests was randomised and in the GFMT only presentation order of the face pairs was randomised. Figure 5.1 shows a trial from both the GFMT and the CFMT.

In the GFMT each face pair was presented on screen individually and participants were asked, on-screen 'Are the images of the same individual?' and selected responses 'Same' or Different' with a mouse. The task was self-paced, with no time limit for responses, and took around 5 minutes to complete.

In the CFMT, faces were shown either individually or as a group to be learned and presentation was timed. Participants were then shown 3 faces, were asked on screen 'Which face is one of the target faces?' and selected the face with a mouse. While presentation of the faces to be learned was timed, the recognition task was self-paced, with no time limit for responses. The CFMT overall took around 15 minutes to complete.

Results

Face Matching Accuracy

Mean face matching accuracy is shown in Table 5.1 with data broken down into same and different face pairs. There was no significant difference between accuracy achieved for same face pairs (M = 90.89, SD = 13.23) and different face pairs (M = 88.04, SD = 13.30); $t(83) = 1.34$, $p > .05$.

Face Matching Decision Times

Decision times are reported here across all response types - correct and incorrect. Mean face matching decision times are shown in Table 5.1 with data broken down into same and different face pairs. There was a significant difference between accuracy achieved for same face pairs (M = 8.05, SD = 3.62) and different face pairs (M = 9.47, SD = 5.23); $t(83) = -2.90$, $p < .01$. On average, participants took longer to reach a decision about different face pairs than same face pairs.

Table 5.1. Summary data for the unfamiliar face matching task broken into same and different face pairs.

GFMT			
Accuracy (%)		Latency (secs)	
Same Face Pairs	Different Face Pairs	Same Face Pairs	Different Face Pairs
90.89 (13.24)	88.04 (13.31)	8.05 (3.62)	9.47 (5.23)

Standard deviations in parentheses

Face Recognition Accuracy

Mean face recognition accuracy is shown in Table 5.2 with data broken down into the different sections of the CFMT: identical, novel and novel + noise. A single factor (face type: identical, novel, novel + noise) within subjects ANOVA showed a significant main effect of face type $F(2,251) = 116.60$, $p < .001$, $\eta_{\rho}^2 < .58$. Tukey HSD revealed significant differences between all pairs of conditions and confirmed that recognising novel instances of faces with noise applied resulted in the lowest accuracy achieved.

Table 5.2. Summary data for the unfamiliar face recognition task broken down into the identical, novel image of faces and novel images shown with noise sections of the CFMT.

CFMT					
Accuracy (%)			Latency (secs)		
CFMT Intro	CFMT Novel	CFMT Novel + Noise	CFMT Intro	CFMT Novel	CFMT Novel + Noise
98.41 (7.35)	78.82 (16.27)	73.41 (19.45)	5 (1.72)	6.93 (2.89)	6.66 (2.39)

Standard Deviations in parentheses

Face Recognition Decision Times

Decision times are reported here across all response types - correct and incorrect. Mean face recognition decision times are shown in Table 5.2, with data broken down into the different sections of the CFMT: identical, novel and novel + noise. A single factor (face type: identical, novel, novel + noise) within subjects ANOVA showed a significant main effect of face type $F(2,251) = 64.50$, $p < .001$, $\eta_{\rho}^2 < .48$. Tukey HSD revealed significant differences between the identical and novel images of faces and novel images of faces shown with noise; however, there was no significant difference between novel faces and novel faces shown with noise.

Correlations

Table 5.3 provides a correlation matrix for performance on the GFMT, same and different face pairs and the CFMT and face recognition sections. As expected performance on the GFMT and the CFMT overall were positively associated. Performance on the GFMT was positively associated with all of the CFMT sub sections and performance on the CFMT was positively associated with both same and different face pairs. However, whilst performance on all of the CFMT subsections was positively associated, performance on same and different face pairs was dissociated $r(84) = -.081$, $p > .05$, suggesting different qualitative processes may be used. Performance on same face pairs was positively associated with all of the face recognition sections: identical $r(84) = .441$, $p < .001$; novel $r(84) = .375$, $p < .001$; and novel + noise $r(84) = .337$, $p < .001$. Whereas, performance on different face pairs was dissociated from performance on the identical face recognition section $r(84) = .139$, $p > .05$ but was positively associated with performance on the novel faces $r(84) = .329$, $p < .001$; and novel + noise $r(84) = .349$, $p < .001$.

Table 5.3. Face matching & face recognition – correlations

	GFMT	GFMT: Same	GFMT: Diff	CFMT	CFMT: Identical	CFMT: Novel	CFMT: Novel + Noise
GFMT	-	.676**	.680**	.579**	.427**	.519**	.506**
GFMT: Same		-	-0.081	.423**	441**	.375**	.337**
GFMT: Diff			-	.362**	0.139	.329**	.349**
CFMT				-	.540**	.914**	.913**
CFMT: Identical					-	.393**	.395**
CFMT: Novel						-	.698**
CFMT: Novel + Noise							-

** Correlation is significant at the 0.01 level (2-tailed).

Power analysis in G Power 3.1 indicated that with a sample size of 84 participants, only r values greater than or equal to 0.342 have the required power for the null hypothesis to be confidently rejected. On this basis, all of the positive associations found can be confidently accepted except the relationship between same face pair processing and recognising novel faces shown with noise. With an r value of .337 this relationship must be taken with caution. Overall, same face pairs show a decrease in association across novel faces and novel faces shown with noise whereas different face pairs show the opposite, suggesting again, that different qualitative processes may be used.

Discussion

This experiment provides a useful confirmation of the findings from Experiment 8. The results demonstrate again that same and different face matching are dissociated, unlike the subtasks of the unfamiliar face recognition task, and that they associate with the subtasks of the unfamiliar face recognition task in quite separate and specific ways. The decision times recorded support these associations and dissociations.

Taken together, these findings suggest that there is both a qualitative difference between processing identical faces and novel faces (with and without noise) and a qualitative difference between same and different face matching. Neither finding is new. Same and different face matching has been shown to dissociate across a range of tasks (Megreya & Burton, 2007). Much earlier research identified that the processes of recognising identical pictures of faces (image recognition) and face recognition as significantly different (Bruce, 1982; Bruce & Young, 1986).   Participants were much slower and less accurate when recognising faces rather than images. This relationship was also replicated in this experiment with participants achieving lower accuracy for novel faces when compared with identical faces and slower decision times when comparing novel faces with identical faces.

It was the change in response time that was key to Bruce (1982) identifying mechanisms had changed between the recognition tasks. This may also be supported by the change in the strength of association shown here. Recognition performance for identical faces is positively associated with recognition performance for novel faces and novel faces shown with noise, but this association is much weaker than the association between novel faces and novel faces shown with noise. The strength of these associations indicate a change in processes has occurred but is not as strong an indicator as dissociation. The closer relationship between novel faces and novel faces shown with noise is borne out in terms of accuracy and latency. Where performance for novel faces plus noise was significantly less accurate than performance for novel faces there was no significant difference between the two in terms of latency. Novel faces plus noise were more difficult to recognise but still used the same underlying processes used to process novel faces.

Face matching accuracy and latency also provide further evidence regarding the use of qualitatively different processes by same and different face matching. Performance across same and different face matching in terms of accuracy is not significantly different, however in terms of decision times there is a significant difference. On average, participants took longer to reach a decision about different face pairs than same face pairs. This suggests that like image and face recognition, different face matching uses qualitatively different processes than same face matching and that these processes are more effortful.

Faster response times for same judgements are not unusual in decision making. Krueger (1978) notes that they are typical. However, they might not be expected, since identifying that two patterns are different only requires one differing aspect to be identified (Nickerson, 1965). Our expectations then would suggest that identifying two images as different should be faster than identifying that they are the same; since establishing they are the same would mean checking all aspects before coming to a conclusion. Krueger (1978) identifies that faster response times may be understood in terms of in terms of a noisy operator theory. As operators with internal noise, viewers are more likely to make invalid featural mismatches than matches. To ensure high levels of accuracy, a viewer is forced to recheck mismatches which results in a longer response time. Krueger (1978) states that this checking is carried out a configural level, giving the example of the stimuli used in his study whereby strings of letters are compared by the configuration of all the letters rather than each letter individually. This would suggest that different pair matching for objects is associated with configural processing resulting in longer decision times.

In unfamiliar face matching tasks where featural and configural changes have been made, participants have been shown to be both faster and more accurate at identifying featural rather than configural changes (Mercure et al., 2008). Mercure et al., (2008) also analysed the impact of this task on ERP components and found that the P2 component was more sensitive to faces with configural changes than faces with featural changes. The authors point out that the P2 component has been associated with the effects of visual cortical feedback, so the stronger activation of the P2 combined with the longer response times might suggest that configural changes may require greater visual analysis. In contrast, fMRI research has shown that the occipital face area (OFA) is activated preferentially for single features, and eye features particularly, whereas the fusiform face area (FFA) and the lateral occipital cortex (LO) showed no such preference. This supports earlier research that suggests the OFA is involved in early feature based analysis whereas the FFA is more concerned with processing the whole, although this may also include the integration of low-level features (Arcurio, Gold & James, 2012). These findings would suggest that the longer decision times identified here are the result of not only a change in strategy but also a move away from featural analysis that comes earlier and is faster, to configural analysis which takes longer.

The findings of Experiment 8 and 9 are novel in that they identify how the dissociated subtasks within the GFMT relate to these qualitative changes in face recognition and how these might be supported by relationships with more general tasks. These relationships are not causal but may predict performance in unfamiliar face matching tasks, particularly when different processing strategies are isolated. In experiment 10, the importance of featural processing when matching same face pairs is tested by comparing participants’ performance across same and different face matching pairs when full faces and the eyes only are shown.

5.3 Experiment 10 – Isolating featural processing

Introduction

Findings from Experiment 8 and 9 support earlier research that suggests that same and different face matching are dissociated and may be carried out in a qualitatively different way (Megreya & Burton, 2007). Correlational and regression analysis from Experiment 8 also suggest that same face matching is associated with and predicted by local processing and different face matching is associated with and predicted by space perception. Within the face matching context these mechanisms may translate as featural processing and configural processing, in particular the second order relations or spatial distances between the features.

These predictions have yet to be tested. However, there are a number of ways stimuli within an unfamiliar face matching task could be altered in order to isolate or prioritise specific ways of processing. Using a feature only set of faces could allow the effects of isolating featural processing to be tested. Since same face matching is associated with featural processing it could be expected that accuracy might be reduced by showing a feature only since the variation between two individuals may not lie in a single feature. However, the reduction in accuracy for same face matching would not be as great as the reduction in accuracy for different face matching. Since different face matching is not associated with featural processing it might be expected that both accuracy and latency would be compromised to a much greater degree if only a single feature was available for comparison. The relationships between same and different face matching might also change. If same and different face matching are dissociated due to their use of different mechanisms, and if both types of matching are forced to use local or featural processing then the dissociation may be resolved.

In the following experiment, the effect of viewing faces as full faces and feature only on same and different face matching is investigated. The eyes were selected for the feature only condition as participants have been shown to be more accurate making eye judgements than nose judgements or mouth judgements when carrying out unfamiliar recognition and unfamiliar simultaneous matching tasks (Tanaka & Farah, 1993; Sergent, 1984). This has been supported by research that suggests that the occipital face area (OFA) is activated for single features and eye features particularly (Arcurio et al., 2012). Viewers appear to preferentially attend to the eyes of faces and this preference only increases with practice, whereas individuals with prosopagnosia do not show this pattern of behaviour (O'Donnell & Bruce, 2001; Vinette, Gosselin & Schyns, 2004; Caldara, Schyns, Mayer, Smith, Gosselin & Rossion, 2005). These findings suggest that the eyes are particularly important in identifying and recognising individuals; as a result, faces presented as eyes only should produce the highest levels of accuracy in an unfamiliar feature matching task.

Method

Participants

Participants were 40 (5 male) students and staff of the University of York, who were over 18, British citizens, and reported normal or corrected to normal vision (mean age = 19.73, range = 18-28). Participants were reimbursed a small fee or course credit for their time.

Materials

96 pairs of unfamiliar faces were used as stimuli from the GFMT long form (Burton et al., 2010) with equal numbers of same and different trials. Feature only trials were created by masking the original face so that only the eyes were seen and to ensure the eyes remained in the original position and the original size.

Procedure

This experiment was conducted in the lab with the experimenter present throughout. Each participant saw 48 face pairs in the Full Face condition and a further 48 face pairs in the Feature Only condition. Presentation of the conditions was blocked. Faces were counterbalanced across the experiment such that each face occurs equally often in each condition and the presentation order was randomised. Figure 5.2 gives an example of faces in each condition. On presentation of each pair of faces participants were asked, on-screen 'Are the images of the same individual?' and selected responses 'Same' or Different' with a mouse. The task was self-paced, with no time limit for responses, and took around 20 minutes to complete.

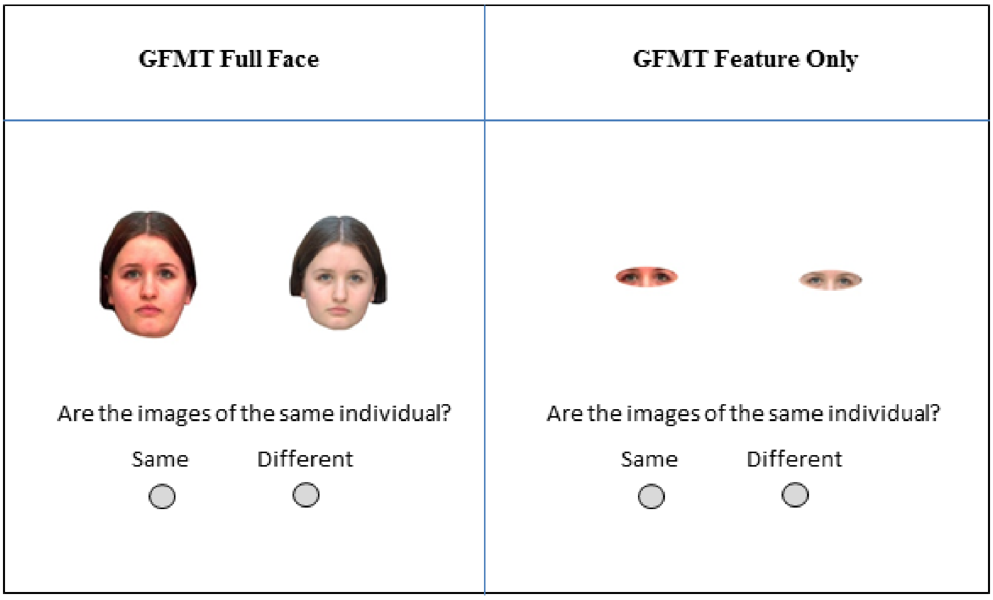

Figure 5.2. Example stimuli from the GFMT 96 Full Face and Feature Only conditions.

Results

Face Matching Accuracy

Mean face matching accuracy is shown in Figure 5.3, with data broken down into same and different trials. A 2 (image type: Full Face/Feature Only) x 2 (match type: same/different) within subjects ANOVA showed a significant main effect of image type $F(1,39) = 236.29$, $p <.001$, $\eta_{\rho}^2 = .86$. However, neither the match type $F(1,39) = 1.21$ $p >.05$, $\eta_{\rho}^2 = .03$ or interaction was significant $F(1,39) = .19$, $p >.05$, $\eta_{\rho}^2 = .00$. Presenting faces as Feature Only leads to a reduction in accuracy overall but in no difference between same and different face pairs

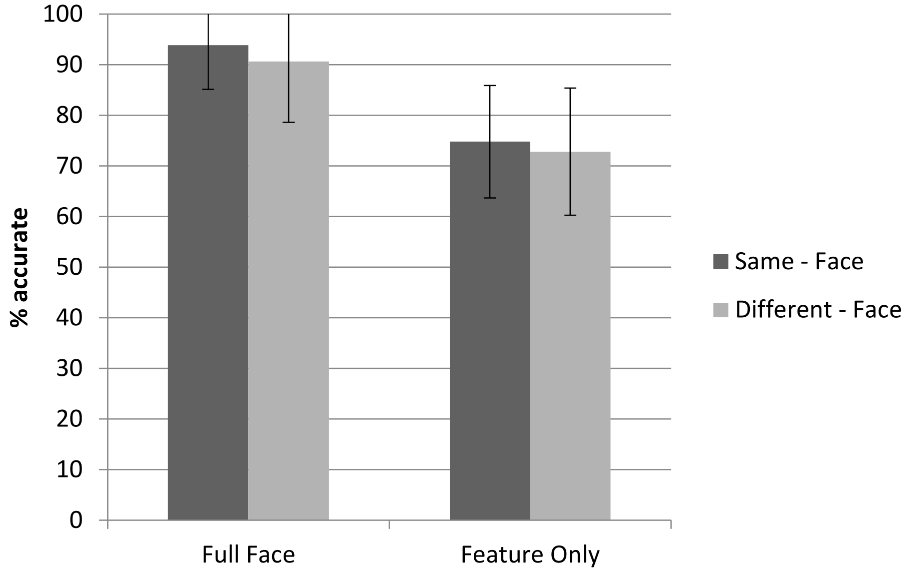

Figure 5.3. Mean face matching accuracy across conditions. Note error bars represent standard deviation.

Face Matching Decision Times

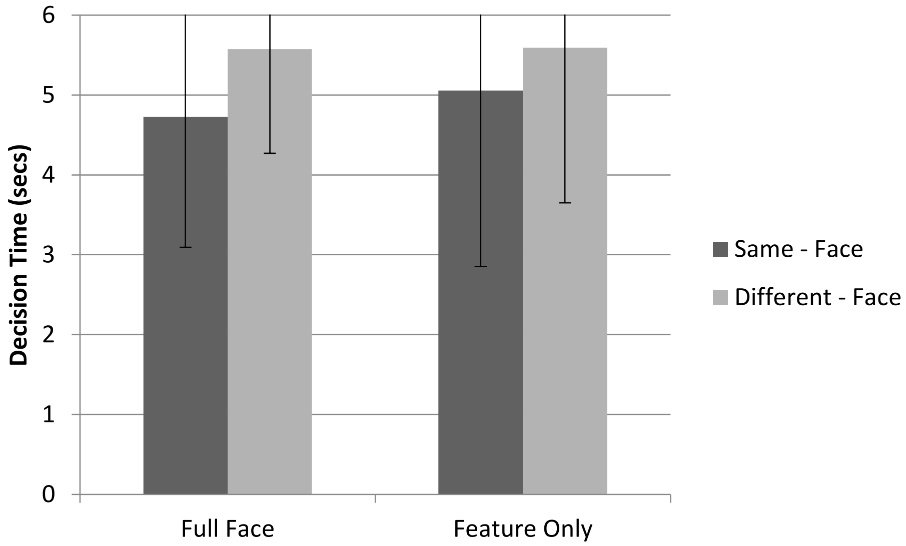

Figure 5.4. Mean face matching decision time across conditions. Note error bars represent standard deviation.

Decision times are reported here across all response types - correct and incorrect. Mean face matching decision time is shown in Figure 5.4, with data broken down into same and different trials. A 2 (image type: Full Face/Feature Only) x 2 (match type: same/different) within subjects ANOVA showed no significant main effect of image type $F(1,39) = 0.59$ $p > .05$, $\eta_\rho^2 = .01$ or interaction $F(1,39) = 1.38$, $p > .05$, $\eta_\rho^2 = .03$. However, there was a significant effect of match type $F(1,39) = 11.26$, $p < .005$, $\eta_\rho^2 = .22$.

Participants took longer to reach a decision about different face pairs than same face pairs whether the faces were shown as Full Face or Feature Only.

Correlations

Table 5.4 provides a correlation matrix for performance on same and different face pairs and their image type. As expected, performance on same and different face pairs was dissociated in the Full Face condition suggesting different qualitative processes may be used, $r(40) = -.155$, $p > .05$. Performance on same and different face pairs was also dissociated in the Feature Only condition $r(40) = -.212$, $p > .05$. This does not support the hypothesis that face matching using Feature Only creates a greater proportion of shared processes (local processing) between same and different face pair processing. Interestingly, performance for both same face pairs and different face pairs was significantly and positively associated in both the Full Face and Feature Only conditions ($r(40) = .460$, $p < .001$, and $r(40) = .433$, $p < .001$, respectively). This suggests that whether the face pairs are Feature Only or Full Face, same and different face pair processing remain related between both conditions.

Table 5.4. Face matching accuracy for faces shown as full faces or feature only – correlations.

	Full Face Same	Full Face Different	Feature Only Same	Feature Only Different
Full Face Same	-	-.155	.460**	-.020
Full Face Different		-	-.212	.433**
Feature Only Same			-	-.281
Feature Only Different				-

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Discussion

This study did not show the expected effects. Unfamiliar face matching overall is compromised by showing a feature only – in this case the eyes. However, there is no difference in the reduction in performance between same and different processing – this is true for both accuracy and decision time. This suggests that same and different face processing use similar resources whether processing Full Faces and a Feature Only. This is supported by the dissociation between same and different face matching holding whether faces are viewed in full or as a feature only. The relationships between faces viewed in full or as a feature only also hold whether same or different face pairs are matched. A dissociation here would also suggest that the strategies being used to match single features were different. There may be two explanations for these findings. First, the prediction may have been incorrect; it may be the case that same and different face matching use dissociated processes but this is not due to using local and configural processing differently. Secondly, the Feature Only condition may not have isolated featural processing and allowed viewers to apply configural strategies. Participants discussed using the distance between the eyes and the eyes and the eyebrows to make their decisions and these strategies have proved successful in unfamiliar face matching tasks (Hosie, Ellis, Haig, 1988; Sekuler, Gaspar, Gold & Bennett, 2004; Sadr, Jarudi, & Sinha, 2003). An experiment that isolates configural processing may help in identifying which of these explanations is most accurate. If viewers are forced to use a configural strategy and same face matching is compromised to the same degree as different face matching then this would suggest that their dissociation is not due to their respective reliance on featural and configural processing.

The following experiments aim to provide further explanation by taking this approach and isolating configural processing.

5.4 Experiment 11 – Isolating Configural Processing

Introduction

Experiment 8 and 9 demonstrated that same and different unfamiliar face matching are dissociated and associate differently with the CFMT subtasks. Experiment 8 also showed that same face matching is associated with local processing whereas different face matching was associated with configural processing. Experiment 10 aimed to test these associations by requiring participants to match faces using a feature only. It was predicted that different face matching would be compromised to a much greater degree than same face matching since participants would be forced to use local processing to match different face pairs. It was also predicted that same and different face matching would also become associated if participants viewed single features only since participants would be forced to use local processing for both types of matching. However, these predictions were not met. Same and different face matching were equally compromised by the feature only condition and performance across both remained dissociated. These results may have been due to a failure in the stimuli to isolate featural processing or by the dissociation between same and different face matching being due to factors other than a differential reliance on local and configural processing. By requiring viewers to use configural processing the reliance on configural processing for different face matching could be tested.

Duchaine and Nakayama (2006a) introduced noise to the CFMT to make the task more challenging and produce an increased reliance on configural processing (McKone, Martini & Nakayama, 2001). Adding noise to the face matching task might then require viewers to become more reliant on configural processing; a strategy that would compromise performance in same face matching to a much greater extent than performance in different face matching. It may also be reasonable to expect that the addition of noise may remove the speed advantage for same face matching if the speed advantage is due to featural processing. In terms of associations, it could be assumed that different face matching would be associated across faces viewed with and without noise since different face matching would primarily use configural processing in both conditions. However, same face matching would not be associated across conditions since face matching without noise primarily uses featural processing and would be forced to change to using primarily configural processing. This change in the underlying mechanism may also result in same and different face matching being associated when noise is applied.

Method

Participants

Participants were 48 (6 male) students and staff of the University of York, who were over 18, British citizens, and reported normal or corrected to normal vision (mean age = 19.25, range = 18-22). Participants were reimbursed a small fee or course credit for their time.

Materials

96 pairs of unfamiliar faces were used as stimuli from the GFMT long form (Burton et al., 2010) with equal numbers of same and different trials. Random heavy noise was added using the ‘Hurl’ filter in Gimp 2.8.14. 50% of the pixels were changed to random colour.

Procedure

This experiment was conducted in the lab with the experimenter present throughout. Each participant saw 48 face pairs in the No Noise condition and a further 48 face pairs in the With Noise condition. Presentation of the conditions was blocked. Faces were counterbalanced across the experiment such that each face occurs equally often in each condition and the presentation order was randomised. Figure 5.5 gives an example of faces in each condition. On presentation of each pair of faces participants were asked, on-screen 'Are the Images of the same individual?' and selected responses 'Same' or Different' with a mouse. The task was self-paced, with no time limit for responses, and typically took around 20 minutes to complete.

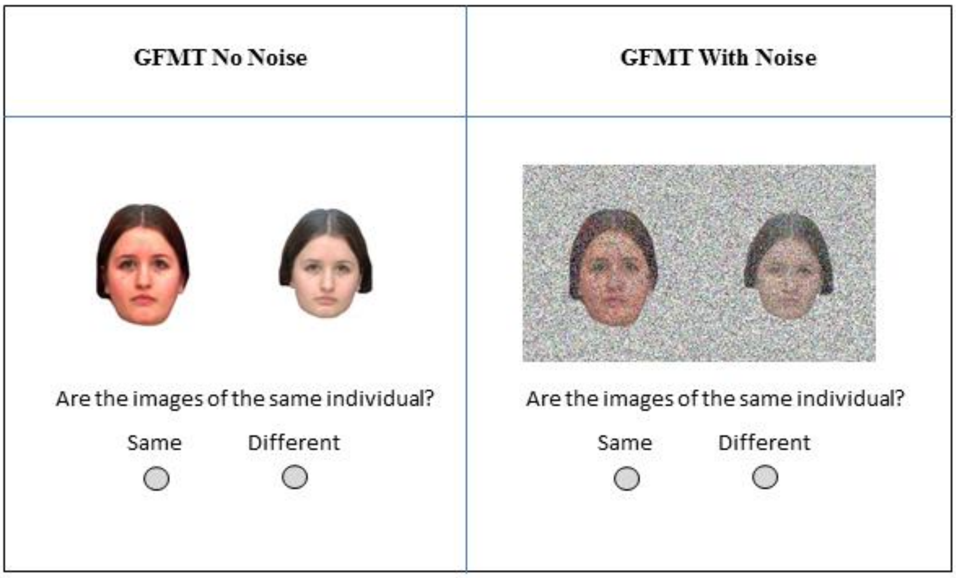

Figure 5.5. Example stimuli from the GFMT without noise and with noise applied

Results

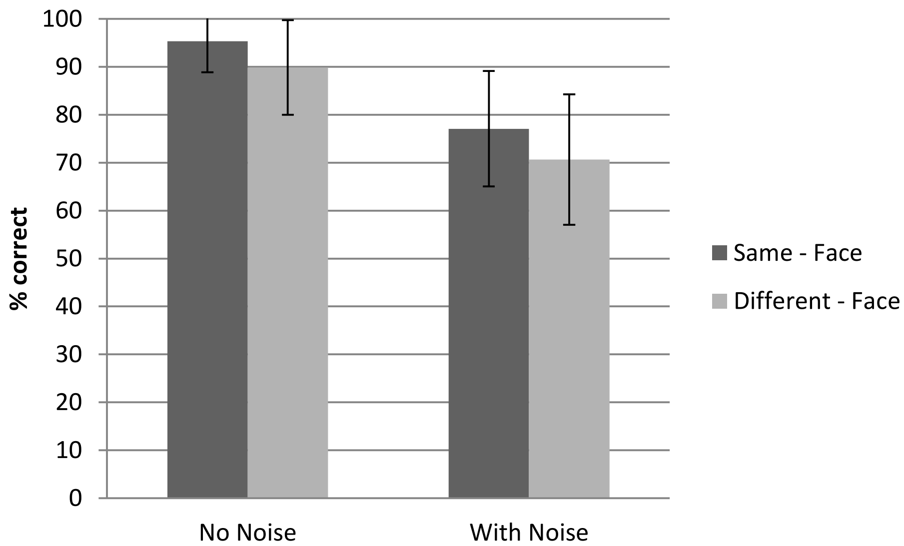

Figure 5.6. Mean face matching accuracy across conditions. Note error bars represent standard deviation.

Face Matching Accuracy

Mean face matching accuracy is shown in Figure 5.6, with data broken down into same and different trials. A 2 (clarity type: no noise/with noise) x 2 (match type: same/different) within subjects ANOVA showed a significant main effect of clarity *F(*1,47) = 291.97, *p* <.001, η_p^2 = .86 and a significant main effect of match type *F(*1,47) = 8.74, *p* <.005, η_p^2 = .02 however the interaction was not significant *F(*1,47) = .12, *p* >.05, η_p^2 = .00. Presenting faces with noise leads to a reduction in accuracy overall. There is a significant difference between accuracy for same and different face pairs, but this does not significantly change whether the faces are shown with noise or without.

Face Matching Decision Times

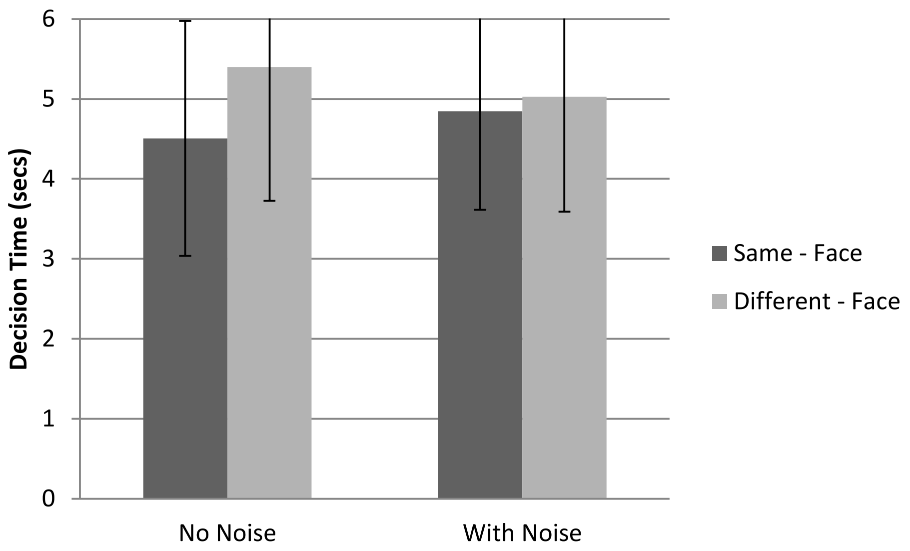

Figure 5.7. Mean face matching decision time across conditions. Note the error bars represent standard deviation.

Decision times are reported here across all response types - correct and incorrect. Mean face matching decision time is shown in Figure 5.7, with data broken down into same and different trials. A 2 (clarity type: no noise/noise) x 2 (match type: same/different) within subjects ANOVA showed no significant main effect of clarity type *F(*1,47) = 0.01, *p* > .05, η_p^2 = .00 however there was a significant main effect of match type *F(*1,47) = 17.59, *p* < .001, η_p^2 = .27 and a significant interaction *F(*1,47) = 12.31, *p* < .005, η_p^2 = .21. Simple main effects analyses showed that match type was significantly different when faces were shown without noise *F(*1,94) = 29.88, *p* <.001, η_p^2 = .24 but not when faces were shown with noise *F(*1,94) = 1.19, *p* >.05, η_p^2 = .01. Participants were faster at identifying same face pairs than different face pairs, but this was only the case when faces were shown without noise.

Correlations

Table 5.5. Face matching accuracy for faces shown with and without noise – correlations.

	No noise Same	No noise Different	Noise Same	Noise Different
No noise Same	-	.119	.206	-.026
No noise Different		-	-.126	.528**
Noise Same			-	-.349*
Noise Different				-

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Table 5.5 provides a correlation matrix for performance on same and different face pairs and their clarity type. As expected performance on same and different face pairs was dissociated in the no noise condition suggesting different qualitative processes may be used, $r(48) = .119$, $p >.05$. However, performance on same and different face pairs was negatively associated in the noise condition $r(48) = -.349$, $p < .05$ this supports the hypothesis that face matching using noise creates a greater proportion of shared processes between same and different face pair processing. Interestingly, performance for same face pairs was dissociated across the no noise and noise conditions $r(48) = .206$, $p > .05$, however, performance for different face pairs was positively associated and $r(48) = .528$, $p < .01$. This suggests that different face pairs may share a large degree of processing across the noise conditions, whereas same face pairs do not.

Discussion

This study shows a number of interesting effects. Performance in same face matching was compromised by the addition of noise to the face pairs. However, performance in different face matching was also compromised and to the same degree. This was not the expected effect of adding noise. However, all of the other predictions were met. The speed advantage for same face matching was lost and the relationships between the different measures changed as expected. Without noise, same and different face matching were dissociated, when noise was added they were associated, suggesting a change in mechanisms underlying these measures. Since different face matching remains associated across the No Noise and With Noise conditions, the underlying mechanisms very likely remain the same in both conditions. However, same face matching was not associated across the No Noise and With Noise conditions, which suggests the underlying mechanism changes when noise is added. These findings, and the absence of the speed advantage for same pair face matching, would suggest that featural processing has been abandoned in favour of configural processing when matching same face pairs shown with noise.

There may be a different explanation to these results. As in experiment 10, the attempt to isolate a particular type of configural processing may have failed. It is difficult to identify the levels of noise required to affect processing. The addition of noise may have compromised both featural and configural processing and so both same and different face matching were compromised. Mckone et al., (2001) demonstrated that adding noise to images reduces higher spatial frequencies and the ability to perceive fine detail such as distances between features. Participants were only able to categorically discriminate between two learned faces when the full face was shown and not when features only were shown. However, as Young et al., (1987) and Hole (1994) noted the contribution of configural and featural processing may well depend on the nature of the task and noise may disrupt an image differently when being processed for similarity rather than identity.

Studies regarding the effects of reducing high spatial frequencies on configural processing have provided mixed results. Goffaux, Hault, Michel, Vuong and Rossion (2005) found that higher spatial frequencies supported the extraction of local features and lower spatial frequencies supported the extraction of configural cues in unfamiliar face matching. However, Boutet, Collin and Faubert (2003) found that there was no difference in the levels of accuracy provided by featural or configural processing for unfamiliar face matching or recognition when faces were shown as low, medium or high spatial frequencies. Blurring has also been used as a method to reduce high spatial frequencies in face identity processing and with mixed results. Collishaw and Hole (2000) found that blurring reduced accuracy for recognising both familiar and unfamiliar faces but that when blurred faces were also scrambled accuracy was reduced to an even greater degree. Since scrambling faces was thought to reduce featural processing the additional reduction in accuracy for faces that had been both blurred and scrambled was identified as a combination of separate and dissociated processes. A later face matching study where blur was added systematically to faces where configural and featural changes were matched found no difference in the decline in accuracy for detecting both types of changes as the image quality degraded (Gilad-Gutnick, Yovel & Sinha, 2012).

While the addition of noise may well have compromised both featural and configural processing, it would not be true to say that the effects were shared equally across same and different face matching. Both were compromised equally in terms of accuracy, but the speed advantage for same face matching was lost. Same and different face matching became associated when noise was added which suggest that the processes underlying the two became shared. Same face matching was dissociated between faces shown with noise and without noise, whereas different face matching was associated. This suggests that it was the mechanisms underlying same face pair matching that changed. This would add weight to the theory that both featural and configural processing are compromised by the addition of noise. Once featural processing is compromised, same face matching may switch or default to configural processing which is also compromised. Both experiments 10 and 11 have sought to investigate the differences between same and different face matching in unfamiliar face matching using a type of manipulation designed to effect only one type of processing. However, there is evidence in both experiments that both types of processing have been manipulated. Experiment 12 uses a final type of manipulation shown to affect configural processing to a greater degree than featural processing – inversion (Freire et al., 2000; Mondloch et al., 2002)

5.5 Experiment 12 – Disrupting configural processing

Introduction

Experiment 8 and 9 demonstrated that same and different face matching were dissociated and related to measures of local processing and configural processing in different ways. Experiments 10 and 11 aimed to investigate these relationships by masking faces and adding noise in order to isolate featural and configural processing. However, it was not clear whether the masking and addition of noise were successful in isolating different types of processing; although experiment 11 indicated that adding noise might have resulted in a shift towards configural processing for same face pair matching. In this experiment, a different approach is taken, investigating the consequences of disrupting configural processing for same and different face matching.

Inversion has been shown to disrupt configural processing to a much greater degree than featural processing (Freire et al., 2000; Mondloch et al., 2002). Megreya and Burton

(2006) investigated the effects of inversion on unfamiliar face matching using a paired items task. They found that hits, or same pairs identified correctly, were significantly lower when inverted but that the difference was not large: 24.5/30 vs. 22.3/30; $t(29)$ 3.1, p <.01. This was surprising as Yin(1969) had shown that when faces were inverted, the reduction in recognition performance was both significant when compared to recognition performance when the faces upright but the reduction was also large. The reduction in performance for faces was significantly greater than the reduction for the objects when stimuli were recognised upright and inverted. On this basis, it might not be unreasonable to expect a larger reduction in performance for Megreya & Burton's (2006) unfamiliar face matching task. Megreya & Burton (2006) also found that the hits and accuracy scores for the inverted version of this task were highly correlated with an upright unfamiliar face matching task using an array, but not with an upright familiar face recognition task. Taken together these findings led them to make the assertion that 'unfamiliar faces are not faces'.

However, the deficit in accuracy in Megreya & Burton's (2006) face matching task for different scores was not reported. The false positives scores (same face pairs identified incorrectly) for the face matching upright and inverted were provided [6.6/30 vs.11.5/30; $t(29)$ 6.4, p <.01] and enables the scores for different face matching to be calculated (23.4./30 vs. 18.5/30). These suggest that accuracy for matching different face pairs may also be significantly lower when the faces are inverted and the reduction in performance significantly lower than the reduction in performance for same face pair matching. This would potentially support the findings in Experiments 8 and 9 that suggests a stronger relationship between different face pair matching and configural processing than same face pair and configural processing.

Megreya and Burton's (2006) findings build on other results that were not investigated but published by Farah et al., (1998). Farah et al. found that inversion reduced accuracy much more for different-face matching than for same-face matching. If this can be shown here, then it may provide further evidence regarding the qualitative differences between same and different face matching. On this basis, it would be predicted that when face pairs were inverted, accuracy for different face matching would be significantly lower than accuracy for same face matching. Decision times would increase to a larger degree for different face matching than for same face matching when the faces were inverted to reflect the greater effort required when configural processing is disrupted. It would also be expected that same and different face matching would become associated during inversion since viewers would not be able to use configural processing relying instead on shared face processing mechanisms also used by same face matching. In this case, same face matching may remain associated across upright and inverted faces, however, different face matching may not. In the following experiment, we investigate the effect of inversion on same and different face matching.

Method

Participants

Participants were 41 (11 male) British students and staff of the University of York, who all reported normal or corrected to normal vision (mean age = 20.98, range = 18-51). Participants were reimbursed a small fee or course credit for their time.

Materials

96 pairs of unfamiliar faces were used as stimuli from the GFMT long form (Burton et al., 2010) with equal numbers of same and different trials.

Procedure

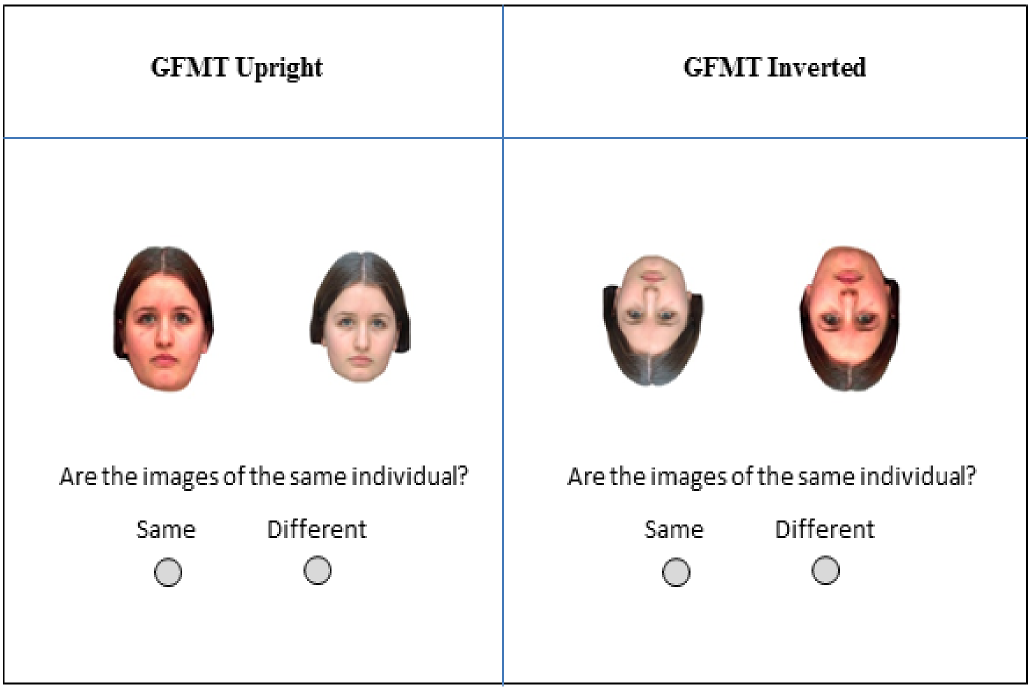

Figure 5.8. Example stimuli from the GFMT 96 upright and inverted.

This experiment was conducted in the lab with the experimenter present throughout. Each participant saw 48 face pairs in the upright condition and a further 48 face pairs in the inverted condition. Presentation of the conditions was blocked. Faces were counterbalanced across the experiment such that each face occurs equally often in each condition and the presentation order was randomised. Figure 5.8 gives an example of faces in each condition. On presentation of each pair of faces participants were asked, on-screen 'Are the images of the same individual?' and selected responses 'Same' or Different' with a mouse. The task was self-paced, with no time limit for responses, and took around 20 minutes to complete.

Results

Face Matching Accuracy

Mean face matching accuracy is shown in Figure 5.9, with data broken down into same and different trials. A 2 (orientation type: upright/inverted) x 2 (match type: same/different) within subjects ANOVA showed a significant main effect of orientation type $F(1,40) = 150.44$, $p < .001$, $\eta_\rho^2 = .79$ and match type $F(1,40) = 34.29$ p < .001, $\eta_\rho^2 = .46$ which was qualified by a significant interaction $F(1,40) = 59.19$, $p < .001$, $\eta_\rho^2 = .60$. Simple main effects analyses showed that orientation type affected both same and different match items $F(1,80) = 6.57$, $p < .01$, $\eta_\rho^2 = .08$; and $F(1, 80) = 194.45$, $p < .001$, $\eta_\rho^2 = .71$, respectively. Inverting the stimuli leads to a reduction in accuracy for both same and different face pairs, but much more so for different face pairs.

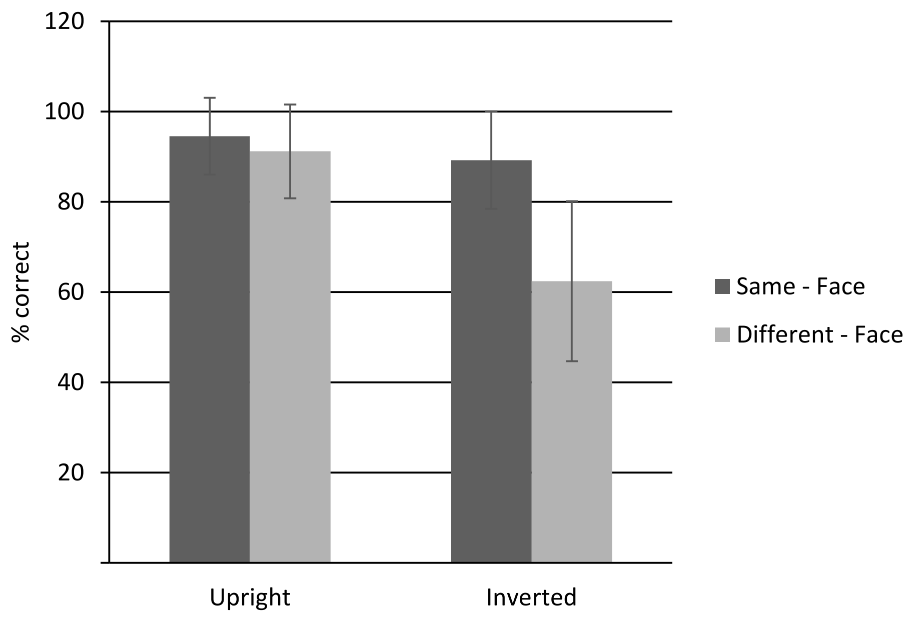

Figure 5.9. Mean face matching accuracy across conditions. Note error bars represent standard deviation.

Face Matching Decision Times

Decision times are reported here across all response types - correct and incorrect. Mean face matching decision time is shown in Figure 6.0, with data broken down into same and different trials. A 2 (orientation type: upright/inverted) x 2 (match type: same/different) within subjects ANOVA showed a significant main effect of orientation type $F(1,40) = 18.53$, $p < .001$, $\eta_\rho^2 = .32$ and match type $F(1,40) = 23.15$ $p < .001$, $\eta_\rho^2 = .37$ with no significant interaction $F(1,40) = 3.27$, $p > .05$, $\eta_\rho^2 = .08$. On average, participants took longer to reach a decision about different face pairs than same face pairs when the faces were both upright and inverted, but both decision times were significantly longer when the faces were inverted.

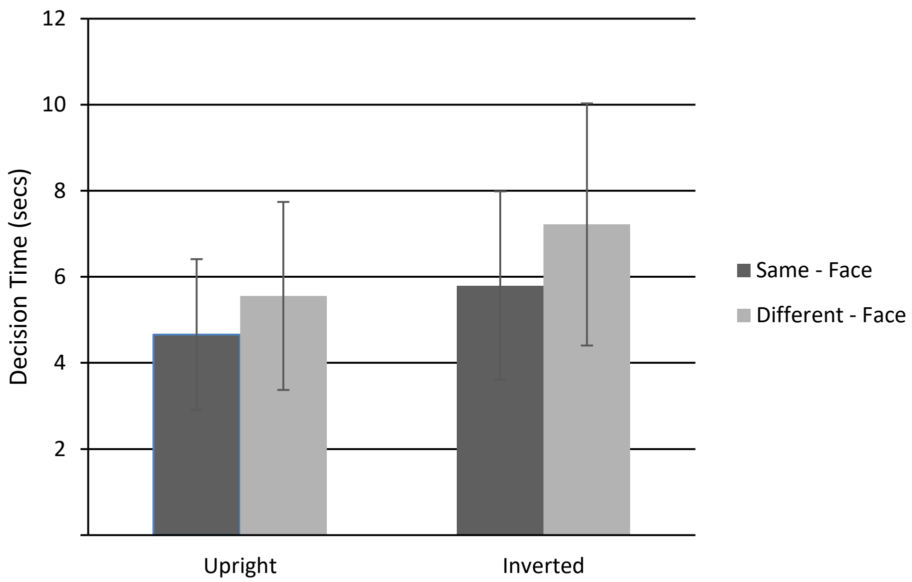

Figure 6.0. Mean face matching decision time across conditions. Note error bars represent standard deviation.

Correlations

Table 5.6 provides a correlation matrix for performance on same and different face pairs and their orientation. As expected, performance on same and different face pairs was dissociated in the upright condition suggesting different qualitative processes may be used. As predicted, performance on same and different face pairs was significantly and negatively associated in the inverted condition $r(41) = -.401$, $p < .05$. This supports the hypothesis that inversion creates a greater proportion of shared processes between same and different face pair processing. Interestingly, performance for both same face pairs and different face pairs was significantly and positively associated in both the upright and inverted conditions (*r*(41) = .477, *p* < .01, and *r*(41) = .465, *p* < .01, respectively). This suggests that whether the face pairs are inverted and upright, same and different face pair processing remain related across both conditions

Table 5.6. Face matching accuracy for upright and inverted faces – correlations.

	Upright Same	Upright Different	Inverted Same	Inverted Different
Upright Same	-	.130	.477**	-.182
Upright Different		-	-.162	.465**
Inverted Same			-	-.401*
Inverted Different				-

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Discussion

This study, like those before, shows a number of interesting effects. Firstly, unfamiliar face matching is compromised by inversion; this has been shown to be the case for both same and different face pairs and in terms of accuracy and decision times. Secondly, accuracy and decision times for different face matching is compromised to a much greater degree supporting the hypothesis that same and face different pairs are processed qualitatively differently. Different face pair processing being more reliant on configural processing than same face pair processing.

The dissociation between performance on the same and different face pairs in the upright face matching task also suggests that these tasks are carried out using qualitatively different processes. Interestingly when inverted, the negative association between the same and different face pairs suggests that these tasks are now using a greater proportion of qualitatively similar processes or a shared process that identifies a face pair as either 'Same' or 'Different'. This supports the hypothesis that configural processing is reduced when faces are inverted and therefore participants have to utilise more of the processes used for same face pair processing. However, the positive association between performance for same face pairs when upright and inverted and for different face pairs when upright and inverted also suggests that a proportion of those qualitatively different processes are still retained for same face matching and different face matching. This suggests that configural processing may not be entirely abandoned (though it is less successful) when the face pairs are inverted but that it is supported to a much greater extent by the processes used in same face pair processing.

The question here then is what are those processes used by same face pair processing? The reduction in accuracy for same face pair processing could be due to same face pair processing also using configural processing but to a lesser degree or that featural processing is also compromised by inversion. The literature is not united on the effect of inversion on featural and configural processing. Some researchers find that configural processing is affected to a much greater degree than featural processing by inversion, and others find that both are affected equally (Freire et al., 2002; Mondloch et al., 2002). Riesenhuber et al., 2004; Yovel & Kanwisher, 2004). It was not the purpose of this experiment to investigate this question, but it may well be that accuracy for same face matching is compromised because of a reliance on featural processing which is compromised and configural processing that is compromised. Certainly experiments 8 and 9 suggest that same face pair matching is associated with both local processing used for objects and types of face recognition requiring greater configural processing. What is much clearer is that different face matching is disproportionately compromised by inversion and this may well be due to a greater reliance on configural processing. This supports the hypothesis that same and different face processing are qualitatively different.

5.6 General Discussion

It was the aim of this chapter to replicate and extend the findings from Chapter 4. The battery of measures used in Chapter 4 demonstrated that same and different face matching were dissociated, as had been shown previously (Megreya & Burton, 2007), and differently associated with subtasks from the CFMT and more general measures. Same face matching showed stronger associations with the identical and novel subtasks of the CFMT than the novel + noise subtask and was associated with local processing. Different face matching showed stronger associations with the novel + noise and novel subtasks of the CFMT and was associated with space perception. The first objective was to test the relationships between the GFMT and the CFMT and extend these findings using latency measures. Experiment 9 replicated the relationships between the GFMT and CFMT found in Chapter 4, providing greater confidence in these findings. The latency measures provided further insight into the processes underlying the tasks. Decision times for the CFMT showed that recognising novel views of faces and novel views of faces shown with noise took significantly longer than recognising identical images, suggesting a change in the underlying mechanisms (Bruce, 1982). Decision times for the GFMT demonstrated that identifying different face pairs took significantly longer than identifying same face pairs; suggesting again a change of mechanism that was supported by the dissociation. Research in both decision-making and face processing has shown faster processing to be associated with featural processing and slower processing to be associated with configural processing (Krueger, 1978; Arcurio et al., 2012; Mercure et al., 2008). These findings add greater weight to the theory that same face matching is associated with featural processing and different face matching is associated with configural processing.

While Experiment 9 provides greater confidence and support for the relationships identified in Experiment 8, it does not provide evidence that same or different face matching is achieved as a result of featural or configural processing. The second objective of this chapter was to investigate a causal relationship between the different types of processing and unfamiliar face matching. This investigation was carried out using the GFMT stimuli and manipulating the images to isolate or compromise different types of processing. In Experiment 10, accuracy for matching full faces and feature only was examined across same and different face matching. The aim of this experiment was to isolate featural processing using stimuli showing only the eyes. It was predicted that different face matching would be compromised to a much greater degree than same face matching in terms of accuracy and latency in the feature only condition. It was also predicted that in the feature only condition same and different face matching might be associated since different face matching would not be able to use configural processing and have to switch to more shared processes with same face matching. However, these predictions were not met. Both same and different face matching were compromised to the same degree in the feature only condition and they remained dissociated across both conditions. This suggested that same and different face matching continued to use featural and configural processing differentially for the feature only condition. Given that the pair of eyes may have provided enough information for participants to continue to use configural processing e.g. distance between the eyes, this would be a very feasible explanation. (Hosie et al., 1988)

Since isolating featural processing proved difficult, a different approach was taken. Experiment 11 aimed to isolate configural processing using random heavy noise. It was predicted that same face matching would be compromised to a much greater degree than different face matching in terms of accuracy and latency when faces were shown with noise. It was also predicted that when faces were shown with noise same and different face matching might be associated since same face matching would not be able to use featural processing and have to switch to configural processing. However, accuracy for same and different face matching was compromised to the same degree when noise was added. This might suggest that same and different face matching made use of configural processing to the same degree. However, decison times for same face matching and different face matching were no longer significantly different suggesting that the featural speed advantage was lost when noise was added. In addition, same and different face matching became associated when noise was added and it was same face matching that was dissociated when comparing faces shown with and without noise. The changing relationships, like the change in latency, suggests that same face matching used a different mechanism when noise was applied. These results were not as clear as hoped. It may well be that the noise compromised both featural and configural processing supporting studies that have shown that removing high level frequencies can affect configural and featural processes equally (Boutet et al., 2003; Gilad-Gutnick et al., 2012).

Rakover's (2002) finding that manipulating configural and featural processes separately is difficult is clearly demonstrated in both Experiment 10 and 11. Even those studies where researchers aim to recruit different types of processing by making featural or configural changes to the images (e.g. replacing features or changing the distance between features (Riesenhuber, 2004) it is difficult to be sure that only one type of change that has been made. Replacing a feature inevitably changes the spatial relations between that feature and surrounding features. Making changes to the stimuli that might indirectly result in recruiting particular processes e.g. adding noise, appears to be equally difficult. This is in part due to the underlying cause of these changes e.g. the face inversion effect, remaining unknown (Farah et al 1995a; Collishaw & Hole, 2000; Goffaux, Duecker, Hausfield, Schiltz & Goebel, 2016).

In Experiment 12, it was hoped that a clearer demonstration of the stronger relationship between different face matching and configural processing could be found by disrupting rather than isolating configural processing. Inversion was used since it has been shown to compromise configural processing to a larger degree than featural processing (Freire et al., 2000; Mondloch et al., 2002). It was predicted that different face matching would be compromised to a much greater degree than same face matching in terms of accuracy, however the speed advantage for same face matching would be retained since featural processing should not be compromised to the same degree by inversion. It was also predicted that in the inverted condition same and different face matching might be associated since different face matching would not be able to use configural processing and have to share processes with same face pair matching. Both predictions were met. Accuracy for different face matching was compromised to a much greater degree than same face matching when faces were inverted. Though both same and different face matching took significantly longer in the inverted condition, the speed advantage for same face matching was retained. Same and different face matching also became associated in the inverted condition. However, interestingly performance in the upright and inverted conditions was associated across both same and different face matching. This suggests that same and different face matching shared greater resources in the inverted condition but also retained enough of their independent processes for those to be shared across inverted and upright stimuli.

Taken together these experiments demonstrate that the qualitative processes underlying same and different face matching are different. However, what those processes are is less clear. Neither experiment 10 nor experiment 11 provide clear information regarding the type of processing underlying same and different face matching. Experiment 10 provides no indication that same face matching uses featural processing to any greater degree than different face matching. It may be that replicating Experiment 10 using scrambled faces, as Collishaw & Hole (2000) did, or a single feature might be more helpful in identifying whether same face pair matching preferentially use featural processing. The left eye has been shown to be highly diagnostic in identifying gender, recognising identity and matching faces (Schyns, Bonnar & Gosselin, 2002; Vinette et al., 2004; Megreya & Havard, 2011). However, as it stands it is not possible to say any more than same face pair matching is associated with local processing.

The findings for Experiments 11 and 12 are more useful. Again, the results of experiment 11 may be confounded by a manipulation that compromises both featural and configural processing. However, if this is case, the loss of the speed advantage for same face matching provides some support for the association with featural matching. Experiment 12 appears to provide some clear indications regarding the extent to which configural processing might underlie different face matching particularly. Accuracy for different face pair matching is compromised to a much greater extent than same face pair matching when faces are inverted, building on the results of earlier research (Farah et al., 1998, Megreya & Burton, 2007). Same and different face matching also become associated when the faces are inverted suggesting both types of matching are using greater shared resources. It is the nature of these resources that is particularly interesting and points to a potential separation of mechanisms between those used specifically for faces and those used for objects.

Experiment 8 demonstrated that the face recognition subtasks remained associated whether the tasks required a greater or lesser reliance on configural processing, as demonstrated by the addition of noise. Of course, the addition of noise may not have only reduced configural processing but also featural processing. However, it is interesting that when eyes only were matched the reduction in accuracy for same face matching did not exceed that of different face matching. Collishaw and Hole (2000) found that when faces were blurred and scrambled the effect was much greater than when faces were either blurred or scrambled, evidence they used to support the dissociation between featural and configural matching. If same face matching uses both featural and configural processing it might be expected that the reduction for same face matching might be greater than for different face matching. However, this was not the case. The stronger relationship between different face matching and recognised faces viewed with noise suggests that noise may reduce both featural and configural processing but configural processing to a larger degree. If this is the case, then it is important that face recognition tasks remain related when more configural processing is required. This supports Yovel and Kanwisher's (2008) finding that configural and featural processing in faces were associated. It is also important that both same and different face matching are associated with face recognition tasks where more configural processing may be required suggesting that both same and different face matching make use of configural and featural processing as used to recognise faces.

Yovel and Kanwisher (2008) also found that configural and featural processing for objects were dissociated. All of the experiments in this chapter demonstrated that same and different face matching when viewing upright faces were dissociated and experiment 8 showed that same face matching was associated with general local processing and different face matching is associated with space perception in objects. This suggests that there are aspects of configural and featural processing that may be specific to objects that are used when matching upright faces. When faces are inverted same and different face matching are forced to use shared resources; those used to recognise faces and those that are compromised by inversion. Inversion breaks holistic processing as demonstrated by Young et al., (1987) and compromises the ability to judge distance between features rather than the features themselves Leder and Bruce (2000). While the ability of inversion to compromise the judgement of distances between features is not without question, many researchers have found matching and recognising faces where spatial distances between features have been altered *is* compromised by inversion (Freire et al., 2000; Mondloch et al., 2002; Tanaka & Sengko 1997; Yovel & Kanwisher, 2004).

The lack of clarity in research overall may be due, in part, to a lack of clarity in the definition of configural processing (Burton, Schweinberger, Jenkins & Kaufmann, 2015). Not only are the distances for comparison often poorly defined, but the term configural processing is used in face processing terms to refer to 3 separable types of processing: 1) holistic processing or viewing the face as a whole or gestalt; 2) 1st order relations or the appearance of the eyes above the nose above the mouth; and 3) 2nd order relations or the distances between the features (Maurer et al., 2002). These terms are applied differently to objects, Diamond and Carey (1986) identified first order relations as the distances between the parts of an object and second order relations only for those objects whose parts shared a typical configuration like a face. To clarify, in these experiments space perception refers to the process of comparing spatial distances between parts of an object that also appears to be used in matching but not recognising faces.

It is also difficult to provide clarity for these terms since though they may be separable they are also highly related. As Maurer et al., (2007) note, viewers process faces holistically even when they are advised not to or when to use holistic processing might be disadvantageous (Hole, 1994; Young et al., 1987). Within face matching specifically this process may be particularly disadvantageous. In experiment 12, participants described a distortion effect that occurred during face matching whereby the eyes appeared to right themselves and participants' attention was drawn towards the eyes downwards to the hair and away from the nose and mouth (Tanaka, Kaiser, Hagen & Pierce, 2014). Tanaka et al. confirmed this viewing pattern in a matching task using inverted faces, which also showed that inversion disrupted the perception of featural, and configural changes in the mouth region. In a serial matching task Haig (1984) identified that participants were particularly sensitive to changes in the vertical position of features, in particular upwards movements of the mouth. If these types of changes are important to face matching then any distortion effect that might occur when the face is inverted may draw attention away from the important areas of comparison. This may also account for the additive effect Collishaw and Hole (2000) found when blurring and inverting faces. Blurring and inverting as independent methods of modifying faces reduced recognition to around 65% whereas blurring plus inversion reduced accuracy to just above chance. In face matching this distortion might be so overwhelming as to also disrupt the configural processing mechanism used for matching faces and the space perception mechanism used for matching objects. This might explain why same and different face matching were associated in the inverted condition since both mechanisms were compromised

There is strong evidence then to support the prediction that different face matching uses configural processing to a larger degree than same face matching. The evidence goes further suggesting that different face matching also adopts a type of configural processing generally used for matching objects that same face matching does not access at all. There is less evidence to support the prediction that same face matching uses local processing to a larger degree than different face matching. However, the association between the two found in Experiment 8 and the absence of a speed advantage for same face matching when noise is applied provides support for further investigation. It is very possible that in the same way that different face matching uses configural processing mechanisms developed for matching faces and objects, same face matching uses local processing mechanisms develop for matching faces and objects. The face recognition system has been optimised to identify individuals as the same if viewed from different angles, identifying lengthening of the nose and changes in location of the features as being a change in head tilt (Collishaw, Hole & Schwaninger, 2005). However, when face matching a longer nose may not be indicative of a change of head tilt, though the head tilts of two faces for comparison may be different, but indicative of the faces belonging to different people. In this circumstance, automatic face processing mechanisms may have to be over ridden by a more manual featural comparison.

There is certainly enough evidence to show that same and different face matching are qualitatively different but also qualitatively similar. Both use underlying face processing mechanisms, some of which, like holistic processing, may have to be subjugated during face matching by object matching mechanisms. Space perception may be particularly useful in analysing images of individuals who are featurally very similar. Local processing may be particularly useful in forcing analysis of images where an individual has changed greatly or is shown from a different viewing angle. Within the face matching and face recognition context, configural processing is generally thought of as automatic with featural processing seen as more piecemeal and analytical (Maurer et al., 2002). However, the results here would suggest that within face matching both types of processing can be automatic and analytical, the difference lies in whether they are the sub types generally used for recognising faces or matching objects.

Further work is certainly required, particularly investigating within person variability and featural analysis. This might be achieved by investigating face matching using only the left eye of each image and ensuring different images of individuals are used for both same and different faces pairs. Studies using a brain imaging approach might also provide further insight into the mechanisms used in same and different face matching, indicating whether face matching shows activation in those areas of the brain associated with face processing or object processing or both. There is also a great deal of variance in face matching that is unaccounted for. However, this is the first time that face matching mechanism has been investigated using natural varying stimuli. Rather than making featural and configural changes to the stimuli the mechanisms as they naturally appear have been tested. The experiments in this chapter have shown that same and different face matching are qualitatively different and as Young et al., (1987) and Hole (1994) noted, the contributions of configural and featural processing to any face processing task certainly would appear to depend on the nature of the task.

Different face matching is associated with, predicted by and affected by manipulations that have been shown to compromise configural processing. In an applied context, this makes the need to identify high performers in both same and different face matching abundantly clear if performance is to be optimised. For those individuals already employed and carrying out the task, training should be provided for same and different face matching as independent tasks requiring a different approach. Since the nature of a face pair can never be assumed these approaches can then be used in addition to the automatic processing underlying face matching and face recognition, with the aim of increasing accuracy overall in unfamiliar face matching.

Chapter 6 – Summary & Conclusions

6.1 General Discussion

The aim of this thesis was to investigate unfamiliar face matching in the applied context. Three objectives were identified: to identify the nature of the unfamiliar face matching task carried out in an ID context; to identify measures and tasks that might predict success in unfamiliar face matching and; to test these predictions through manipulations of the stimuli.

The forensic literature identifies that analysis of evidence rarely takes place without a wider context. This wider context e.g. information about the crime, the perpetrator or the validity of other associated evidence can bias decision makers. When examining individual pieces of evidence such as fingerprints, forensic experts are unduly influenced by this context and make decisions that fit the evidence. This influence, or confirmation bias, is found throughout the forensic process from forensic experts to lay witnesses and judges alike (for review see: Kassin, Dror & Kukucka, 2013). This bias has also been seen in judgements about faces. When viewers were asked to compare a facial composite with individuals in a line-up, they were much more likely to identify an individual as most similar to the composite if they were told that the individual has been selected previously by witnesses to the crime (Charman, Gregory & Carlucci, 2009). Face matching in the applied context is rarely carried out in isolation, with faces examined by practitioners who have access to further information about the individuals or evidence surrounding the case (Heyer & Semmler, 2013). This suggests that practitioners may well be vulnerable to confirmation bias. Even within simple ID checks, viewers may be influenced by biographical information about the individual.

Research into unfamiliar face matching has traditionally and almost exclusively used faces in isolation. This approach is useful, precisely because it reduces potential confounds and identifies that even in optimum viewing conditions participants are typically very poor at unfamiliar face matching (Bruce et al., 1999; Burton et al., 2010). Where faces have been shown in credit cards or ID cards participants have not been asked to review any associated data, but to face match only (Kemp et al., 1997; Bindemann & Sandford, 2011). Chapters 2 and 3 of this thesis aimed to investigate the consequences of embedding faces in an ID context; requiring participants to match face pairs and check biographical information as required in the applied context.

Across all of the experiments in Chapter 2 face matching was more difficult when one of the faces was embedded in a passport frame. The presence of an ID context biased viewers to identify a face pair as showing the same person. This occurred regardless of whether the biographical information was valid or invalid, suggesting a form of 'halo effect' whereby the validity of the passport context biases any analysis of the face pair. This confirms that unfamiliar face matching in an ID context is different to matching unfamilar faces in isolation. The mechanisms underlying this bias are not clear and certainly require further investigation. However the implications are important for the applied context as this bias could result in individuals using a stolen passport gaining access to countries and services illegally.

The second important finding across the experiments in Chapter 2 was that data checking for invalid biographical information when shown with a face pair was very poor. Data checking for the stimuli used in Chapter 2 was measured in Chapter 3 when the stimuli was shown in isolation and within a passport frame. Accuracy for estimating the age and identifying the gender of the faces was high when the faces were shown in isolation. As was accuracy for identfying spelling errors in the places of birth. Accuracy for checking biographical information dropped when the information was placed in a passport frame with a single face. However, accuracy for checking the biographical information was compromised even further when the passport frame was shown with a face pair. These findings clearly demonstrated that viewers' ability to notice incorrect personal information was severely damaged by the presence of a face pair. Not only this, but the final experiment in Chapter 3 demonstrated that the nature of the face pair also biased accuracy when checking the biographical information. Biographical data checking was vulnerable to confirmation bias. Valid biographical information was more likely to be assumed to be incorrect when shown with a different face pair. Even more concerning was the finding that invalid biographical information was more likely to be missed if the faces matched. In the applied context this would lead to a fraudulent passport being accepted if the photo had been replaced; even if the biographical information described someone of the opposite gender or who was 20 years older.

It is important to note that these biases occurred whatever order the tasks were given and whether the tasks appeared on the same screen or separate screens. They were also present whether the comparison face was shown above the passport (as it would be in a comparison with a live person) or beside the passport (as it would be when renewing a passport). Reductions in acccuracy for both face matching and data checking occurred consistently, despite both tasks being entirely independent of the other. The reduction in performance in data checking when carried out in the presence of a face pair was not only unexpected but was also much greater than the reduction in face matching accuracy in the presence of the ID context. The reduction occurred even when participants were explicitly told to compare the data with the face in the passport frame only. A potential explanation of this finding is that, as found in other studies, faces retain attention (Jenkins et al., 2003; Lavie et al., 2003). With fewer attentional resources, the viewer is unable to detect invalid data and is further compromised by having a biased perspective when doing so.

Taken together, Chapters 2 and 3 demonstrate that face matching in the applied context is clearly different from matching faces in isolation. Unfamiliar face matching in the ID context is vulnerable to bias – both a halo effect and confirmation bias - and as a result is more difficult than matching faces alone. This suggests that the error rates predicted for the applied context may also be underestimated and highlights an even greater need for high performers to be placed in face matching roles. It would be an important extension of this work to identify whether individuals working in roles checking ID day to day are also vulnerable to the same biases. The evidence from the wider forensic literature would suggest that they may well be since both forensic finger print and handwriting examiners have been shown to be influenced by contextual information (Dror & Charlton, 2006; Kukucka & Kassin, 2014).

The wider forensic literature has identified a number of methods of reducing or minimising forensic confirmation bias in practice: working linearly, blind testing and optimising the use of technology (Kassin et al., 2013). Heyer and Semmler (2013) have indicated that these methods could be both beneficial and easily applied in the forensic face matching context. Linear working would require facial analysis experts to evaluate face images in isolation and document their findings before comparing them with target images – allowing greater objectivity. Blind testing ensures facial experts evaluate and compare face images without access to any other information about the crime e.g. eyewitness identification or expert decisions about evidence. Careful ranking of potential candidates in face recognition systems and reducing lists sizes would improve the accuracy of decisions made by practitioners. In an ID check, the image and biographical information are not easily separated. For passport examiners and border control officers this may be possible to an extent with the stages of identity confirmation being separated across different screens. However, the experiments in Chapter 2 have shown that even when face matching and data checking are carried out on separate screens the face pair continues to influence data checking accuracy. For those individuals checking ID for purchases or employment there are no other options than to view a face embedded in a passport frame. There may be potential to reduce bias by making viewers aware of the potential for confirmation bias so that they might be especially careful to review the faces and the biographical information separately. However, as Heyer and Semmler (2013) note, the effectiveness of this strategy is, as yet, unknown.

It is also important to note that with the exception of Experiment 1 the prevalence rate of errors (different face pairs and invalid biographical data) was 50%. This was to measure performance in the tasks where participants viewed stimuli in optimum detection conditions. Even in optimum viewing conditions participants were poor at spotting different face pairs and very poor at identifying errors in the biographical data. However, varying prevalence of targets in visual search tasks between 1% and 50% has shown that lower prevalence results in lower levels of accuracy when spotting targets (Wolfe et al., 2005; Wolfe, et al., 2007). This effect has also been found in unfamiliar face matching tasks by Papesh & Goldinger (2014) who identified the lower the number of identity mismatches or different face pairs, the more frequently they were undetected. It is possible then that the error detecting rates found in chapters 2 and 3 may be much lower in the applied context where the current numbers of fraudulent passports detected (0.15%, Her Majesty's Passport Office, 2014) suggest the prevalence rates are lower than 1%. This not only highlights the need to examine the effects of prevalence on performance for both face matching and biographical data in the passport context but also the importance of identifying ways in which performance can be improved.

The findings of Chapters 2 and 3 and the prevalence of confirmation bias have highlighted the need for high performers in roles where unfamiliar face matching is required. However, earlier research has shown that those we might expect to excel in unfamiliar face matching do not necessarily have any better face matching ability than the general population. Only one controlled study has shown an advantage for expertise in unfamiliar face matching (White et al., 2015). The majority have shown that experts, such as passport officers, do not perform significantly better than naïve viewers on unfamiliar face matching tasks and that time on the job bears no relationship to performance overall (Heyer, 2013; White et al., 2014a). The effects of training have also been shown to be negligible. Individuals with extensive training in facial identification have been seen to be no better than lay viewers in matching unfamiliar faces and an evaluation of the face shape training method for facial image comparison showed no advantage for training (Lee et al., 2009, Towler et al., 2014). This is in stark contrast to findings in the wider forensic literature. Training has been shown to improve fingerprint analysis (Schiffer & Champod, 2007) and fingerprint experts perform substantially better than novices in fingerprint matching tasks (Thompson, Tangen & McCarthy, 2013). In the forensic handwriting analysis field, experts have been shown to outperform lay viewers in matching handwriting, signatures and identifying simulated handwriting (Kam, Fielding & Conn, 1997; Fielding, Gummadidala & Conn, 2001; Kam, Abichandi & Hewett, 2015).

This difference in the effects of expertise and training on unfamiliar face matching when compared to other perceptual matching tasks in the forensic literature might suggest that faces are more difficult stimuli to process. This may of course be true, however, even in the unfamiliar face matching literature there have been wide ranges in performance amongst passport officers and the general population alike (White et al., 2014a; Burton et al., 2010). Some individuals perform exceptionally well on unfamiliar face matching tasks regardless of their job role or experience. The same has been seen in unfamiliar face recognition tasks. Those individuals with superior face recognition abilities or 'super recognisers' are not generally employed in roles where they have greater interaction with individuals or face related tasks. The first super recognisers to be formally tested were a PhD student, a homemaker, a municipal employee and a computer programmer (Russell et al., 2009). Equally, those in roles requiring greater interaction with individuals and face related tasks e.g. the Metropolitan police, are not necessarily super recognisers (Davis et al., in press). This would suggest that individuals with superior face matching and face recognition skills do exist, but that their abilities are unrelated to experience or training.

Studies with super recognisers have also shown that superior performance in one face processing task e.g. unfamiliar face recognition, is generally associated with superior performance in another e.g. unfamiliar face unfamiliar (Robertson et al., 2016; Bobak et al., 2016a; Bobak et al., 2016b). However, these findings are not as clear-cut as they may appear. Further work looking at the individual performance of super recognisers across face processing tasks also suggests that they may not always be consistently related. Studies have shown that super recognisers are heterogeneous; with up to 50% of participants failing to outperform controls in unfamiliar face matching tasks (Bobak et al., 2016c; Davis et al., in press).

Individuals with very poor face processing ability, or prosopagnosia, have also been shown to be relatively heterogeneous. Prosopagnosia can be broadly defined as a disorder that is specific to faces (Susilo & Duchaine, 2013a; Susilo et al., 2013b). However, prosopagnosia can be broken down into sub types by cause (acquired or congenital) and by consequence (affecting perception of faces or affecting memory for faces). Even within one type of prosopagnosia, such as congenital prosopagnosia, the consequences or behavioural deficits for individuals with the disorder can be varied (Yovel & Duchaine, 2006). Nevertheless, studies with individuals with prosopagnosia have both informed, and been supported by, research with the general population. Studies with prosopagnosic participants have demonstrated that performance in familiar and unfamiliar face tasks is dissociated and that face recognition is dissociated from object recognition (Malone et al., 1982; Farah et al., 1995b). These findings have been supported by Megreya and Burton (2006) who found that unfamiliar face matching was dissociated from familiar face recognition in the general population and Kanwisher et al., (1997) who showed that the fusiform face area (FFA) was specialised for face recognition in individuals without face processing deficits.

Individuals with prosopagnosia have relatively spared perceptual abilities and memory for objects and scenes. Individuals with Williams Syndrome (WS) have often been presented as examples of the other side of this dissociation. With often quite profound cognitive impairment in terms of their IQ and visual spatial skills they have been shown to perform as well as age matched controls on face recognition tasks (Bellugi et al., 1996; Wang & Bellugi, 1994; Bellugi et al., 1999). However, work that is more recent has suggested that both children and adults with WS perform poorly in face matching tasks requiring configural analysis (Karmiloff-Smith, 1997; Dimitriou et al.,2015). These findings suggest that face processing may not be as specific as first thought. Performance on face processing tasks may not always be associated and identifying configural changes in a face matching task may be influenced by more general visual spatial skills.

While broadly agreeing that face processing is specific, research at both ends of the face recognition spectrum also provides evidence to the contrary that is worthy of further research. Research with the general populations has also shown considerable support for the face specificity hypothesis (Kanwisher, 2000; 2010). However, support for the expertise hypothesis is just as prevalent. Studies have shown that the FFA is also recruited by individuals with expertise in birds, cars and newly learned greebles when recognising objects from those classes (Gauthier et al., 2000; Gauthier et al., 1999). Other researchers have taken an individual differences approach and have shown that unfamiliar face recognition performance is associated with more general abilities such as memory or empathy (Woodhead & Baddeley, 1981; Bate et al., 2010).

Unfamiliar face matching has been the focus of less research. Only one published study has investigated the nature of unfamiliar face matching specifically using an individual differences approach with a more representative sample. Megreya and Burton (2006) invited students to participant in a battery of face processing and more general tasks. They found that unfamiliar face matching was associated with unfamiliar face recognition but not familiar face recognition. Unfamiliar face matching was also associated with perceptual speed, a visual short-term memory task and a familiar object matching task. These results supported the dissociation of familiar and unfamiliar faces found by Malone et al., (1982) and suggested that unfamiliar face matching did not use the face specific processes underlying familiar face recognition. Rather, unfamiliar face matching made use of more general mechanisms underlying object matching. The results from this study were so persuasive that Megreya and Burton (2006) argued that unfamiliar faces were not faces.

Taken together the findings regarding the face specific nature of unfamiliar face matching provide a mixed picture. Studies with individuals with Williams Syndrome and super recognisers at the individual level have found that unfamiliar face matching and unfamiliar face recognition are dissociated (Karmiloff-Smith, 1997; Bobak et al., 2016c; Davis et al., in press). Other research with individuals with prosopagnosia and the general population have shown a dissociation between familiar and unfamiliar face tasks (Malone et al., 1982; Megreya & Burton, 2006). In addition both Karmiloff-Smith (1997) and Megreya and Burton found that performance in unfamiliar face matching was associated with performance in more general tasks. In contrast, studies with super recognisers at the general level have shown that superior performance is associated across familiar and unfamiliar face recognition tasks and face recognition and face matching tasks (Russell et al., 2009; Robertson et al., 2016; Bobak et al., 2016a, 2016b; 2016c; Davis et al., in press). It is against this context that the objective for Chapter 4 was established - to identify measures and tasks that might predict success in unfamiliar face matching. A battery of tests was assembled online so that participants from the general population could complete a range of face processing tasks and other cognitive and personality measures. This would allow associations and dissociations to be found; not only between face processing tasks but between specific face processing tasks and the more general measures.

The results of the face battery confirmed that face processing is specific. With the exception of the face detection task, all of the face processing tasks (an unfamiliar face matching task, an unfamiliar face recognition task and a familiar face recognition task) were associated. This was the first time that this had been shown in the general population. Face processing as a whole was not associated with any of the general measures. However, there were some specific associations. Unfamiliar face recognition was associated with the Silhouettes task, a task in which participants were required to recognise silhouettes of objects shown from unusual perspectives. This could suggest a general reliance on memory or processes for transforming images since many of the test faces in the recognition task were also shown from novel perspectives. Unfamiliar face matching was again associated with a familiar object matching task suggesting that both shared underlying mechanisms. Perhaps most interestingly, same and different face matching were not only dissociated, as had been shown previously (Megreya & Burton, 2007), but they associated differentially with the other face processing tasks and more general tasks. Same face pair matching was positively associated with all of the face processing tasks (with the exception of face detection) and local processing. Different face pair matching was only associated with the unfamiliar face recognition task and space perception. The strongest association between different face matching and the unfamiliar face recognition task was with faces shown with noise, a task designed to isolate configural processing.

These results demonstrated that while face processing tasks are associated there are degrees of association. For example, the unfamiliar face tasks were associated with each other to a much greater degree than they were with familiar face recognition task. This might explain why the associations across the familiarity spectrum are not always found for face tasks. These results also suggest that while success in unfamiliar face matching might be predicted by performance in any face identity task, greater accuracy and confidence can be gained from this prediction if it is using unfamiliar face recognition tasks. Performance in unfamiliar face matching can also be predicted by performance in local processing and space perception. This seems counter-intuitive since high levels of performance in configural or global processing is generally understood to underlie face recognition rather than local processing (Maurer et al., 2002). However, it may explain the dissociation between same and different face pair matching, since featural or local processing and configural or spatial relations have been found to be associated for faces but dissociated for objects (Yovel & Kanwisher, 2008).

Chapter 4 provided a number of novel findings regarding unfamiliar face matching which were important to test in the context of an unfamiliar face matching task. This was the purpose of Chapter 5 and the third objective of this thesis. Three experiments were designed to test the predictions of the face battery through manipulations of the stimuli in the unfamiliar face matching task. Two experiments aimed to isolate either featural (local) or configural (space perception) processing and a third to disrupt configural processing to identify if same and different face processing were differentially affected. The experiments isolating features for unfamiliar face matching using a pair of eyes and adding noise to isolate configural processing did not provide conclusive evidence that same and different face matching used local and configural processing differently. The use of pairs of eyes as a single feature may have allowed configural processing between the eyes and eyebrows. The use of noise to disrupt configural processing may also have disrupted featural processing. These experiments confirmed as others had previously that isolating processing mechanisms is difficult (Rakover, 2002). However, disrupting configural processing using inversion did create different results for same and different face matching. Accuracy for same face pair matching was only reduced by a very small extent by inversion, whereas accuracy for different face pair matching was reduced by over 40%. When upright same and different face matching were dissociated, when inverted they were associated suggesting a change in underlying mechanisms. The results of Chapter 5 confirmed the qualitative difference between same and different face matching and provides evidence that different face matching differentially makes use of configural processing or space perception. This adds greater weight to evidence supporting the prediction of different face matching ability not only by performance in unfamiliar face recognition but also in space perception.

In the theoretical context the findings from this thesis are both novel and informative. They suggest that contrary to Megreya and Burton's (2006) view, unfamiliar faces are faces. Chapters 2 and 3 demonstrate that unfamiliar faces, like familiar faces retain attention. Chapter 4 demonstrates that unfamiliar face matching is associated with familiar face recognition and makes use of face specific mechanisms. However, both Chapter 4 and Chapter 5 provide evidence for unfamiliar face matching also making use of more general mechanisms. The more general mechanisms are used to support the limited face specific processes.

Taken together these findings suggest that unfamiliar face matching may be highly image bound but may not be as limited to using pictorial codes as Bruce and Young (1986) identified. In order for same pair faces to be matched when images are unconstrained (in the case for the GFMT taken with different cameras), some kind of transformation has to be possible to confirm that they are the same identity. This may require making use of shallower or more limited structural codes to model the within person variability shown in the images provided. The use of these limited structural codes may also explain the positive but relatively weaker nature of the association found between unfamiliar face matching and familiar face recognition.

At least some then of the mechanisms underlying unfamiliar face matching are face specific. The face specificity hypothesis maintains that the face processing system has been optimised over its evolutionary history to process faces (Kanwisher, 2000). It must be remembered that face matching is new from an evolutionary perspective. It is thought that the first photo ID was used for exhibitors at the 1876 Centennial Exposition in Philadelphia Pennsylvania (Hall, Dodds & Triggs, 1993). However, photographs only became more wide spread for identification purposes when photographs became a requirement of the British Passport in 1915 (Parkinson, 2015). If face processing is specific and optimised over time, this has only given humans 100 years to adapt to the requirements of unfamiliar face matching. The archaeological evidence points to a separation between chimpanzees and humans around 4.4 million years ago, with modern humans emerging around 50,000 years ago (Klein, 1995). Defined as the first humans demonstrating 'human uniqueness' modern humans developed a set of unique behaviours and cognitive abilities not seen in primates. These behaviours and abilities saw modern humans interacting on much more complex levels beyond close kinship groups to both compete and co-operate e.g. share and trade resources (Hill, Barton & Hurtado, 2009). At the very least, this might suggest that familiar face recognition has been 4.4 million years in the making and unfamiliar face recognition 50,000 years. Unfamiliar face matching has not been a requirement long enough for the face processing system to be optimised to meet those requirements. It makes evolutionary sense that individuals are using the best of both face and object processing systems to match images and individuals.

The use of both face specific and more general mechanisms in unfamiliar face matching could be modelled using functional components from Bruce and Young's (1986) framework for familiar face recognition as shown in figure 6.1. Many of the components from the original Bruce and Young (1986) framework are not required, such as the Person Identity Node, as very little semantic information is known about the faces. When matching unfamiliar faces, the most efficient and accurate method would be to identify if any of the faces are familiar. This would result in expression independent descriptions being created from the view centred descriptions and being compared with the face recognition units of familiar faces to identify if the faces were known/familiar. This would be carried out using holistic processing, identified as automatic by (Young et al., 1987).   If recognition were not achieved, then this automatic process would have to be overridden and local processing and space perception mechanisms used via directed visual processing. These general measures would be used to evaluate the view centred descriptions and allow comparison between them. The results of this comparison could be contrasted with more automatic decision-making carried out between expression-independent descriptions. This would allow the optimum same and different face matching decisions to be made. As a result, the most accurate structural codes could be developed for individual(s) identified.

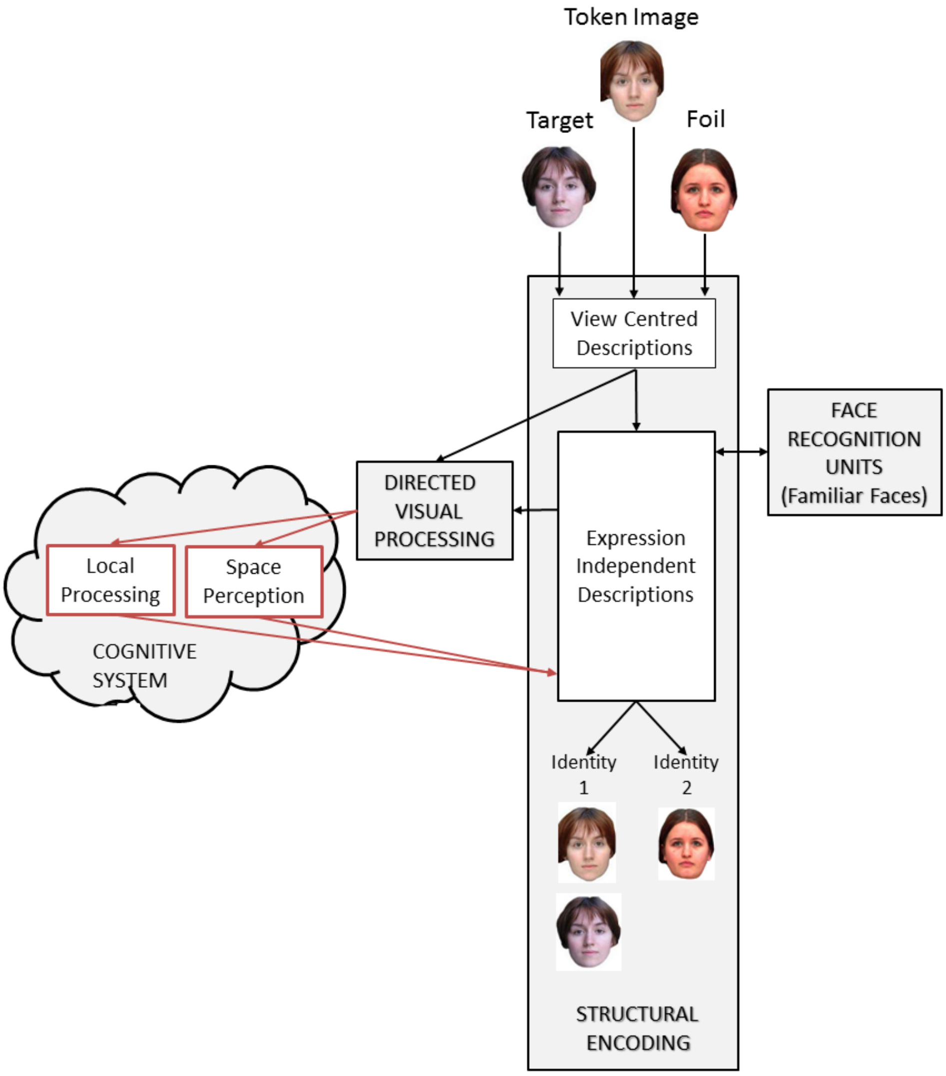

Figure 6.1. A functional model of unfamiliar face matching.

This unfamiliar face matching model would also fit with the perceptual models proposed by Valentine (1991) and the perceptual and cognitive model proposed by Burton, Bruce and Hancock (1999b). Both models use the concept of face space in which faces are represented as dimensions or codes in space. Since this encoding is required for the first recognition part of the unfamiliar face matching model and both the Valentine (1991) and Burton et al., (1999b) model rely on initial holistic coding of the face there are no obvious challenges to either of these models by the unfamiliar face matching model.

The findings of Chapter 5 also require further investigation and replication to confirm this model. This is particularly true in terms of the relationship between same and different face matching and local processing. It may be useful for example to try to isolate local processing by scrambling the faces or using the left eye of each face only (Collishaw & Hole 2000; Megreya & Havard, 2011). It may also be useful to employ a brain imaging approach to investigate a neural model of unfamiliar face matching. Brain imaging such as fMRI studies may identify whether those regions of the brain associated with local and configural processing in faces and/or objects are differentially activated when matching same and different face pairs (Yovel & Kanwisher, 2008).

In the applied context, the findings from this thesis suggest that individuals with high levels of accuracy in unfamiliar face matching could be recruited from pools of super recognisers or those with super recogniser ability. Since unfamiliar face matching and unfamiliar face recognition are associated the likelihood of finding high performers in this group would be greater than in the general population. Even so, these individuals' unfamiliar face matching score may not always be associated with their unfamiliar face recognition ability. Super recognisers should still be screened using a test of unfamiliar face matching and the suitability for a face matching role assessed on their performance using their scores for both same and different face matching.

Optimising the performance of individuals already in the role may provide another way of improving unfamiliar face matching in the applied context. Working in pairs and feedback have been shown to improve unfamiliar face matching performance but the gains made are limited (White, 2014c; Dowsett & Burton 2015). However, the findings that unfamiliar face matching ability can be predicted by performance in local processing and space perception tasks suggest that new areas for training might be available.

When perceiving the world, adults see the forest before the trees (Navon 1977; Krakowski et al., 2016). Using a negative priming task Poirel et al., (2014) have shown that this global precedence in the perception of objects can be overcome through inhibition or executive control. Training has been shown to improve executive function (for review see Karbach & Verhaeghen, 2014). This suggests individuals may be able to improve their ability to prioritise local processing. Training in spatial skills has also shown that significant gains in spatial ability can be made (for review see Uttal et al., 2013). This review also demonstrates that training using video games has great potential for improving spatial skills which would also be easily implemented in the applied context. Increasing local processing and space perception ability may have the potential to improve unfamiliar face matching. Since the nature of each face pair is unknown it may appear difficult to know when and how to apply these strategies. However, carrying out local (featural) analysis on the right side of the faces and space perception (configural) analysis on the left side may not only provide a clear methodology but also make best use of hemispheric specialisation (Christie, Ginsberg, Steedman, Fridriksson, Bonilha & Rorden, 2012).

Appendices

Appendix 1 – Biographical Information used in Experiment 1 (Invalid data highlighted)

Surname	Given Names	Date of Birth	Sex	Place of Birth
SMITH	ELIZABETH	1 JUN /JUIN 93	F	BIRMINGHAM
JONES	SARAH	2 JUN /JUIN 93	F	LEEDS
BROWN	EMILY	3 JUN /JUIN 93	F	GLASGOW
WILLIAMS	LUCY	4 JUN /JUIN 93	F	SHEFFIELD
WILSON	EMMA	5 JUN /JUIN 93	F	BRADFORD
THOMSON	HANNAH	6 JUN /JUIN 93	F	MANCHESTER
ROBERTSON	CHARLOTTE	7 JUN /JUIN 93	F	EDINBURGH
TAYLOR	MARIA	8 JUN /JUIN 93	F	LIVERPOOL
CAMPBELL	RACHEL	9 JUN /JUIN 93	F	KIRKLEES
DAVIES	ANNA	10 JUN /JUIN 93	F	FIFE
STEWART	ALICE	11 JUN /JUIN 93	F	CROYDON
WILSON	LILY	12 JUN /JUIN 93	F	BARNET
ANDERSON	CATHERINE	13 JUN /JUIN 93	F	CARDIFF
EVANS	GRACE	14 JUN /JUIN 93	F	EALING
MACDONALD	HELEN	15 JUN /JUIN 93	F	NORTH LANARKSHIRE
THOMAS	KATE	16 JUN /JUIN 93	F	WAKEFIELD
SCOTT	PAUL	17 JUN /JUIN 93	M	WIRRAL
JOHNSON	RYAN	18 JUN /JUIN 93	M	WIGAN
REID	SCOTT	19 JUN /JUIN 93	M	COVENTRY
ROBERTS	SEAN	20 JUN /JUIN 93	M	SOUTH LANARKSHIRE
MURRAY	SIMON	21 JUN /JUIN 93	M	DUDLEY
WALKER	ALEXANDER	22 JUN /JUIN 93	M	ENFIELD
TAYLOR	ANDREW	23 JUN /JUIN 93	M	BRENT
WRIGHT	DAVID	24 JUN /JUIN 93	M	BROMLEY
CLARK	JAMES	25 JUN /JUIN 93	M	SANDWELL
ROBINSON	JOHN	26 JUN /JUIN 93	M	NEWHAM
MITCHELL	MICHAEL	27 JUN /JUIN 93	M	WANDSWORTH
THOMPSON	ROBERT	28 JUN /JUIN 93	M	LAMBETH
ROSS	THOMAS	29 JUN /JUIN 93	M	DONCASTER
WHITE	WILLIAM	30 JUN /JUIN 93	M	SOUTHWARK
WALKER	CHARLES	1 JUL /JUIL 93	M	STOCKPORT
HUGHES	CHRISTOPHER	2 JUL /JUIL 93	M	BELFAST
PATERSON	DANIEL	3 JUL /JUIL 93	M	NEWCASTLE UPON TYNE
EDWARDS	GEORGE	4 JUL /JUIL 93	M	REDBRIDGE
YOUNG	JOSEPH	5 JUL /JUIL 93	M	BOLTON
GREEN	MARK	6 JUL /JUIL 93	M	LEWISHAM
WATSON	PETER	7 JUL /JUIL 93	M	NORTH TYNESIDE
HALL	BENJAMIN	8 JUL /JUIL 93	M	HILLINGDON
MORRISON	HARRY	9 JUL /JUIL 93	M	SEFTON

Surname	Given Names	Date of Birth	Sex	Place of Birth
WOOD	MATTHEW	10 JUL /JUIL 93	M	WALSALL
MILLER	OLIVER	11 JUL /JUIL 93	M	WALTHAM FOREST
TAYLOR	MARC	12 JUL /JUIL 93	M	LEWISHAM
ANGELO	BRENDA	13 JUL /JUIL 93	F	HARINGEY
BOULSTRIDGE	HILARY	14 JUL /JUIL 93	F	GREENWICH
BUNGARD	SYLVIA	15 JUL /JUIL 93	F	TOWER HAMLETS
BURSNELL	JILL	16 JUL /JUIL 93	F	HOUNSLOW
CABRERA	JOYCE	17 JUL /JUIL 93	F	ABERDEENSHIRE
CHAISTY	HAZEL	18 JUL /JUIL 93	F	ROSSENDALE
CLAYWORTH	ANNETTE	19 JUL /JUIL 93	F	CASTLEREAGH
DENIAL	GAIL	20 JUL /JUIL 93	F	SOUTH BUCKS
NOWAK	ALEKSANDRA	21 JUL /JUIL 93	F	WARSAW
DOMVILLE	GERALDINE	22 JUL /JUIL 93	F	WEYMOUTH AND PORTLAND
DUA	DIANA	23 JUL /JUIL 93	F	BOSTON
EDESON	CAROLYN	24 JUL /JUIL 93	F	BALLYMENA
GARROTT	PENELOPE	25 JUL /JUIL 93	F	TORRIDGE
GASPAR	JEANETTE	26 JUL /JUIL 93	F	NORTH WARWICKSHIRE
GAUGE	THERESA	27 JUL /JUIL 93	F	FERMANAGH
GELSON	JOY	28 JUL /JUIL 93	F	MALDON
HAPPER	MACKENZIE	29 JUL /JUIL 93	M	CORBY
HAWA	LUCA	30 JUL /JUIL 93	M	ADUR
HELLING	KEIRAN	31 JUL /JUIL 93	M	FOREST HEATH
HOLLINGBERRY	JUDE	1 AUG /AOÛ 93	M	ARMAGH
HOWSHAM	HARRISON	2 AUG /AOÛ 93	M	COLERAINE
HUSHER	RORY	3 AUG /AOÛ 93	M	MERTHYR TYDFIL
HUTH	ROY	4 AUG /AOÛ 93	M	DUNGANNON
MURPHY	JACK	5 AUG /AOÛ 93	M	DUBLIN
KINLAN	RUSSELL	6 AUG /AOÛ 93	M	OADBY AND WIGSTON
LE FEUVRE	LEWIS	7 AUG /AOÛ 93	M	CRAVEN
LEATHERBY	SCOTT	8 AUG /AOÛ 93	M	WEST DEVON
LOWSLEY	SEAN	9 AUG /AOÛ 93	M	ANTRIM
MARDLING	SEBASTIAN	10 AUG /AOÛ 93	M	SOUTH LAKELAND
MCCART	SETH	11 AUG /AOÛ 93	M	RICHMONDSHIRE
MCCALMAN	SHAY	12 AUG /AOÛ 93	M	RYEDALE
MCKIDDIE	SONNY	13 AUG /AOÛ 93	M	CLACKMANNANSHIRE
MCQUILLEN	STANLEY	14 AUG /AOÛ 93	M	OMAGH
MEATH	TAYLOR	15 AUG /AOÛ 93	M	MELTON
MUSTOW	THEO	16 AUG /AOÛ 93	M	BANBRIDGE
NANA	THEODORE	17 AUG /AOÛ 93	M	CHRISTCHURCH
PEPALL	TOBY	18 AUG /AOÛ 93	M	MAGHERAFELT
PERDUE	TOM	19 AUG /AOÛ 93	M	PURBECK
RAVENSDALE	TOMMY	20 AUG /AOÛ 93	M	STRABANE
RUKIN	VINCENT	21 AUG /AOÛ 93	M	CARRICKFERGUS
SHELSHER	WAYNE	22 AUG /AOÛ 93	M	COOKSTOWN
SILSBURY	ZAC	23 AUG /AOÛ 93	M	WEST SOMERSET

Surname	Given Names	Date of Birth	Sex	Place of Birth
FRASER	**PAUL**	01 JUN /JUIN 93	F	LIMAVADY
LEWIS	**STEPHEN**	2 JUN /JUIN 93	F	LARNE
DAVIDSON	**ADAM**	3 JUN /JUIN 93	F	BALLYMONEY
MARTIN	**ALAN**	4 JUN /JUIN 93	F	EILEAN SIAR
GRAY	**ANGER**	5 JUN /JUIN 93	F	SHETLAND ISLANDS
JACKSON	**EGG**	6 JUN /JUIN 93	F	ORKNEY ISLANDS
CLARKE	**ORANGE**	7 JUN /JUIN 93	F	MOYLE
SMITH	**JOKING**	8 JUN /JUIN 93	F	BIRMINGHAM
JONES	SOPHIE	**41 JUN /JUIN 93**	F	LEEDS
BROWN	ZOE	**42 JUN /JUIN 93**	F	GLASGOW
WILLIAMS	JULIA	**43 JUN /JUIN 93**	F	SHEFFIELD
WILSON	ALEXANDRA	**44 JUN /JUIN 93**	F	BRADFORD
THOMSON	JENNIFER	**45 JUN /JUIN 93**	F	MANCHESTER
BROWN	HEATHER	**12 JUN / JUIN 24**	F	EDINBURGH
ROBERTSON	STEPHANIE	**13 JUN / JUIN 35**	F	LIVERPOOL
TAYLOR	DOROTHY	**14 JUN / JUIN 27**	F	KIRKLEES
CAMPBELL	**WHEEL**	15 JUN / JUIN 93	M	FIFE
DAVIES	**CAFETIERE**	18 JUN /JUIN 93	M	CROYDON
STEWART	**ICEBERG**	19 JUN /JUIN 93	M	BARNET
WILSON	**FORK**	20 JUN /JUIN 93	M	CARDIFF
BROWN	**MELON**	21 JUN /JUIN 93	M	EALING
FRASER	**DAISY**	22 JUN /JUIN 93	M	NORTH LANARKSHIRE
LEWIS	**ANNE**	23 JUN /JUIN 93	M	WAKEFIELD
DAVIDSON	**ANN**	24 JUN /JUIN 93	M	WIRRAL
MARTIN	**LAURA**	25 JUN /JUIN 93	M	WIGAN
GRAY	**RUTH**	26 JUN /JUIN 93	M	COVENTRY
JACKSON	**CAROLINE**	27 JUN /JUIN 93	M	SOUTH LANARKSHIRE
CLARKE	**LISA**	28 JUN /JUIN 93	M	DUDLEY
KHAMBAITA	MARTIN	29 JUN /JUIN 93	M	ENFIELD
SELVARATNAM	OWEN	30 JUN /JUIN 93	M	BRENT
UPADHYAD	PHILIP	08 JUN /JUIN 93	M	BROMLEY
VALJI	RICHARD	9 JUN /JUIN 93	M	SANDWELL
VIRJI	COLIN	10 JUN /JUIN 93	M	NEWHAM
DISSANAYAKE	HENRY	11 JUN /JUIN 93	M	WANDSWORTH
SMITH	IAN	**12 JUN /JUIN 25**	M	LAMBETH
JONES	JACK	**13 JUN /JUIN 36**	M	DONCASTER
BROWN	JACOB	**14 JUN /JUIN 28**	M	SOUTHWARK
WILLIAMS	JAMIE	**15 JUN /JUIN 31**	M	STOCKPORT
WILSON	JONATHAN	**30 FEB/ FEV 93**	M	BELFAST
THOMSON	KENNETH	**31 FEB/ FEV 93**	M	NEWCASTLE UPON TYNE
BROWN	LOUIS	**32 FEB/ FEV 93**	M	REDBRIDGE
ROBERTSON	LUKE	**33 FEB/ FEV 93**	M	BOLTON

Appendix 2 – Biographical Information used in Experiment 2 & 3 (Invalid data highlighted)

Surname	Given Names	Date of Birth	Sex	Place of Birth
ANDERSON	CATHERINE	13 JUN /JUIN 93	F	CARDIFF
EVANS	GRACE	14 JUN /JUIN 93	F	EALING
MACDONALD	HELEN	15 JUN /JUIN 93	F	NORTH LANARKSHIRE
THOMAS	KATE	16 JUN /JUIN 93	F	WAKEFIELD
SCOTT	PAUL	17 JUN /JUIN 93	M	WIRRAL
JOHNSON	RYAN	18 JUN /JUIN 93	M	WIGAN
REID	SCOTT	19 JUN /JUIN 93	M	COVENTRY
ROBERTS	SEAN	20 JUN /JUIN 93	M	SOUTH LANARKSHIRE
MURRAY	SIMON	21 JUN /JUIN 93	M	DUDLEY
WALKER	ALEXANDER	22 JUN /JUIN 93	M	ENFIELD
FRASER	**PAUL**	01 JUN /JUIN 93	F	LIMAVADY
BROWN	ZOE	**30 FEB /FEV 93**	F	EILEAN SIAR
GRAY	**ANGER**	05 JUN /JUIN 93	F	SHETLAND ISLANDS
BROWN	HEATHER	**12 JUN / JUIN 24**	F	BIRMINGHAM
FRASER	**DAISY**	15 JUN / JUIN 93	M	FIFE
WILSON	JONATHAN	**30 FEB/ FEV 93**	M	CARDIFF
SMITH	IAN	**12 JUN /JUIN 25**	M	COVENTRY
JONES	JACK	**13 JUN /JUIN 36**	M	SOUTH LANARKSHIRE
WILSON	**FORK**	20 JUN /JUIN 93	M	DONCASTER
SELVARATNAM	OWEN	30 JUN /JUIN 93	M	BOLTON
ANGELO	BRENDA	13 JUL /JUIL 93	F	HARINGEY
BOULSTRIDGE	HILARY	14 JUL /JUIL 93	F	GREENWICH
DENIAL	GAIL	20 JUL /JUIL 93	F	SOUTH BUCKS
NOWAK	ALEKSANDRA	21 JUL /JUIL 93	F	WARSAW
HAPPER	MACKENZIE	29 JUL /JUIL 93	M	CORBY
LEATHERBY	SCOTT	8 AUG /AOÛ 93	M	WEST DEVON
MCKIDDIE	SONNY	13 AUG /AOÛ 93	M	CLACKMANNANSHIRE
NANA	THEODORE	17 AUG /AOÛ 93	M	CHRISTCHURCH
PERDUE	TOM	19 AUG /AOÛ 93	M	PURBECK
SHELSHER	WAYNE	22 AUG /AOÛ 93	M	COOKSTOWN

Appendix 3 – Biographical Information used in Experiment 4 (Invalid data highlighted)

Surname	Given Names	Date of Birth	Sex	Place of Birth
JAMES	SAMANTHA	**27 JUL /JUIL 73**	F	WOLVERHAMPTON
REES	RACHAEL	**28 JUL /JUIL 73**	F	STOKE-ON-TRENT
RUSSELL	LIAM	**20 AUG /AOÛ 73**	M	TELFORD
GIBSON	KIERAN	**21 AUG /AOÛ 73**	M	WARRINGTON
PRICE	NATALIE	**31 JUL /JUIL 73**	F	LANCASTER
PHILLIPS	NICOLE	**1 AUG /AOÛ 73**	F	BARNSLEY
SUTHERLAND	ADAM	**01 JUN /JUIN 73**	M	PRESTON
CRAIG	NATHAN	**21 FEB /FEV 73**	M	PETERBOROUGH
FRASER	**JOSHUA**	4 AUG /AOÛ 93	**M**	ABERDEEN
GRAY	**SHAUN**	5 AUG /AOÛ 93	**M**	CAMDEN
BURNS	**LISA**	09 JUN /JUIN 93	**F**	BATH
MUIR	**BETHANY**	12 JUN /JUIN 93	**F**	ST. HELENS
HENDERSON	**PETER**	8 AUG /AOÛ 93	**M**	ROCHDALE
KERR	**DEAN**	9 AUG /AOÛ 93	**M**	SWINDON
BURTON	**CLAIRE**	03 JUN /JUIN 93	**F**	BASILDON
WATT	**NICOLA**	04 JUN /JUIN 93	**F**	COLCHESTER
GRAHAM	ERIN	19 JUL /JUIL 93	F	**PLLYMOUTH**
MARTIN	KATIE	20 JUL /JUIL 93	F	**EDINBORUG**
CAMERON	MARK	12 AUG /AOÛ 93	M	**WESTMINTSER**
DUNCAN	JACK	13 AUG /AOÛ 93	M	**LUUTON**
SANDERS	LOUISE	23 JUL /JUIL 93	F	**SOUTHAMTPON**
LEWIS	HOLLY	24 JUL /JUIL 93	F	**SWASNEA**
GRANT	CAMERON	16 AUG /AOÛ 93	M	**OXFROD**
ALLAN	CONNOR	17 AUG /AOÛ 93	M	**CNATERBURY**
SMITH	ELIZABETH	1 JUN /JUIN 93	F	BIRMINGHAM
JONES	SARAH	2 JUN /JUIN 93	F	LEEDS
PATERSON	DANIEL	3 JUL /JUIL 93	M	NEWCASTLE UPON TYNE
EDWARDS	GEORGE	4 JUL /JUIL 93	M	REDBRIDGE
WILSON	EMMA	5 JUN /JUIN 93	F	BRADFORD
THOMSON	HANNAH	6 JUN /JUIN 93	F	MANCHESTER
WATSON	BENJAMIN	7 JUL /JUIL 93	M	BRISTOL
HALL	HARRY	8 JUL /JUIL 93	M	HILLINGDON
BROWN	EMILY	3 JUN /JUIN 93	F	GLASGOW
WILLIAMS	LUCY	4 JUN /JUIN 93	F	SHEFFIELD
YOUNG	JOSEPH	5 JUL /JUIL 93	M	BOLTON
GREEN	MARTIN	6 JUL /JUIL 93	M	LEWISHAM
ROBERTSON	CHARLOTTE	7 JUN /JUIN 93	F	CAMBRIDGE
TAYLOR	MARIA	8 JUN /JUIN 93	F	LIVERPOOL
MORRISON	MATTHEW	9 JUL /JUIL 93	M	IPSWICH
WOOD	OLIVER	10 JUL /JUIL 93	M	WALSALL
CAMPBELL	RACHEL	9 JUN /JUIN 93	F	NOTTINGHAM

Surname	Given Names	Date of Birth	Sex	Place of Birth
DAVIES	ANNA	10 JUN /JUIN 93	F	SUNDERLAND
MILLER	PAUL	11 JUL /JUIL 93	M	KENSINGTON AND CHELSEA
JOHNSTON	RYAN	12 JUL /JUIL 93	M	TOWER HAMLETS
MACDONALD	HELEN	15 JUN /JUIN 93	F	GLASGOW
THOMAS	KATE	16 JUN /JUIN 93	F	WAKEFIELD
O'NEILL	SIMON	15 JUL /JUIL 93	M	BRIGHTON
DOHERTY	ALEXANDER	16 JUL /JUIL 93	M	LEICESTER

Appendix 4 – Biographical Information used in Experiment 7 (Invalid data highlighted)

Surname	Given Names	Date of Birth	Sex	Place of Birth
SMITH	ELIZABETH	1 JUN /JUIN 93	F	BIRMINGHAM
JONES	SARAH	2 JUN /JUIN 93	F	LEEDS
BROWN	EMILY	3 JUN /JUIN 93	F	GLASGOW
WILLIAMS	LUCY	4 JUN /JUIN 93	F	SHEFFIELD
WILSON	EMMA	5 JUN /JUIN 93	F	BRADFORD
THOMSON	HANNAH	6 JUN /JUIN 93	F	MANCHESTER
ROBERTSON	CHARLOTTE	7 JUN /JUIN 93	F	CAMBRIDGE
TAYLOR	MARIA	8 JUN /JUIN 93	F	LIVERPOOL
CLARK	JAMES	25 JUN /JUIN 93	M	SANDWELL
ROBINSON	JOHN	26 JUN /JUIN 93	M	NEWHAM
MITCHELLEL	MICHAEL	27 JUN /JUIN 93	M	WANDSWORTH
THOMPSON	ROBERT	28 JUN /JUIN 93	M	LAMBETH
ROSS	THOMAS	29 JUN /JUIN 93	M	DONCASTER
WHITE	WILLIAM	30 JUN /JUIN 93	M	SOUTHWARK
KELLY	CHARLES	1 JUL /JUIL 93	M	STOCKPORT
HUGHES	CHRISTOPHER	2 JUL /JUIL 93	M	BELFAST
CAMPBELL	RACHEL	9 JUN /JUIN 93	F	NOTTINGHAM
DAVIES	ANNA	10 JUN /JUIN 93	F	SUNDERLAND
STEWART	ALICE	11 JUN /JUIN 93	F	CROYDON
JACKSON	LILY	12 JUN /JUIN 93	F	BARNET
ANDERSON	CATHERINE	13 JUN /JUIN 93	F	CARDIFF
EVANS	GRACE	14 JUN /JUIN 93	F	EALING
MACDONALD	HELEN	15 JUN /JUIN 93	F	GLASGOW
THOMAS	KATE	16 JUN /JUIN 93	F	WAKEFIELD
PATERSON	DANIEL	3 JUL /JUIL 93	M	NEWCASTLE UPON TYNE
EDWARDS	GEORGE	4 JUL /JUIL 93	M	REDBRIDGE
YOUNG	JOSEPH	5 JUL /JUIL 93	M	BOLTON
GREEN	MARTIN	6 JUL /JUIL 93	M	LEWISHAM
WATSON	BENJAMIN	7 JUL /JUIL 93	M	BRISTOL
HALL	HARRY	8 JUL /JUIL 93	M	HILLINGDON
MORRISON	MATTHEW	9 JUL /JUIL 93	M	IPSWICH
WOOD	OLIVER	10 JUL /JUIL 93	M	WALSALL
SCOTT	CHLOE	17 JUN /JUIN 93	F	WIRRAL
JOHNSON	MEGAN	18 JUN /JUIN 93	F	WIGAN
REID	REBECCA	19 JUN /JUIN 93	F	COVENTRY
ROBERTS	AMY	20 JUN /JUIN 93	F	HARROGATE
MURRAY	LAUREN	21 JUN /JUIN 93	F	DUDLEY
WALKER	SHANNON	22 JUN /JUIN 93	F	KINGSTON UPON THAMES
TAYLOR	SOPHIE	23 JUN /JUIN 93	F	NORWICH
WRIGHT	DANIELLE	24 JUN /JUIN 93	F	BROMLEY
MILLER	PAUL	11 JUL /JUIL 93	M	KENSINGTON AND CHELSEA

Surname	Given Names	Date of Birth	Sex	Place of Birth
JOHNSTON	RYAN	12 JUL /JUIL 93	M	TOWER HAMLETS
MOORE	SCOTT	13 JUL /JUIL 93	M	FALKIRK
SMYTH	SEAN	14 JUL /JUIL 93	M	YORK
O'NEILL	SIMON	15 JUL /JUIL 93	M	BRIGHTON
DOHERTY	ALEXANDER	16 JUL /JUIL 93	M	LEICESTER
QUINN	ANDREW	17 JUL /JUIL 93	M	GREENWICH
MURPHY	DAVID	18 JUL /JUIL 93	M	DURHAM
GRAHAM	ERIN	19 JUL /JUIL 93	F	**PLLYMOUTH**
MARTIN	KATIE	20 JUL /JUIL 93	F	**EDINBORUG**
HAMILTON	KIRSTY	21 JUL /JUIL 93	F	**MILNOT KEENES**
MITCHELL	LAURA	22 JUL /JUIL 93	F	**WARKICW**
SANDERS	LOUISE	23 JUL /JUIL 93	F	**SOUTHAMTPON**
LEWIS	HOLLY	24 JUL /JUIL 93	F	**SWASNEA**
MORGAN	JENNIFER	25 JUL /JUIL 93	F	**LINLCON**
GRIFFITHS	MORGAN	26 JUL /JUIL 93	F	**SOLHULLI**
CAMERON	MARK	12 AUG /AOÛ 93	M	**WESTMINTSER**
DUNCAN	JACK	13 AUG /AOÛ 93	M	**LUUTON**
HUNTER	ROSS	14 AUG /AOÛ 93	M	**STAFDROF**
BELL	LEWIS	15 AUG /AOÛ 93	M	**BOURNMEOUHT**
GRANT	CAMERON	16 AUG /AOÛ 93	M	**OXFROD**
ALLAN	CONNOR	17 AUG /AOÛ 93	M	**CNATERBURY**
BLACK	CALLUM	18 AUG /AOÛ 93	M	**KINGSTON UOPN HULL**
MCLEAN	JORDAN	19 AUG /AOÛ 93	M	**ETERX**
JAMES	SAMANTHA	**27 JUL /JUIL 73**	F	WOLVERHAMPTON
REES	RACHAEL	**28 JUL /JUIL 73**	F	STOKE-ON-TRENT
JENKINS	HEATHER	**29 JUL /JUIL 73**	F	SOUTHEND-ON-SEA
OWEN	JADE	**30 JUL /JUIL 73**	F	SALFORD
PRICE	NATALIE	**31 JUL /JUIL 73**	F	LANCASTER
PHILLIPS	NICOLE	**1 AUG /AOÛ 73**	F	BARNSLEY
MOSS	JESSICA	**2 AUG /AOÛ 73**	F	TRAFFORD
DRISCOLL	STEPHANIE	**3 AUG /AOÛ 73**	F	OLDHAM
RUSSELL	LIAM	**20 AUG /AOÛ 73**	M	TELFORD
GIBSON	KIERAN	**21 AUG /AOÛ 73**	M	WARRINGTON
WALLACE	CRAIG	**22 AUG /AOÛ 73**	M	GATESHEAD
GORDON	KYLE	**23 AUG /AOÛ 73**	M	DUNDEE
MARSHALL	JAMIE	**01 JUN /JUIN 73**	M	STOCKTON-ON-TEES
STEVENSON	DYLAN	**02 JUN /JUIN 73**	M	RICHMOND UPON THAMES
SUTHERLAND	ADAM	**01 JUN /JUIN 73**	M	PRESTON
CRAIG	NATHAN	**21 FEB /FEV 73**	M	PETERBOROUGH
FRASER	**JOSHUA**	4 AUG /AOÛ 93	**M**	ABERDEEN
GRAY	**SHAUN**	5 AUG /AOÛ 93	**M**	CAMDEN
DAVIDSON	**DARREN**	6 AUG /AOÛ 93	**M**	ROTHERHAM
MCDONALD	**SAMUEL**	7 AUG /AOÛ 93	**M**	NORTHAMPTON
HENDERSON	**PETER**	8 AUG /AOÛ 93	**M**	ROCHDALE
KERR	**DEAN**	9 AUG /AOÛ 93	**M**	SWINDON

Surname	Given Names	Date of Birth	Sex	Place of Birth
SIMPSON	**JONATHAN**	10 AUG /AOÛ 93	**M**	BLACKPOOL
FERGUSON	**LUKE**	11 AUG /AOÛ 93	**M**	ISLINGTON
MCKENZIE	**ZOE**	05 JUN /JUIN 93	**F**	NEWPORT
KENNEDY	**NATASHA**	06 JUN /JUIN 93	**F**	NEW FOREST
BURNS	**LISA**	09 JUN /JUIN 93	**F**	BATH
MUIR	**BETHANY**	12 JUN /JUIN 93	**F**	ST. HELENS
BURTON	**CLAIRE**	03 JUN /JUIN 93	**F**	BASILDON
WATT	**NICOLA**	04 JUN /JUIN 93	**F**	COLCHESTER
MACMILLAN	**GEMMA**	43 JUN /JUIN 93	**F**	DARLINGTON
ONEILL	**VICTORIA**	44 JUN /JUIN 93	**F**	PETERBOROUGH

References

Anastasi, J. S., & Rhodes, M. G. (2006). Evidence for an own-age bias in face recognition. *North American Journal of Psychology, 8*, 237–252.

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

Arcurio, L. R., Gold, J. M., & James, T. W. (2012). The response of face-selective cortex with single face parts and part combinations. *Neuropsychologia*, *50*(10), 2454-2459.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*(1), 54-75.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning the effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241-248.

Barton, J. J. (2009). What is meant by impaired configural processing in acquired prosopagnosia? *Perception*, *38*(2), 242-260.

Bate, S. Parris, B. Haslam, C. & Kay, J. (2010). Socio-emotional functioning and face recognition ability in the normal population. *Personality and Individual Differences*, *48*(2), 239-242.

Behrmann, M. Avidan, G. Marotta, J. J. & Kimchi, R. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioural findings. *Journal of Cognitive Neuroscience*, *17*(7), 1130-1149.

Bellugi, U., Klima, E. S., & Wang, P. P. (1996). Cognitive and neural development: Clues from genetically based syndromes. In D. Magnussen (Ed.), *The life-span development of individuals: Behavioral, neurobiological, and psychosocial perspectives* (pp. 223–243). The Nobel Symposium. New York, NY: Cambridge University Press

Bellugi, U., Lichtenberger, L., Mills, D., Galaburda, A., & Korenberg, J. R. (1999). Bridging cognition, the brain and molecular genetics: evidence from Williams syndrome. *Trends in Neurosciences*, *22*(5), 197-207.

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & George, M. S. (2000). I. The neurocognitive profile of Williams Syndrome: a complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, *12* (Supplement 1), 7-29.

Benton, A., Hamsher K., Varney N. R. & Spreen O. (1983) Benton Test of Facial Recognition. Oxford University Press, New York.

Bevan, T., (Producer), Fellner E., (Producer), Hayward, D., (Producer), Mackintosh, C., (Producer), & Hooper (Director). (2012) *Les Miserables* [Motion Picture]. United States: Universal Pictures.

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

Bindemann, M., Sandford, A., Gillatt, K., Avetisyan, M., & Megreya, A. M. (2012). Recognising faces seen alone or with others: Why are two heads worse than one? *Perception, 41*(4), 415-435.

Bindemann, M. & Lewis, M. B. (2013). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin & Review*, *20*(6), 1140-1145.

Bindemann, M., Attard, J., & Johnston, R. A. (2014). Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychology*, *1*(1) 986903.

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016a). Solving the border control problem: evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PloS ONE*, *11*(2), e0148148.

Bobak, A. K., Hancock, P. J., & Bate, S. (2016b). Super-recognisers in Action: Evidence from Face-matching and Face Memory Tasks. *Applied Cognitive Psychology*, *30*(1), 81-91.

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016c). An in-depth cognitive examination of individuals with superior face recognition skills. *Cortex*, *82*, 48-62.

Boehm, S. G., Klostermann, E. C., Sommer, W., & Paller, K. A. (2006). Dissociating perceptual and representation-based contributions to priming of face recognition. *Consciousness and cognition*, *15*(1), 163-174.

Bombari, D., Mast, F. W., & Lobmaier, J. S. (2009). Featural, configural, and holistic face-processing strategies evoke different scan patterns. *Perception*, *38*(10), 1508-1521.

Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and spatial frequency information. *Perception & Psychophysics*, *65*(7), 1078-1093.

Brants, M., Wagemans, J., & de Beeck, H. P. O. (2011). Activation of fusiform face area by Greebles is related to face similarity but not expertise. *Journal of Cognitive Neuroscience*, *23*(12), 3949-3958.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, *62*(3), 311-318.

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

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

Bruce, V., Ellis, H. D., Gibling, F., & Young, A. (1987). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *41*(4), 510-520

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

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

Bruce, V., & Young, A. W. (2012). Face perception. Psychology Press.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*(06), 547-558.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*(3), 361-380.

Burton, A. M., & Bruce, V. (1992). I recognize your face but I can't remember your name: A simple explanation? *British Journal of Psychology*, *83*(1), 45-60.

Burton, A. M., & Bruce, V. (1993). Naming faces and naming names: Exploring an interactive activation model of person recognition. *Memory*, *1*(4), 457-480.

Burton, A. M., Kelly, S. W., & Bruce, V. (1998). Cross-domain repetition priming in person recognition. *Quarterly Journal of Experimental Psychology*, *51A*(3), 515–529.

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

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

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*(1), 286-291.

Burton, A. M., (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1467-1485.

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), 482-496.

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, *40*(1), 202-223.

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

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *Journal of Cognitive Neuroscience*, *17*(10), 1652-1666.

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological bulletin, 56(2), 81–105.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*(4275), 312-314.

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90-98.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Carson, D. R., & Burton, A. Mike. (2001). Semantic priming of person recognition: Categorical priming may be a weaker form of the associative priming effect. *The Quarterly Journal of Experimental Psychology: Section A, 54*(4), 1155-1179.

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. *Journal of Experimental Psychology: Applied, 15*(1), 76

Christie, J., Ginsberg, J. P., Steedman, J., Fridriksson, J., Bonilha, L., & Rorden, C. (2012). Global versus local processing: seeing the left side of the forest and the right side of the trees. *Frontiers in Human Neuroscience*, *6*, 28.

Civil Aviation Authority. (2014). Summary of Reporting at UK Airports 2014. Accessed 29/02/2016. Retrieved from: http://www.caa.co.uk/Data-and-analysis/UK-aviation-market/Airports/Datasets/UK-Airport-data/Airport-data-1990-onwards/

Clutterbuck, R. & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*(8), 985-994.

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

Collishaw, S. M., Hole, G. J., & Schwaninger, A. (2005). Configural processing and perceptions of head tilt. *Perception*, *34*(2), 163-168.

Courtois, M. R. & Mueller, J. H. (1981). Target and distractor typicality in facial recognition? *Journal of Applied Psychology*, *66*(5), 639-645.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience*, *13*(1), 89-109.

Davidoff, J., & Landis, T. (1990). Recognition of unfamiliar faces in prosopagnosia. *Neuropsychologia*, *28*(11), 1143-1161.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *85*.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23*(4), 482–505.

Davis, J. P., & Valentine, T. (2015). Human verification of identity from photographic images. Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV, 209-238.

Davis, J., Lander, K., Evans, R., & Jansari, A. (In Press) Investigating predictors of superior face recognition ability in police super-recognisers. *Applied Cognitive Psychology*. (Online).

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587-605.

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

Dimitriou, D., Leonard, H. C., Karmiloff-Smith, A., Johnson, M. H., & Thomas, M. S. (2015). Atypical development of configural face recognition in children with autism, Down's syndrome and Williams syndrome. *Journal of Intellectual Disability Research, 59*(5), 422-438.

Donnai, D., & Karmiloff-Smith, A. (2000). Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics, 97*(2), 164-171.

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

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *The Quarterly Journal of Experimental Psychology*, *69*(1), 1-10.

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification, 56*(4), 600.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74-78.

Duchaine, B. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia, 41*(6), 713-720.

Duchaine, B. C., & Nakayama, K. (2004). Developmental prosopagnosia and the Benton Facial Recognition test. *Neurology, 62*(7), 1219-1220.

Duchaine, B., & Nakayama, K. (2006a). 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*(4), 576-585.

Duchaine, B. C. & Nakayama, K. (2006b). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, *16*(2), 166-173.

Duchaine, B. Germine, L. & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*(4), 419-430.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, *1*, 393-416.

Edgin, J. O., & Pennington, B. F. (2005). Spatial cognition in autism spectrum disorders: Superior, impaired, or just intact? *Journal of Autism and Developmental Disorders*, *35*(6), 729-745.

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

Ellis, H. D., & Florence, M. (1990) Bodamer's (1947) paper on prosopagnosia. *Cognitive Neuropsychology*, 7(2), 81-105.

Enlow, D. H. (1982) *Handbook of Facial Growth 2nd edition.* Philadelphia, PA: W B Saunders Company.

Experian, PKF Littlejohn & University of Portsmouth. (2016.) Annual Fraud Indicator 2016. Accessed 11/08/2016. Retrieved from: http://www.port.ac.uk/media/contacts-and-departments/icjs/ccfs/Annual-Fraud-Indicator-2016.pdf

Facial Identification Scientific Working Group. (2012). Recommendations for a Training Program in Facial Comparison. Accessed 12/08/2016. Retrieved from: https://www.fiswg.org/document/viewDocument?id=25

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

Farah, M. J., Levinson, K. L., & Klein, K. L. (1995b). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia*, *33*(6), 661-674.

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

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is" special" about face perception? *Psychological review*, *105*(3), 482.

Federal Bureau of Investigation. (2016). ‘FBI Budget Request for the fiscal year 2017’ James. B Comey, Federal Bureau of Investigation, News, Testimony. Accessed

29/02/2016. Retrieved from https://www.fbi.gov/news/testimony/fbi-budget-request-for-fiscal-year-2017.

Fender, M. (2008). Spelling Knowledge and Reading Development: Insights from Arab ESL Learners. *Reading in a Foreign Language*, *20*(1), 19-42.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.

Fielding, G., Gummadidala, K., Kam, M., & Conn, R. (2001). Signature authentication by forensic document examiners. *Journal of Forensic Science*, *46*(4), 884-888.

Finger, S. (2001) *Origins of Neuroscience*: A History of Explorations into Brain Function. Oxford: Oxford University Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627-635.

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

Frontex. (2015). Annual Risk Analysis. Accessed 11/08/2016. Retrieved from: http://frontex.europa.eu/assets/Publications/Risk_Analysis/Annual_Risk_Analysis_2015.pdf

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 104-115.

Galli, G., Feurra, M., & Viggiano, M. P. (2006). "Did you see him in the newspaper?" Electrophysiological correlates of context and valence in face processing. *Brain research*, *1119*(1), 190-202.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*(6), 568-573.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191-197.

Gauthier, I. Curby, K. M. Skudlarski, P. & Epstein, R. A. (2005). Individual differences in FFA activity suggest independent processing at different spatial scales. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(2), 222-234.

George, P.A. & Hole, G.J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception, 24,* 1059–1073.

George, P. A., & Hole, G. J. (2000). The role of spatial and surface cues in the age-processing of unfamiliar faces. *Visual Cognition*, *7*(4), 485-509.

Gilad-Gutnick, S., Yovel, G., & Sinha, P. (2012). Recognizing degraded faces: The contribution of configural and featural cues. *Perception*, *41*(12), 1497-1511.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*(1), 32-41

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

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., & Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, *81*, 1-11.

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological Science*, *23*(4), 427-434.

Göritz, A. S. (2007). Using online panels in psychological research. *The Oxford handbook of Internet psychology*, 473-485.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception*, *42*(11), 1158-1165.

Hall, R., Dodds, G., & Triggs, S. (1993). *The World of William Notman*. David R. Godine.

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

Harel, A. (2016). What is special about expertise? Visual expertise reveals the interactive nature of real-world object recognition. *Neuropsychologia*, *83*, 88-99.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223-233.

Her Majesty’s Passport Office. (2013). Her Majesty’s Passport Office, Business Plan 2013- 14, 2013. Accessed 29/02/2016. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/204928/Tagged_IPS_bp_v25_WEB.pdf.

Her Majesty’s Passport Office. (2014). Her Majesty's Passport Office Annual Report and Accounts 2013-14. Accessed 29/02/2016. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/332680/HMPO_ARA_2013-14_v5_1_FINAL_WEB.pdf

Heyer, R. L. (2013). *Understanding One-to-Many Unfamiliar Face Matching in the Operational Context : The Impact of Candidate List Size, Expertise, and Decision Aids on the Performance of Facial Recognition System Users*. University of Adelaide.

Heyer, R. L., & Semmler, C. (2013). Forensic confirmation bias: the case of facial image comparison. *Journal of Applied Research in Memory and Cognition, 2*, 68-70.

Hildebrandt, A. Wilhelm, O. Schmiedek, F. Herzmann, G. & Sommer, W. (2011). On the specificity of face cognition compared with general cognitive functioning across adult age. *Psychology and Aging, 26*(3), 701-715.

Hill, K., Barton, M., & Hurtado, A. M. (2009). The emergence of human uniqueness: Characters underlying behavioral modernity. *Evolutionary Anthropology: Issues, News, and Reviews, 18*(5), 187-200.

Hoffman, C. & Kagan, S. (1977). Field dependence and facial recognition. *Perceptual and motor skills*, *44*(1), 119-124.

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

Home Office. (2016). Home Secretary launches new joint fraud taskforce. Accessed 04/08/2016. Retrieved from: https://www.gov.uk/government/news/home-secretary-launches-new-joint-fraud-taskforce

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

House of Commons, Committee of Public Accounts. (2007). Identity and Passport Service: Introduction of ePassports. Forty-ninth Report of Session 2006-07. Accessed 24/08/2016. Retrieved from: http://www.publications.parliament.uk/pa/cm200607/cmselect/cmpubacc/362/362.pdf

Huang, L. Song, Y. Li, J. Zhen, Z. Yang, Z. & Liu, J. (2014). Individual differences in cortical face selectivity predict behavioral performance in face recognition. *Frontiers in Human Neuroscience*, *8*, 483.

Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *Behavioral and Brain Sciences, 8*(04), 689-705.

Hunt, T. (1928). The measurement of social intelligence. *Journal of Applied Psychology*, *12*(3), 317–334.

International Civil Aviation Organisation. (2015). Document 9303. Machine Readable Travel Documents, Seventh Edition. Accessed 13/08/2015. Retrieved from: http://www.icao.int/publications/pages/publication.aspx?docnum=9303

Jenkins, R., Lavie, N., & Driver, J. (2003). Ignoring famous faces: Category-specific dilution of distractor interference. *Perception & Psychophysics*, *65*(2), 298–309.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323.

John, O. P. Donahue, E. M. & Kentle, R. L. (1991). *The Big Five Inventory – Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

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

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *4*(2), 89-108.

Kagan, J. (1965). Reflection-impulsivity and reading ability in primary grade children. *Child Development*, 609-628.

Kaltwasser, L. Hildebrandt, A. Recio, G. Wilhelm, O. & Sommer, W. (2014). Neurocognitive mechanisms of individual differences in face cognition: a replication and extension. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 861-878.

Kam, M., Fielding, G., & Conn, R. (1997). Writer identification by professional document examiners. *Journal of Forensic Sciences, 42,* 778–786.

Kam, M., Abichandani, P., & Hewett, T. (2015). Simulation Detection in Handwritten Documents by Forensic Document Examiners. *Journal of Forensic Sciences*, *60*(4), 936-941.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience.* 17, 4302–4311.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*, 759-763.

Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences, 107*(25), 11163-11170.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *361*(1476), 2109-2128.

Karbach, J., & Verhaeghen, P. (2014). Making working memory work a meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, *25*(11), 2027-2037.

Karmiloff-Smith, A. (1997). Crucial differences between developmental cognitive neuroscience and adult neuropsychology. *Developmental Neuropsychology*, *13*(4), 513-524.

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

Kihlstrom, J.F., & Cantor, N. (2000). "Social intelligence," in *Handbook of Intelligence*, Ed R. J .Sternberg (Cambridge: Cambridge University Press), 359-379.

Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory & Cognition*, *12*(1), 60-70.

Klein, R. G. (1995). Anatomy, behavior, and modern human origins. *Journal of World Prehistory*, *9*(2), 167-198.

Konar, Y. Bennett, P. J. & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, *21*(1), 38-43.

Krakowski, C. S., Poirel, N., Vidal, J., Roëll, M., Pineau, A., Borst, G., & Houdé, O. (2016). The forest, the trees, and the leaves: Differences of processing across development. *Developmental Psychology*, *52*(8), 1262.

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

Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review*, *85*(4), 278

Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behavior*, *38*(3), 256.

Lander, K., & Bruce, V. (2003). The role of motion in learning new faces. *Visual Cognition*, *10*(8), 897-912.

Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological science*, *14*(5), 510-515.

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

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

Leslie, L., & Thimke, B. (1986). The use of orthographic knowledge in beginning reading. *Journal of Literacy Research*, *18*(3), 229-241.

Levy, B. A., Newell, S., Snyder, J., & Timmins, K. (1986). Processing changes across reading encounters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(*4), 467-478.

Levy, B. A., Di Persio, R., & Hollingshead, A. (1992). Fluent rereading: Repetition, automaticity, and discrepancy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 957-971.

Li, J. Tian, M. Fang, H. Xu, M. Li, H. & Liu, J. (2010). Extraversion predicts individual differences in face recognition. Communicative & Integrative Biology, 3(4), 295-298.

Lowry, P. B., D'Arcy, J., Hammer, B., & Moody, G. D. (2016). "Cargo Cult" science in traditional organization and information systems survey research: A case for using nontraditional methods of data collection, including Mechanical Turk and online panels. The Journal of Strategic Information Systems, 25(3), 232-240.

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

Marotta, J. J., McKeeff, T. J., & Behrmann, M. (2002). The effects of rotation and inversion on face processing in prosopagnosia. *Cognitive Neuropsychology*, *19*(1), 31-47

Martens, M. A., Wilson, S. J., & Reutens, D. C. (2008). Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of Child Psychology and Psychiatry*, *49*(6), 576-608.

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

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

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

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

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

Megreya, A. M., & Havard, C. (2011). Left face matching bias: Right hemisphere dominance or scanning habits? *Laterality*. *16*(1), 75-92.

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology*, *25*(1), 30-37.

Menneer, T., Stroud, M. J., Cave, K. R., Li, X., Godwin, H. J., Liversedge, S. P., & Donnelly, N. (2012) Search for two categories of target produces fewer fixations to target-color items. *Journal of Experimental Psychology: Applied, 18(4),* 404-418.

Mercure, E., Dick, F., & Johnson, M. H. (2008). Featural and configural face processing differentially modulate ERP components. *Brain Research*, *1239*, 162-170.

Messick, S. & Damarin, F. (1964). Cognitive styles and memory for faces. *The Journal of Abnormal and Social Psychology*, *69*(3), 313.

Millis, S. R., Malina, A. C., Bowers, D .A. & Ricker, J .H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale-III. *Journal of Clinical & Experimental Neuropsychology* 21, 87–93.

Milne, E. & Szczerbinski, M. (2009). Global and local perceptual style, field-independence, and central coherence: An attempt at concept validation. *Advances in Cognitive Psychology*, *5*, 1.

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

Mooney, C. M. (1957). Age in the development of closure ability in children .*Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *11*(4), 219-226.

Moore, R. M., & Johnston, R. A. (2013). Motivational incentives improve unfamiliar face matching accuracy. *Applied Cognitive Psychology*, *27*(6), 754-760.

Moyse, E. (2014). Age Estimation from Faces and Voices: A Review. *Psychologica Belgica*, *54*(3), 255-265

Mueller, J. H. Bailis, K. L. & Goldstein, A. G. (1979). Depth of processing and anxiety in facial recognition. *British Journal of Psychology*, *70*(4), 511-515.

Navon, D. (1977) Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.

Nickerson, R. S. (1965). Response times for" same"-" different" judgments. *Perceptual and Motor Skills*, *20*, 15-18.

Nowicki, S. Winograd, E. & Millard, B. A. (1979). Memory for faces: A social learning analysis. *Journal of Research in Personality*, *13*(4), 460-468.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory & Cognition*, *26*(1), 146-160.

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

Office for National Statistics (2013), Table PPO1UK, 2011 Census: Population Estimates by single year of age and sex for Local Authorities in the United Kingdom. Accessed 21/11/2015. Retrieved from: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-294277

Papesh, M. H., & Goldinger, S. D. (2014). Infrequent identity mismatches are frequently undetected. Attention, Perception, & Psychophysics, 76(5), 1335-1349.

Parkinson, J. (2015) BBC News Magazine. Accessed on the 20/05/2016. Retrieved from: http://www.bbc.co.uk/news/magazine-30988833

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, *116*(2), 220-244.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(4), 406.

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition*, *4*(4), 409-438.

Poirel, N., Krakowski, C. S., Sayah, S., Pineau, A., Houdé, O., & Borst, G. (2014). Do you want to see the tree? Ignore the forest. *Experimental Psychology, 61,* 205-214.

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

Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, *4*(2), 4.4.

Renzi, C., Schiavi, S., Carbon, C. C., Vecchi, T., Silvanto, J., & Cattaneo, Z. (2013). Processing of featural and configural aspects of faces is lateralized in dorsolateral prefrontal cortex: a TMS study. *Neuroimage, 74*, 45-51.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape–based model of vision. *Proceedings of the Royal Society of London B: Biological Sciences, 271* (Suppl 6), S448-S450.

Ritchie, K. L., & Burton, A. M. (2016). Learning faces from variability. *The Quarterly Journal of Experimental Psychology,* 1-9.

Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. *Cognition*, *141,* 161-169.

Robertson, D. J. Noyes, E. Dowsett, A. Jenkins, R. & A. Mike Burton. (2016). Face Recognition by Metropolitan Police Super-Recognisers. *PLoS ONE 11(2)*: e0150036

Rose, S. A. & Feldman, J. F. (1995). Prediction of IQ and specific cognitive abilities at 11 years from infancy measures. *Developmental Psychology, 31*, 685-696.

Rose, S. A. Feldman, J. F. & Jankowski, J. J. (2003). Infant visual recognition memory: Independent contributions of speed and attention. *Developmental Psychology, 39*, 563-571.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252-257.

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, *32*(3), 285-293.

Scheck B., Neufeld P. & Dwyer J. (2000) *Actual Innocence*. New York: Random House.

Schiffer, B., & Champod, C. (2007). The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation. *Forensic Science International*, *167*(2), 116-120.

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*(3), 405-410.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402-409.

Scott, L. S., & Nelson, C. A. (2006). Featural and configural face processing in adults and infants: A behavioral and electrophysiological investigation. *Perception, 35*(8), 1107-1128.

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

Sergent, J. (1984). Configural processing of faces in the left and the right cerebral hemispheres. *Journal of Experimental Psychology: Human Perception and Performance, 10*(4), 554-572.

Sergent, J., Ohta, S. & MacDonald, B. (1992). Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain, 115*(1), 15–36.

Shakeshaft, N. G. & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences, 112(41),* 12887-12892.

Smalarz, L., Madon, S., Yang, Y., Guyll, M., & Buck, S. (2016). The Perfect Match: Do Criminal Stereotypes Bias Forensic Evidence Analysis? *Law and Human Behav*ior, 40(4), 420–429.

Susilo, T., Yovel, G., Barton, J. J., & Duchaine, B. (2013a). Face perception is category-specific: Evidence from normal body perception in acquired prosopagnosia. *Cognition*, *129*(1), 88-94.

Susilo, T., & Duchaine, B. (2013b). Advances in developmental prosopagnosia research. *Current opinion in neurobiology*, *23*(3), 423-429.

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

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

Tanaka, J. W., Kaiser, M. D., Hagen, S., & Pierce, L. J. (2014). Losing face: impaired discrimination of featural and configural information in the mouth region of an inverted face. *Attention, Perception, & Psychophysics*, *76*(4), 1000-1014.

Tanaka, J. W., & Sung, A. (2016). The "eye avoidance" hypothesis of autism face processing. *Journal of autism and developmental disorders*, *46*(5), 1538-1552.

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013). Expertise in fingerprint identification. *Journal of Forensic Sciences*, *58*(6), 1519-1530.

Thorndike, E. L. (1920). A constant error in psychological ratings. Journal of Applied Psychology, 4(1), 25-29.

Tom, G., Pettersen, P., Lau, T., Burton, T., & Cook, J. (1991). The role of overt head movement in the formation of affect. *Basic and Applied Social Psychology*, *12*, 281–289.

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

UK Border Agency. (2013). UK Border Agency Annual Reports and Accounts 2012-2013. Accessed 29/02/2016. Retrieved from: https://www.gov.uk/government/publications/uk-border-agency-annual-report-and-accounts-2012-to-2013--2

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352.

Valentine, T., & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica*, *61*(3), 259-273.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*(4), 471-491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43(2), 161-204.

Valentine, T. (2001). Face-space models of face recognition. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*, 83-113.

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes. *Cognitive Science*, *28*(2), 289-301.

Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age, *Psychology and Aging, 25*, 265– 277.

Wang, P. P., & Bellugi, U. (1994). Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology*, *16*(2), 317-322.

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169-177.

Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. British Journal of Psychology, 73, 117-129.

Warrington, E. K. (1984). *Recognition Memory Test: Rmt. (Words). Test Booklet 1*. NFER: Nelson Publishing Company.

Warrington, E.K. & James, M. (1991). *The Visual Object and Space Perception Battery.* Bury St. Edmunds: Thames Valley Test Company.

Wechsler, D. (1997). Wechsler Memory Scale—Third Edition (WMS-III) Administration and Scoring Manual. San Antonio, TX: NCS Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale.* 4th Edn. San Antonio, TX: The Psychological Corporation.

Weibert, K. & Andrews, T. J. (2015). Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces. *Neuropsychologia*, *75*, 588-596.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, *36*(3), 1060-1084.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2013). Face recognition deficits in autism spectrum disorders are both domain specific and process specific. PLoS ONE, *8*(9), e74541

Wells, G. L., & Petty, R. E. (1980). The effects of overt head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology, 1,* 219–230.

Wells G.L., Small, M., Penrod, S.D., Malpass, R.S., Fulero, S.M. & Brimacombe C.A.E. (1998). Eyewitness identification procedures: recommendations for line-ups and photospreads. *Law and Human Behavior, 22*(6), 603–7.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*(1), 277-295.

White, D., Kemp, R. I. Jenkins, R. Matheson, M. & Burton, A. M. (2014a). Passport Officers' Errors in Face Matching. *PloS ONE, 9*(8), 1-8.

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014b). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, *20*(2), 166-173.

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014c). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100-106.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. In *Proc. R. Soc. B* (Vol. 282, No. 1814, p. 20151292). The Royal Society.

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*(3), 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 & Justice*, *49*(3), 191-196

Wilmer, J. B., Germine, L. T., & Nakayama, K. (2014). Face recognition: a model specific ability. *Frontiers in Human Neuroscience*, *8*, 769.

Wilson, B. A. Evans, J. J. Alderman, N. Burgess, P. W. & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. *Methodology of Frontal and Executive Function*, 239-250.

Witkin, H. A. & Goodenough, D. R. (1977). Field dependence revisited. *ETS Research Bulletin Series*, *2*, i-53.

Wolf, J. M., Tanaka, J. W., Klaiman, C., Cockburn, J., Herlihy, L., Brown, C., ... & Schultz, R. T. (2008). Specific impairment of face-processing abilities in children with autism spectrum disorder using the Let's Face It! skills battery. *Autism Research*, *1*(6), 329-340.

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

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, *136*(4), 623.

Woodhead, M. M. & Baddeley, A. D. (1981). Individual differences and memory for faces, pictures, and words. *Memory & Cognition*, *9*(4), 368-370.

World Health Organization. (1992). The ICD-10 Classifications of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organization.

Yarmey, A. D. (1973). I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory & Cognition*, *1*(3), 287-290.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141-145.

Young, A. W., McWeeny, K. H., Ellis, A. W., & Hay, D. C. (1986). Naming and categorizing faces and written names. *The Quarterly Journal of Experimental Psychology*, *38*(2), 297-318.

Young, A. W. Hellawell, D. & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *42*(11), 1166-1178.

Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron*, *44*(5), 889-898.

Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *18*(4), 580-593.

Yovel, G., & Kanwisher, N. (2008). The representations of spacing and part-based information are associated for upright faces but dissociated for objects: Evidence from individual differences. *Psychonomic Bulletin & Review*, *15*(5), 933-939.

Yovel, G. Wilmer, J. B. & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, *8*, 562.