Theoretical and real-world applications of superior face recognition

Anna Katarzyna Bobak

A thesis submitted in partial fulfilment of the requirements of Bournemouth University for the degree of Doctor of Philosophy

July 2016

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr Sarah Bate, for her on-going support and believing in me when things got tough. I would also like to extend my thanks to Dr Ben Parris and Professor Changhong Liu for their advice throughout my studies. I also offer my sincere appreciation to Bournemouth University for giving me this incredible opportunity to pursue a career as a researcher. My special thanks are extended to all participants -this project would have not been possible without their help and dedication.

My deepest gratitude goes to my Mum and Grandparents- thank you, not only for your unconditional love and support throughout my studies, but also for bringing me up who I am today. I love you. To my cousin Gabi- thank you for all the chats and for being the voice of common sense when I needed it the most.

My warmest thanks go to my friend, Kinga. Thank you for the long chats, visits, and immediate support when things were not looking up, but also for celebrating all the achievements with me. To Magda, Mikael, and Milton, thank you for being there for me and never stopping to believe that I have got what it takes. To my dear friends Evaggelia, Dina, and George, thank you for your unconditional support and for making me feel like a part of your family.

My completion of this thesis would not have been possible without the caring friendships of those up north: Ola, Dagmara, Patryk, Ania, Clare, Ailsa, and Emma- thank you for being there for me along the way. To Jenny, Jamie, and Helen- thank you for keeping me sane at home and being such a positive influence in my life. I would also like to thank Kate and Charlie for their friendship and care. Furthermore, I would like to thank

Becky, Ruth, and Andy for their warm support and for motivating me to swim, bike, and run in the past year.

My friends, Simon, Nabil, Abby, Michele, Jess, Lesley, Becca, Sarah, and Emmy deserve a special mention. You made the PhD journey much more enjoyable and less lonely. I am also extremely grateful to Martin and Jamie for their continuous support with my project.

Last but not least, I would like to thank the late Dr Robbie Cooper for sparking my interest in faces and for inspiring me to become a scientist. This one is for you.

Copyright Statement

Abstract

While previous work has identified the existence of people with extraordinary face recognition skills (so-called "super-recognisers"; SRs), the cognitive and perceptual underpinnings of the ability are unknown. This thesis addresses this issue, using behavioural and eye-movement measures. It also evaluates the methods used to identify SRs, their role in more applied national security settings, and ways of improving face recognition in typical perceivers. The first set of studies offers an in-depth cognitive and perceptual examination of six SRs using a case-series approach. This investigation revealed that while SRs are a heterogeneous group, they consistently show enhanced holistic processing. A second set of studies examined the eye-movements of SRs in a standard face memory task and a more ecologically valid free-viewing task. In both experiments SRs spent more time looking at the nose (i.e. the centre of faces) than typical perceivers, countering previous work that suggests the eye region is critical in facial identification. A subsequent study was aimed at establishing the UK-specific norms for dominant tests of face recognition and face perception, using a large sample of young British adults. Results suggested that females are better at face recognition than males, and that country-specific control norms are needed for these neuropsychological tests. A fourth set of studies looked at the performance of SRs on more applied face recognition tasks, replicating face matching and recognition scenarios. Results strongly suggested that some SRs are best-suited to particular tasks, and when identified correctly would make extremely valuable employees in national security settings. A final study examined if face matching and face recognition skills can be improved in typical perceivers via intranasal inhalation of the nonapeptide oxytocin, yet neither process was improved following this intervention. The theoretical and practical implications resulting from all these investigations are discussed, particularly in relation to our understanding of the typical face-processing system, and in making practical recommendations for the implementation of super recognition in national security settings.

Table of Contents

List of Figures

List of Tables

Author's Declaration

I hereby declare that the work presented in this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Chapter 1: Introduction

Human faces convey an array of socially salient information such as identity, gender, and emotional state. The ability to extract this information is critical for appropriate social functioning. While most people have similar levels of experience with faces, there are still considerable individual differences in their ability to recognise facial identity (e.g. Bate, Parris, Haslam, & Kay, 2010; Bowles et al. 2009). These differences range from individuals who are remarkably good at face recognition (so-called "super recognisers", SRs: Russell, Duchaine, & Nakayama, 2009; Bobak, Hancock, & Bate, 2016) to those affected by developmental prosopagnosia (DP). This latter group of people experience severe difficulties in face recognition, in the absence of neurological damage or illness, lower-level visual or intellectual impairments, and concurrent socio-emotional difficulties (Bate & Cook, 2012; Bate et al., 2014a; Jones & Tranel, 2001; Susilo & Duchaine, 2013).

While a considerable amount of research has examined the correlates of face recognition in both the typical population (e.g., Bowles et al., 2009; Wilmer et al., 2010) and those with face recognition deficits (e.g., Barton, 2008; Behrmann, Avidan, Marotta, & Kimchi, 2005; Le Grand et al., 2006), comparatively little work has focused on the upper end of the face recognition spectrum by examining SRs. The term was first coined by Russell et al. (2009) who identified four people with extraordinary face recognition skills. This group of individuals outperformed control participants on tests of face memory, face perception, and familiar face recognition. However, very little subsequent work has been published on super recognition, and it is unknown whether the superior abilities of SRs extend beyond facial identity processing, nor have the underlying mechanisms of super recognition been identified. From a more applied perspective, it is clear that SRs could potentially be invaluable in policing and national security scenarios. Yet, no published work to date has explored this possibility, and it is likely that developments in our theoretical understanding of super recognition are required before their more applied potential can be realized. This thesis sets out to address these issues.

1. THE FACE RECOGNITION SPECTRUM

There are large individual differences in the ability to recognize (Bowles et al., 2010; Russell et al., 2009) and perceive (Megreya & Bindemann, 2013; Megreya & Burton, 2006) faces, and particular difficulties are associated with the processing of unfamiliar facial stimuli (see Hancock, Bruce, & Burton, 2000 for a review). A large body of evidence suggests that familiar and unfamiliar faces are processed in a qualitatively different manner. For instance, some people with acquired prosopagnosia are impaired at familiar face recognition but can match unfamiliar faces (e.g., Bauer, 1984; Benton & Van Allen, 1972; Bruyer et al., 1983; Tranel, Damasio, & Damasio, 1988), while others can recognise familiar faces, but are unable to match images of unfamiliar faces (e.g., Young, Newcombe, DeHaan, Small, & Hay, 1993). Malone, Morris, Kay and Levin (1982) identified two patients with damage to occipital areas of the brain. Over time familiar face recognition improved in one patient, but he remained impaired at unfamiliar face matching. Conversely, the other patient remained unable to recognise familiar faces, but was able to match faces that were unfamiliar.

Familiar and unfamiliar faces are also associated with different regional sampling. Ellis, Shepherd, and Davies (1979) found that while familiar face recognition was most efficiently achieved via sampling of the internal facial regions (i.e. the eyes, nose and mouth), unfamiliar face recognition was equally as accurate when the internal or the external features were examined. More recently, Megreya and Burton (2006) examined the recognition of upright and inverted familiar and unfamiliar faces. They found a strong correlation between the matching of familiar and unfamiliar faces, but only when the familiar faces were inverted. These findings suggest that the two types of facial stimuli are processed in a qualitatively different manner.

In addition to the qualitative differences in familiar and unfamiliar face processing, the matching and recognition of unfamiliar faces is much less stable than it is for familiar faces. Viewpoint, expression, and context (for a review see Johnston & Edmonds, 2009) all have a considerably detrimental effect on unfamiliar face recognition, but only marginally affect the recognition of familiar faces. This phenomenon can be explained by recent research suggesting that more stable representations of an unfamiliar face can be established following increased exposure to a variety of images of that person (e.g. Burton, Kramer, Ritchie, & Jenkins, 2016; Ritchie & Burton, In press). These more detailed representations may be immune to the detrimental effect of the changes listed above, whereas more limited representations of unfamiliar faces may largely rely on pictorial information that prevents recognition under novel viewing conditions.

Unfamiliar face recognition also seems to be related to a number of personality factors. For instance, empathy has been found to be positively related to accuracy of face recognition (Bate et al., 2010). Specifically, Bate and colleagues reported that people high in empathy are better at recognising newly learned faces than those who score low on the empathy scale. This may be because the additional information about others' emotional state aids their encoding of new faces. Nonetheless, it is also possible that the naturally occurring high empathy makes these individuals allocate more attention to faces. Furthermore, Hills, Eaton, and Pake (2016) reported that psychometric schizotypy, a cluster of traits related to difficulties in social situations (e.g. anxiety), is negatively related to face recognition accuracy. In addition, three studies have shown a direct link between general anxiety and face recognition. In the first report (Mueller, Bailiss, & Golstein, 1979), the authors divided participants into low and high anxiety groups and reported that those low in anxiety performed better in a face recognition task. Furthermore, Megreya and Bindemann (2013) demonstrated that neuroticism and anxiety are negatively correlated with face matching ability, but only in female observers. Another study by Davis, McKone, Dennett et al. (2011) investigated the relationship between social and trait anxiety and the Cambridge Face Memory Test (CFMT) and found that poorer performance on the CFMT was correlated with a significant increase in participants' social, but not trait anxiety. The authors argued that this suggests successful facial recognition is crucial for social interactions. Indeed, low performance on the CFMT may be related to humans' perceptual learning, where a particular skill is developed with gradual exposure to (and increased exposure with) a stimulus. As such, frequent social interaction would lead to increased expertise in the within-category discrimination of faces, whilst individuals who do not have that expertise in facial recognition due to increased social anxiety and avoidance of social situations are naturally more likely to underperform at a face recognition task. This account can be supported by studies showing that gregariousness is related to individuals' face recognition ability through exposure (Arnell & Dube, 2015; Li, Tian, Fang et al., 2010) and that, conversely, shy children are less sensitive to cues necessary for face recognition (Brunet, Mondloch, & Schmidt, 2009). On the other hand, it is also possible that those with poor face recognition skills acquire social anxiety due to problems in everyday life caused by the failure in recognising faces of colleagues, family and friends.

While the factors identified in this section have been found to have small effects on face recognition ability, it remains unclear why some people excel at this task. It is possible that they possess all or many of these factors, or that their skills are simply underpinned by enhancements in visuo-cognitive processes alone.

2. THE IDENTIFICATION OF SUPER RECOGNITION

Temporarily placing aside the issue of the underpinnings of super recognition, another fundamental practical issue is concerned with the identification or "diagnosis" of superior face recognition skills – a topic that has received very little attention to date. The two existing SR papers have primarily identified their SR participants using a cut-off of two standard deviations above the control mean on the long form of the Cambridge Face Memory Test (CFMT+; Russell, Chatterjee, & Nakayama, 2012; Russell et al., 2009). The standard form of this test (the CFMT: Duchaine & Nakayama, 2006) is extensively used to examine individual differences in unfamiliar face recognition skills in both typical perceivers (Bowles et al., 2009; Richler, Cheung, Gauthier, 2011; Wilmer et al., 2010) and those suspected to have DP (Bate et al., 2014; Bate et al., 2008; Russell et al., 2012; Russell et al., 2009), whereas the extended version of the test overcomes ceiling effects associated with the earlier version. It is generally well accepted that the standard form of the CFMT has excellent psychometric properties, with particularly high reliability (Bowles et al. 2009; Duchaine & Nakayama, 2006).

Russell and colleagues (2009) also used a second test to identify SR participants: a "before they were famous" test which presents photographs of celebrities that were taken some time before they became famous. Unsurprisingly, the SRs also performed well on this test, but it is very difficult to use a famous face test as a reliable diagnostic indicator. Indeed, the level of exposure to target faces and to similar tests (these often appear on social media) cannot be controlled between participants. Although Russell et al. (2009) reported positive correlations between this test and performance on the CFMT and CFMT+, a sampling error makes interpretation of these findings difficult, if not impossible. Namely, four of the 29 participants were SRs in Russell et al.'s study - while there are no published reports on the prevalence of super recognition in the general population, it is highly unlikely that such a high proportion of individuals would possess extraordinary face recognition skills. Hence, the top end of the score distribution in Russell et al.'s study is artificially inflated, and the conclusion that the famous face test correlates with the CFMT and CFMT+ should be seen as tentative. In any case, given most people are excellent at familiar face recognition and individual differences in the face recognition skills of typical perceivers are much better documented in tests of unfamiliar face recognition (Bowles et al., 2009; Richler et al., 2011; Wilmer et al., 2010), there is not currently a strong rationale for using famous face tests to identify SRs.

Russell and colleagues (2009) also examined the perception of facial identity (i.e. by presenting images simultaneously for comparison, placing no demands on face memory) in their four SR participants, using the Cambridge Face Perception Test (CFPT; Duchaine et al., 2007). While Russell et al. make the case that their SRs also outperformed control participants on this test, it should be noted that only a group-based comparison was offered as opposed to the single-case analyses that are typically presented in cognitive neuropsychological investigations (e.g. Bate et al., 2008, 2014). However, it is near impossible for individuals to significantly outperform controls on this test using single-case comparisons given the large variation in control performance and the resulting large standard deviation. Nevertheless, it is of note that examination of the raw data (see Figure 5, Russell et al., 2009) indicates that only some SRs performed above the control mean on the CFPT. This data raises the possibility that the superior face recognition skills of SRs are not always associated with superior face perception skills.

3. THE COGNITIVE UNDERPINNINGS OF SUPER RECOGNITION

The possibility that facial identity perception may not always be facilitated in SRs is important from a theoretical perspective. Cognitive neuropsychological investigation of cases of both developmental (e.g. Bate et al., 2009; Eimer, Gosling, & Duchaine, 2012; Garrido, Duchaine, & Nakayama, 2008) and acquired (e.g. Bate et al., 2015; Rezlescu, Pitcher, & Duchaine, 2012) prosopagnosia have aided the development of cognitive theories of typical face-processing, and tested their assumptions. For instance, the dominant model posited by Bruce and Young (1986) suggests that face-processing is a hierarchical sequential process (see Figure 1), whereby an initial stage of visual analysis is proceeded by the structural encoding of an incoming facial representation. At this stage, the view-dependent representation of the image is transformed into a view-independent representation, in preparation for identity recognition. Once the view-independent representation is constructed, it is compared to all stored representations of known faces in the face recognition units (FRUs). If a familiarity match is achieved, the relevant person identity node (PIN) is activated, and biographical information about that person is retrieved. Finally, the name of the person is accessed. Meanwhile, other perceptual aspects of the view-dependent representation (e.g. emotional expression) are thought to be processed independently to identity recognition.

Investigations using individuals with prosopagnosia suggest that face-processing can be interrupted at different stages of Bruce and Young's model, and that these patterns of impairment may relate to different subtypes of the condition. These findings broadly relate to deficits in face perception that are thought to occur at the level of structural encoding (e.g. Eimer, 2000; Eimer & McCarthy, 1999), and higher-order deficits affecting only face recognition that have been linked to impairments at the level of the FRUs or the PINs (Rezlescu et al., 2012). This latter group of individuals have also contributed to a key double dissociation that has bolstered the theory that facial identity and facial expression are processed independently. That is, neuropsychological patients with prosopagnosia alongside intact facial expression recognition skills appear to present with the reverse pattern of impairment to individuals who cannot recognise facial expression but can recognise facial identity.

Cognitive theories of face-processing can therefore be adopted to predict potential patterns of performance in SRs. Specifically, super recognition (a) may be underpinned by enhanced processing at dissociable stages of the face-processing model, and (b) if the enhancement is at the level of structural encoding, the processing of other aspects of face perception (e.g. facial expression) may also be heightened. Conversely, if the enhancement is at the latter stages of processing (i.e. at the level of the FRUs or PINs) only the recognition of facial identity will be facilitated. While the consistency of the cognitive presentation of SRs provides a novel means to test the predictions of theoretical models of face-processing, no in-depth study has addressed these issues to date.

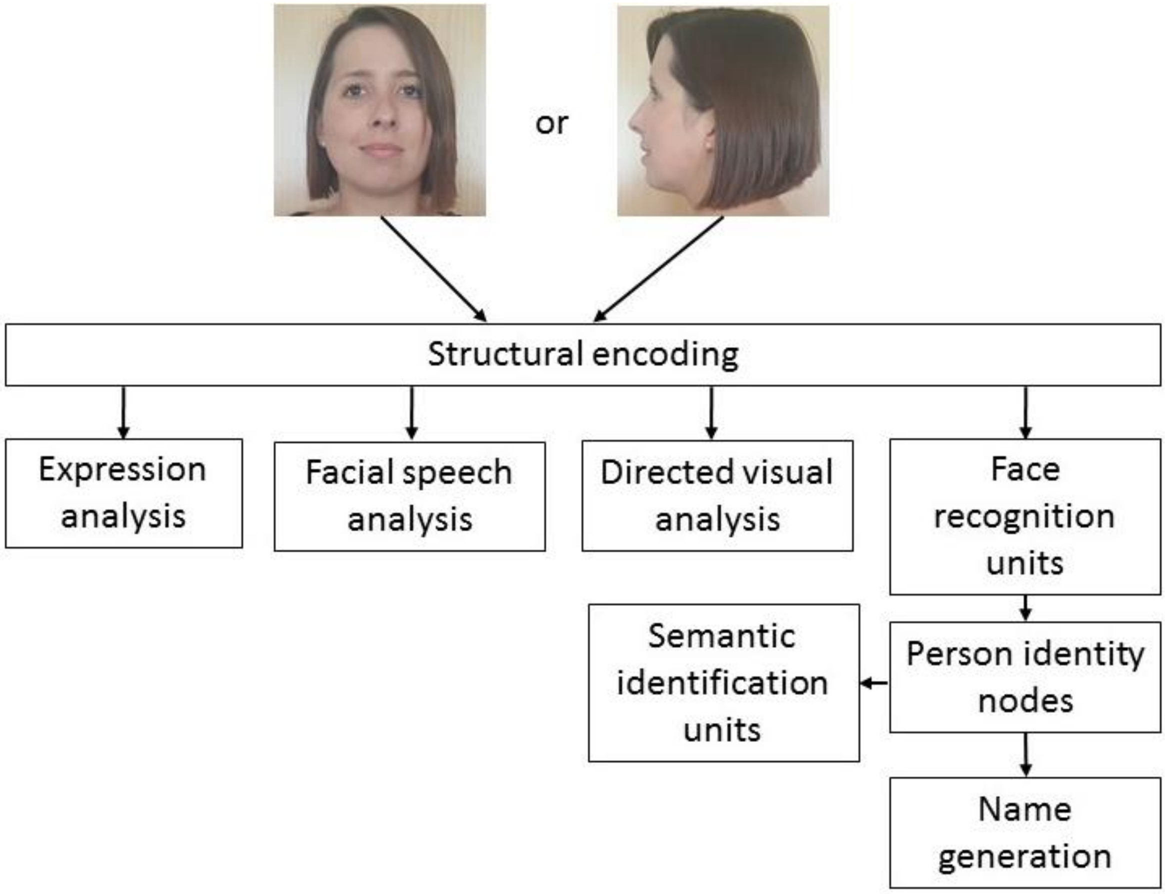

Figure 1. Adaptation of the Bruce & Young's (1986) sequential model of face recognition.

4. PROCESSING STRATEGIES IN SUPER RECOGNITION

While superior face recognition skills may be underpinned by one or more facilitation within the theoretical framework offered by Bruce and Young (1986), another possibility is that SRs use more efficient processing strategies to extract facial information. This may be reflected in their use of the "configural" or "holistic" processing strategy that is thought to be optimal for successful face recognition (e.g. Richler, Cheung, & Gauthier, 2011), or it may be that they are drawn towards specific regions of the face that hold information that is critical for identification.

Configural and holistic processing

Numerous reports indicate that faces are processed in a different manner to objects (McKone & Robbins, 2011; Rossion, 2013). For example, faces are thought to be processed holistically – that is, information is thought to be integrated from across the face rather than being broken down into individual parts (Piepers & Robbins, 2012; Rossion, 2013; see Maurer, Le Grand, & Mondloch, 2002; Richler, Palmeri, & Gauthier, 2012, for a review of different meanings of "holistic processing"). Furthermore, evidence suggests that we are highly sensitive to relational or configural information in faces (Maurer et al., 2002; Piepers & Robbins, 2012): this means that we are more sensitive to subtle variations in the spacing of facial features compared to those of other objects (e.g. Robbins, Shergill, Maurer, & Lewis, 2011; Yovel & Kanwisher, 2008). There is a long-standing belief that the use of these holistic and/or configural processing styles may underlie our proficiency in face recognition (e.g., Richler, Cheung, & Gauthier, 2011a; Rossion, 2013), and many studies attempting to explain group or individual differences in face processing have examined indicators of configural or holistic processing (e.g., DPs and controls: DeGutis, Cohan, Mercado, Wilmer, & Nakayama, 2012; Palermo et al., 2011; children and adults: Crookes & McKone, 2009; Mondloch, Le Grand, & Maurer, 2002; individual differences: DeGutis, Wilmer, Mercado, & Cohan, 2013; Richler et al., 2011a). Given the apparent importance of holistic and configural information in face recognition, it is possible that super recognition is underpinned by proficiencies in these face-specific perceptual processes.

Some preliminary evidence supports this hypothesis. The SRs reported by Russell et al. (2009) showed a larger face inversion effect (a difference in performance between upright and inverted faces) than control participants. The inversion effect is thought to reflect the fact that face-specific perceptual processes – namely, holistic and configural processing – are specialised for upright faces, and are disturbed or reduced in inverted faces (Maurer et al., 2002; Richler, Mack, Palmeri, & Gauthier, 2011b). Therefore, a larger inversion effect is thought to reflect stronger holistic or configural processing, and the fact that SRs showed superior performance for upright faces but relatively normal performance for inverted faces indicates that they may be particularly good at these tasks. Further, Sekiguchi (2011) examined eye-movements in a group of participants with typical face recognition abilities, and found that individuals with higher face memory (who were not SRs) tended to make more saccades between the two eyes than participants with poorer face memory (who did not have prosopagnosia). Sekiguchi suggested that this finding may also reflect the importance of configural processing (i.e. processing the specific spatial relationships between the eyes and other facial features) in face recognition.

The use of regional facial information

The finding reported by Sekiguchi (2011) is of interest because it suggests that information from the eye region may be critical for optimal facial identification performance. Many other studies have reported that the eye region is particularly pivotal for the recognition of facial identity, given typical perceivers fixate on the eyes to a greater extent than any other facial region (e.g. Schyns, Bonnar & Gosselin, 2002; Slessor, Riby & Finnerty, 2013). These findings are bolstered by reports that individuals with acquired prosopagnosia spend less time examining the inner features of the face (i.e. the eyes, nose and mouth) than controls, (e.g. Caldara, Schyns, Mayer, Smith, Gosselin, & Rossion, 2005; Lê, Raufaste, & Demonet, 2003; Lê, Raufaste, Roussel, Puel, & Demonet, 2003; Stephan & Caine, 2009), and one study has reported the same effect in DP (Schwarzer et al., 2007). More specifically, some studies suggest that participants with acquired prosopagnosia (Caldara et al., 2005; Bate et al., 2015; Stephan & Caine, 2009; Van Belle, Ramon, Lefevre, & Rossion, 2010) spend less time examining the eyes and more time examining the mouth than control participants.

However, other lines of evidence suggest the critical measure is not the proportion of dwell time spent on the eyes, but the time spent examining the nose. Hsiao and Cottrell (2008) reported that the optimal viewing position in face recognition (i.e. the location of the first fixation that a person makes to a face) is to the left of the centre of the nose. In addition, the preferred landing position (i.e. the location that participants fixate the most) is around the centre of the nose, rather than within the eye region. Similarly, Peterson and Eckstein (2012) found that the optimal viewing position on a range of face-processing tasks was below the eyes and towards the left side of the nose – in a remarkably similar position to that observed by Hsiao and Cottrell. Both sets of authors suggest that this viewing position may be the optimal location for holistic processing of the entire face to occur.

No study to date has examined eye-movement strategies in SR participants. Investigation of this issue does not only have the potential to inform us about the underpinnings of superior face recognition, but may also inform the key theoretical debate regarding the importance of the eyes versus the nose in facial identification.

5. ARE SUPER RECOGNISERS USEFUL IN APPLIED SECURITY SETTINGS?

The above review clearly indicates that investigation of super recognition is an innovative means to garner further insight into our theoretical understanding of the typical face recognition system. However, it is also of practical interest to examine whether SRs may play an important role in policing and security settings, given excellent face recognition skills are pivotal in some occupations. This is particularly pertinent given increasing findings that (a) typical perceivers (and even those with many years of experience in relevant roles) tend to make many errors in applied face-processing tasks, and (b) the limited success that has been observed in studies that have attempted to improve face recognition skills in the typical population.

The performance of typical perceivers on applied tasks of unfamiliar face-processing

Numerous studies have illustrated the difficulties that typical perceivers encounter in their recognition of unfamiliar faces, and many of these studies have adopted paradigms that mirror face recognition scenarios in security or policing settings. Photographic ID is the most ubiquitous means of evaluating one's identity. It is commonly used to buy age restricted items, travel, to access one's work place or to verify identity by the law enforcement agencies. Overall there is a considerable reliance on this method, despite consistent demonstrations that the matching of unfamiliar faces is a very difficult task. Specifically, in the Glasgow Face Matching Test (GFMT, Burton, White, McNeill, 2010), where participants have to compare two simultaneously presented images and decide whether they show the same person or not, average accuracy is 80% on the short and 89% on the long version of this task. This relatively high error rate occurs even though both images from "match" trials were taken on the same day from a frontal viewpoint, and no time restriction is imposed on the participant. Importantly, the range of performance on the long version of this test was reported to be between 62% and 100%, suggesting large individual differences within the general population.

In another study, Bindeman, Aveytisan and Rakow (2012) administered the GFMT to participants over three (Experiment 1) and five (Experiment 2) consecutive days. The authors reported overall accuracy on a par with the normative data provided by Burton and colleagues (2012), but also highlighted that while some individuals performed consistently over successive attempts, the accuracy of others was subject to considerable fluctuations. Interestingly, when the same set of face pairs was presented on each of the three days (Experiment 1), participants' judgements were not stable in that the same image pairs were often classified differently on each day.

Another occupational setting where unfamiliar face recognition ability is pivotal involves work with video footage or the recognition of missing or wanted persons by officers on patrol. Studies have consistently shown that while such tasks are easy to perform with personally familiar faces (Bruce et al., 2001; Burton et al., 1999), the accuracy of identification for previously studied unfamiliar faces is close to chance in typical observers (Burton et al., 1999). In a seminal study, Burton and colleagues (1999) asked participants to encode ten identities from video footage and later to identify them from photographs. The identities used in the study were of Psychology lecturing staff and the experimental groups included Psychology students, students from non-Psychology courses and police officers. The group who were personally familiar with the identities presented in the video footage (i.e. the Psychology students) were highly accurate in making familiarity judgements for the studied identities, while participants in the unfamiliar groups (students and police staff) were significantly less accurate in their judgements. Interestingly, there were no differences in performance between those who were untrained and inexperienced in face recognition assignments (i.e. students) and police officers who presumably should be acquainted with this type of task.

As discussed above, one possible reason for these low identification rates is that merely pictorial encoding occurs when only one view of the face is available. Specifically, Bruce and Young's (1986) modular sequential model of face recognition posits that faces are initially encoded using pictorial view dependent representations, but with gradual exposure to many views of the same face, this view-dependent information is transformed into more stable view-independent representations. This theoretical account is supported by a plethora of empirical evidence suggesting that variability is critical in the learning of new faces (Andrews, Jenkins, Cursiter, & Burton, 2015; Burton, Kramer, Ritchie, & Jenkins, 2016; Dowsett, Sandford, & Burton, 2016; Ritchie & Burton, In press).

These findings clearly suggest that familiarity, likely achieved by the building of stable face representations, increases the likelihood of a positive identification of a face. However, in real life, perpetrators of crimes or missing persons are often unfamiliar to those searching for them, and the number of images available for comparison is restricted by the availability of CCTV footage, the number of images in a database, or the number and quality of photographs provided by family and friends of a missing person.

The considerable individual differences in face matching performance have important implications for work assignments in national security settings. Pertinently, passport controllers, victim identification officers and CCTV operators perform facial image comparisons on an everyday basis. What is more, routine ID checks accompanying the purchase of age-restricted items are performed by individuals drawn from the typical population. Pertinently, while the images used in the aforementioned studies of individual differences can be described as "optimal" for face matching to be performed accurately, in real life one's appearance is subject to changes in age, weight, hair style, lighting, and so on. Indeed, in a study by Megreya, Sandford, and Burton (2013), participants' performance on a version of the GFMT (Experiment 2) was 90% when images were taken on the same day, but declined to 70% when images were taken several months apart.

In more realistic settings, Kemp, Towell, and Pike (1997) examined fraud detection from credit cards in a supermarket setting. Cashiers were assigned to a task of matching photographic credit cards (with images sized 2cm x 2cm) to their bearers at the checkout. When cards were used falsely and the identity of the shopper did not match the credit card, accuracy amongst experienced cashiers was as low as 50%. This was despite the fact that the supermarket employees were briefed before the study and promised a bonus of 50% above their statuary reimbursement if they performed the checks accurately and in a timely fashion. As such, participants in Kemp et al.'s study were presumably motivated to be vigilant and perform well. It is possible that in real-life situations and in the absence of these financial inducements accuracy would be even lower.

When IDs are used for international border crossing, successful matching of photographic documents to their holders is a matter of national security. In a recent study White, Kemp, Matheson, Jenkins, and Burton (2014) showed that when compared to unqualified students, experienced passport control officers make a comparable amount of errors when fraudulent mock IDs are produced. Officers were reported to make 10% of mistakes, even though (1) the photographs for mock IDs were taken only a few days prior to the testing session and (2) the foils (photographs depicting a person different to the ID bearer) were chosen opportunistically from students who volunteered to take part in the study, rather than being carefully chosen according to pre-rated similarity to the holder. It is conceivable that if the photographs were taken many months apart, the accuracy would decline similarly to performance in the paradigms used by Megreya and colleagues (2013).

Together, the above literature illustrates that, across many tasks, unfamiliar face recognition is a difficult and error-prone process. It is therefore conceivable that identification of SRs as potential candidates for some real-word occupations would be invaluable for national security. Whether the existing means of identifying SRs are useful for more applied face recognition scenarios is as of yet unknown, and of course if theoretical investigations do uncover different subtypes of super recognition, the same individuals may not necessarily be suited to every task.

Can face recognition skills be improved?

In order for SRs to be used for more applied roles they must be readily available. The precise definition and prevalence of “true” SRs has not yet been determined, raising the possibility that there may not be many of these individuals available for employment. An alternative is that typical perceivers may be trained to become SRs, or at least to undergo some improvement in their face recognition skills. However, the study described above by White et al. (2014) casts doubt on whether this is possible, given they found no difference in face recognition performance according to officers’ years of relevant experience. Three other studies have reported that trained experts excel at face matching tasks in comparison to typical perceivers (Norell et al., 2014; White, Dunn, Schmid,, & Kemp, 2015a; White, Phillips, Hahn, Hill, O’Toole, 2015b). However, baseline face recognition ability was not examined in those participants and it is unclear whether they have a natural ability to process unfamiliar faces or if their superior performance is a result of training and experience. It is thus possible that these “experts” are aware of their extraordinary face recognition ability and self-select for assignments involving facial image comparison. It is also important to note that neither of the recent studies with experts reported case-by-case analyses or the variation in performance within the experimental groups. As such, these studies shed little light on the processes underpinning superior face matching in the expert groups.

While the error-prone performance in various face matching paradigms has been well-documented in the literature, only recently have efforts been made to enhance it, mostly with limited success. Two investigations concentrating on facial features (Woodhead, Baddeley, & Simmonds, 1979) and face shape (Towler, White, Kemp, 2014) were not able to improve face matching performance at all. Attempts to improve face matching with trial-to-trial feedback have yielded mixed results. While White, Kemp, Jenkins and Burton (2014) reported improvement in face matching performance generalising to novel faces, Alenezi and Bindemann (2013) found that trial-to-trial performance feedback merely inhibits performance decline, but does not lead to an overall increase in matching accuracy. Moore and Johnson (2013) investigated the impact of food incentive, on performance in a face matching task. They found that the overall discriminability increased when participants were made aware of a sweet food incentive, and participants also became more conservative in their responding, i.e. the improvement was driven by the increased accuracy on mismatched trials. Other attempts at improving face matching performance included caricaturing (McIntyre, Kittler, Hancock, & Langton, 2012) and redesigning the ID to include multiple photographs of the holder (White, Burton, Jenkins, & Kemp, 2014), all yielding rather limited results.

A recent study by Dowsett and Burton (2015) examined the impact of working in pairs on face matching performance. In a series of experiments, the authors showed consistent increase in face matching accuracy and further individual improvement in performance, particularly in those whose scores were initially low. Most importantly, the effect of working in pairs was transferable to individual performance on a new set of images, a finding providing a potential path for future regimes. The longevity of this effect is, however, unclear as all participants were tested on the same day. Furthermore, in the current economic climate and governmental agencies affected by austerity measures, increase in staffing is an unlikely step. As such, ways of improving individual performance are a pivotal research avenue.

One such attempt was recently reported by Bate et al. (2014) where participants inhaled a nasal spray with oxytocin or placebo before completing the one-in-ten task (Bruce et al., 1999), a well-established line-up matching paradigm containing target-present and target-absent trials. While, participants in the oxytocin condition had better accuracy on target-present trials, they were also more prone to make positive identifications in target-absent trials, a potentially costly mistake, when made within the national security or forensic sectors. Bate et al. (2014) suggested that it is possible that the aspects that contribute to oxytocin's facilitative effect when administered before encoding in face memory tasks, such as increasing the saliency of studied faces, are disadvantageous in line-up scenarios. In target absent line-ups, when many faces are presented simultaneously, participants could be mistaking saliency for familiarity and choosing a foil that seems as the most similar to the target. This is in contrast to the placebo condition, where without the saliency enhancing effect of oxytocin, participants responded in a more conservative way. This latter methodology therefore may still have potential for improved face recognition when the hormone is inhaled before recall, yet this possibility had not yet been explored.

6. THESIS OVERVIEW AND AIMS

In sum, four key questions are addressed in this thesis. First, what are the underpinnings of super-recognition? Second, how prevalent is super-recognition, as measured by the currently available tests of face recognition (the CFMT+, Russell et al., 2009) and face perception (CFPT, Duchaine et al., 2007)? Third, how well do super-recognisers perform on tests of face recognition and face matching resembling real-life scenarios? Finally, is it possible to instantaneously improve face recognition and face matching ability in typical observers?

To answer these questions, studies in Chapter 2 offer an in-depth examination of the cognitive and perceptual abilities of SRs. In addition, eye-movement patterns are examined in Chapter 3 and compared to those of typical perceivers and individuals with developmental prosopagnosia. Chapter 4 investigates the distribution of face recognition ability and scrutinises the currently used cut-offs for the “diagnosis” of super recognition. Chapters 5 and 6 examine the performance of SRs on applied tasks of face matching and memory. Specifically, Chapter 5 looks at the recognition of faces from CCTV and matching performance in the well-established one-in-ten task (Bruce et al., 1999). Chapter 6 further examines the ability of super-recognisers to match faces in tasks resembling passport control and other types of facial image comparison. Chapter 7 investigates the ostensible role of oxytocin in face recognition ability by applying it to the real-life scenarios investigated in the previous two chapters. The final chapter pulls together all these findings and discusses their implications for the diagnosis of superior face recognition, our theoretical understanding of the face-processing system, and practical scenarios in forensic and security settings.

Chapter 2: An in-depth cognitive examination of individuals with superior face recognition skills

Chapter (except for Study 2) published as:

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (In press). An In-depth Cognitive Examination of Individuals with Superior Face Recognition Skills. *Cortex*.

1. INTRODUCTION

Russell, Duchaine, and Nakayama (2009) presented the first report of super recognition, describing four individuals who outperformed control participants on tests of face memory, face perception, and familiar face recognition. Specifically, Russell and colleagues reported the face memory and face perception scores of these super-recognisers (SRs) and found that they significantly outperformed typical observers on both types of tasks (in group-level analyses). Given only one other paper to date has investigated super recognition (and that paper focused on the role of surface and reflectance processing in super-recognition; Russell, Chattejee, & Nakayama, 2012), there has been very little work examining the skills that underpin super recognition. In particular, it is unclear whether it is underpinned by enhancements to more generalized mechanisms, specific stages of the face recognition process, or specific processing strategies. Investigation into these issues is of clear theoretical importance, particularly as SRs (akin to those with prosopagnosia) may not represent a homogenous group of individuals in terms of their cognitive presentation or the mechanisms underpinning their superior skills.

This chapter extends the existing SR literature by reporting a detailed neuropsychological assessment of six individuals who meet the previously published criteria for super recognition. Four questions are addressed via a battery of neuropsychological and cognitive tests. First, more general perceptual and cognitive processing mechanisms are examined, to investigate whether enhancements in these processes support superior face recognition skills. Much research supports the hypothesis that face recognition is a highly specialised process involving a number of dedicated neural circuits (Haxby, Hoffman, & Gobbini, 2000; Gobbini & Haxby, 2007), and this theoretical standpoint is supported by findings that some individuals with developmental (Duchaine & Nakayama, 2005; Jones & Tranel, 2001) and acquired (Busigny, Joubert, Felician, Ceccaldi, & Rossion, 2010; de Renzi & di Pellegrino, 1998) prosopagnosia only have difficulties in the recognition of faces. Further, existing work has failed to find a relationship between face recognition skills in the typical population and performance on tests of non-facial visual memory (e.g. an abstract art memory test) or verbal memory (e.g. verbal paired-associates test) (Wilmer et al., 2012; Wilmer et al., 2010). However, no work to date has examined the domain-specificity of super recognition, and it is possible that particularly good general perceptual or mnemonic abilities could support the exceptional face recognition skills observed in these individuals. Alternatively, if it is found that the exceptional skills of SRs are restricted only to the processing of faces, this would further support the face-specificity hypothesis.

Second, we investigate whether SRs are only proficient at facial identity recognition, or whether their skills extend to other aspects of face-processing (e.g. the recognition of emotional expression). This speaks to important theoretical questions concerning the structure and function of the face-processing system, given dominant cognitive theories posit that the face-processing pathway is composed of a set of hierarchical sub-processes (e.g. Bruce & Young, 1986). Individuals at the other end of the face recognition spectrum, i.e. those with prosopagnosia (Bate & Bennetts, 2014; Bate & Cook, 2012; Bennetts, Butcher, Lander, & Udale, 2015), can present with or without impairments in face perception (Dalrymple, Garrido, & Duchaine, 2014). These findings have aided the development of dominant models of face-processing by indicating that the face-processing pathway can be lesioned at different locations (i.e. at an early stage involving structural encoding or a later stage involving retrieval); yet the presumed hierarchical nature of the framework explains why the hallmark deficit in facial identity recognition presents even in the former group of individuals. Likewise, it follows that super recognition may result from relatively early enhancements affecting all aspects of face perception (i.e. judgments that use view-dependent representations, such as age, gender or expression), facial identity perception alone (i.e. after view-independent representations have been created), or from later enhancements affecting only memory for faces. Examination of the face perception abilities of SRs will therefore have important theoretical implications by presenting a novel means to evaluate current theoretical frameworks.

Third, the processing strategies used by SRs are examined, to investigate whether these are different or merely enhanced in comparison to typical perceivers. Indeed, while SRs are clearly better than controls at face recognition in quantitative terms, it is unknown whether this heightened performance is underpinned by qualitative differences in processing strategy, whereby they use different types of visual information or different information processing styles. Alternatively, SRs may adopt the same processing strategies as typical perceivers, but in a heightened or more efficient manner. It is well-accepted that configural or holistic processing strategies underpin typical face recognition and many researchers agree that the composite task is the most robust measure of this process in group studies (Richler, Floyd, & Gauthier, 2014). Although this task has not yet been administered to SRs, existing work indicates that it does correlate with face recognition abilities in the general population (Richler et al., 2011; Wang, Li, Fang, Tian, Liu, 2012, c.f. Konar, Bennett, & Sekuler, 2010), and that the composite face effect is reduced in people with prosopagnosia (Avidan, Tanzer, & Behrman, 2011; Palermo et al., 2011; but see Susilo et al., 2010). To date though, no studies have addressed this question directly in SRs, and it remains unclear whether these face-specific processes underpin superior face recognition skills. It should be noted, though, that performance on composite tasks widely varies even in the typical population, and it can be very difficult to reliably detect individual differences in performance using single-case comparisons (Rossion, 2013).

Hence, while any significant individual differences using case-by-case analyses may be insightful, null effects are more inconclusive.

It is important to note that holistic processing also occurs on a more general scale (e.g., integrating many different objects into a coherent visual scene). Manipulating this general process by asking individuals to focus on local details (e.g. the small letters in a Navon stimulus) can be detrimental to face recognition, possibly because it encourages piecemeal, non-holistic processing (e.g. Gao, Flevaris, Robertson, & Bentin, 2011; Macrae & Lewis, 2002). Building on this work, some research into prosopagnosia has established that some people with face recognition deficits show a general bias towards the processing of local details, and this correlates with their reduced holistic processing of faces (Avidan et al., 2011; but see Duchaine, Yovel, & Nakayama, 2007). This work suggests that it may be variation in this more general holistic processing ability, rather than a face-specific process per se, that underpins individual differences face recognition abilities. Once again, though, this issue has not been addressed in the SR population. Examination of face-specific configural and holistic processing, alongside more general holistic processing tendencies (also sometimes referred to as "global processing biases", e.g., Behrmann et al., 2005; Duchaine et al., 2007), would therefore provide insight into whether SRs show specific qualitative differences in processing strategies, and whether these effects are domain-specific or reflect more general perceptual processes.

Finally, we pull our findings together to examine whether SRs show a consistent pattern of enhanced abilities, or whether these individuals vary in their cognitive presentation as has been observed at the bottom end of the face-processing spectrum (i.e. in DP).

2. CASE DESCRIPTIONS

Following widespread media coverage about super recognition, the six individuals described in this chapter contacted our laboratory. DF is an 18 year-old right-handed male Engineering student, TP is a 35 year-old right-handed male IT manager , GK is a 33 year-old right-handed male university lecturer, JN is a 35 year-old right-handed female sourcing consultant, CH is a 27 year-old right-handed male lawyer, and CW is a 21 year-old Psychology graduate.

In an initial informal interview, all the SRs described extraordinary face recognition skills that had been present from an early age. They reported that they are able to recognise people even after a brief encounter or after many years have passed (for instance, childhood friends): “I recently saw a girl who I taught for a couple of swimming lessons when I was a teenager. I recognised her immediately, despite the fact that I had not seen her since she was 6, and she is now 18” (CH). Following existing procedure, each participant was screened using the CFMT+ (Russell et al., 2009). The initial three stages of this test replicate the standard version of the CFMT (Duchaine & Nakayama, 2006). In the encoding stage, participants view each of six novel faces from three different viewpoints, and complete three test trials per face where they are asked to select the encoded identity from a triad of faces. In the second stage, participants are asked to select the encoded identities from novel viewpoints or lighting conditions over 30 triads. The third stage is similar to the second, but the test faces are overlaid with visual noise to make recognition more difficult (24 trials). The CFMT+ then adds a fourth, more challenging stage to the original test: participants are asked to identify the learnt faces from profile images which now display hair, are heavily cropped, or show a different emotional expression (30 trials, see Figure 1).

All six SRs achieved CFMT+ scores that are above the previously-used cut-off of 90/102 (Russell et al., 2009, 2012) on this test (see Table 1). However, we also collected our own control data (N = 30, 15 female; mean age = 25.9 years, SD = 4.5) to ensure that we were comparing our SRs to an appropriately matched control group. Single case statistics showed that all the SRs but one (TP) significantly outperformed the control group: CW and GK, $t(32) = 2.66$, $p = .01$, $Z_{cc} = 2.70$, 95% CI [1.917, 3.474];

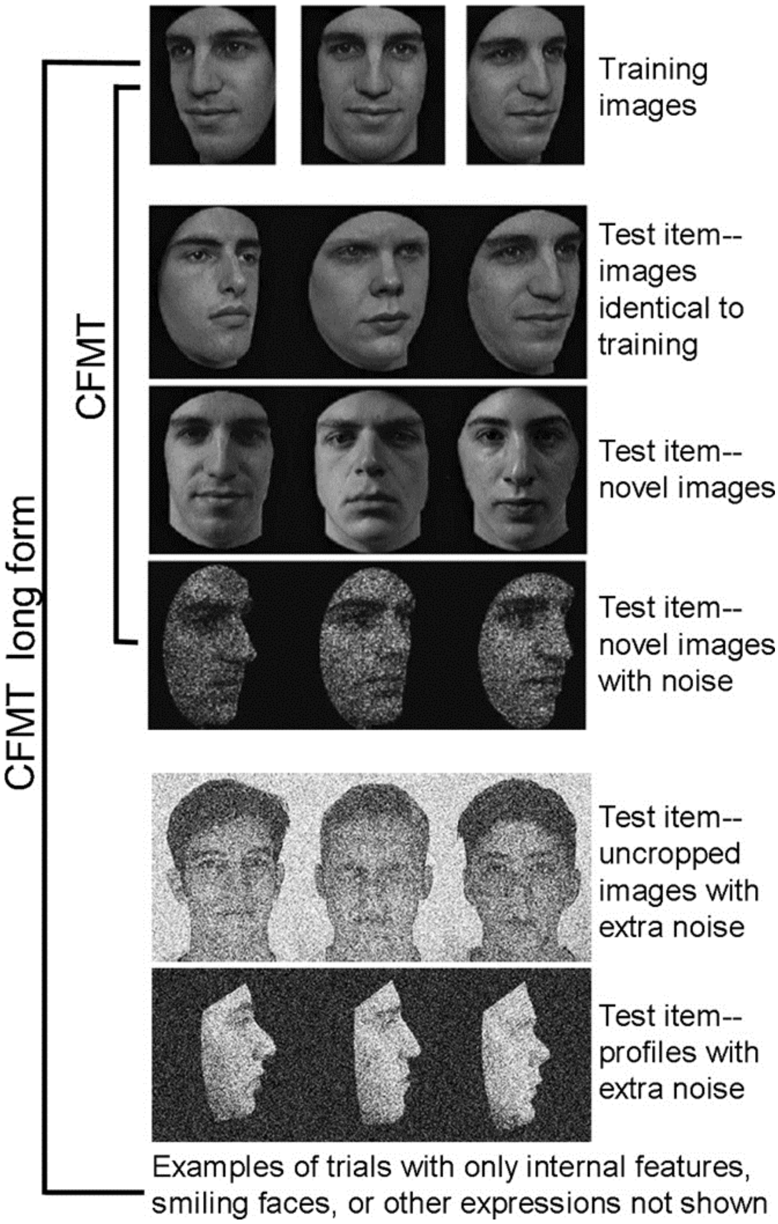

Figure 1. The structure of the CFMT+ (Russell et al., 2009)

estimated % population below their scores = 99.37 and JN, CH and DF, $t(32) = 2.40$, $p = .02$, $Z_{cc} = 2.445$ (95% CI: 1.718 – 3.160); estimated % population below their scores = 98.86. Given TP reached the criteria for super recognition based on previously published control data (Russell et al., 2009) and on additional tests of face recognition (see Chapter 5), we still included him in our sample for this investigation.

All SRs reported normal or corrected-to-normal vision. General intelligence was assessed using the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II, Wechsler, 2011). One SR performed within the “average” range (JN), whereas TP, DF, CH and GK were within the “superior” range (see Table 1). Due to limited time availability, CW’s intelligence was estimated using the WTAR (Holdnack, 2001). Similarly to JN, he scored within the “average” range. While CH excelled at the verbal component of the measurement, DF and JN showed a clear advantage on the performance rather than verbal sub-tests. Conversely, both TP and GK performed similarly on the two sub-tests. This variation in IQ is in line with findings that face recognition ability is domain-specific and unrelated to general intelligence (Wilmer et al., 2009; Zhu et al., 2010).

For each of the investigations below, performance of the SRs was compared to controls using at least two tests to address each theoretical question. For each individual test, a subset of individuals were extracted from a control group containing 30 gender- and age-matched participants (19 female, M age = 32.1, SD = 9.3; see Table 1).

Table 1. Demographical and background neuropsychological information about the SR participants, presented in comparison to controls. Values for the performance of the SR participants on the CFMT+ are expressed in the number of SDs away from the control mean.

	Controls			Super-Recognisers					
	Mean	SD	N	CH	DF	JN	GK	CW	TP
Age	32.1	9.3	30	27	18	35	33	21	35
Gender	19 (F)	-	30	M	M	F	M	M	M
Handedness	3L	-	30	R	R	R	R	R	R
WASI-II[a]:									
Verbal	-	-	-	148	114	99	118	-	127
Performance	-	-	-	111	131	116	119	-	127
Full-2 IQ	-	-	-	134	125	108	121	-	130
WTAR[b]	113.8	8.2	30	-	-	-	-	115	-
CFMT+[c]	68.4/102	11.7	30	2.4*	2.4*	2.4*	2.7*	2.7*	2.0

* indicates participant significantly differed to controls using Crawford et al.’s (2010) modified *t*-tests for single-case comparisons ($p < .05$)

[a]Wechsler Abbreviated Scale of Intelligence, Second Edition (Wechsler, 2011) – this more thorough assessment of IQ was carried out with available SRs; [b]Wechsler Test of Adult Reading (Wechsler, 2001) – this quick IQ screen was used with controls to ensure they were appropriately matched to the SRs and with CW due to time constraints; [c]Cambridge Face Memory Test - Long Form (Russell et al., 2009) – this test was used to confirm superior face recognition skills in the SRs and typical skills in the controls

These individuals had typical face recognition skills (as confirmed by their performance on the CFMT+: see Table 1). Note that a larger control sample is reported for the CFPT, due to the larger variability in the typical population on this test (see below). All control participants presented with normal visual acuity and contrast sensitivity. Not all control participants completed all tests due to time constraints and some computer errors (the N for individual tests is presented in Tables 1-4; gender was approximately equal for each test). For each test, the SRs were compared to the controls on a single case level, using modified *t*-tests for single case comparisons (SINGLIMS, Crawford, Garthwaite, & Porter, 2010) or Revised Standardised Differences Tests (RSDT, Crawford et al., 2010) as appropriate. This is a particular strength of this work as previous studies (Russell et al. 2009; Russell et al., 2012) have only used group-based statistics to analyse the performance of a smaller number of SRs. Informed consent was obtained from all participants, and ethical approval for the study was granted by the departmental ethics committee.

3. STUDY 1: IS SUPER RECOGNITION FACE-SPECIFIC?

As discussed above, previous work examining super recognition has focused exclusively on their face recognition performance, and it remains possible that the skill is supported by enhancements in more generalized cognitive, perceptual or mnemonic skills. An initial investigation sought to address this issue by examining performance on two different object-processing tests: one assessing matching skills, and the other memory skills.

3.1. Matching test

An object and face matching test was created to assess whether SRs show superior object processing skills compared to typical participants. Participants completed a sequential same/different matching task with faces, hands, and houses (see Figure 2). Each trial consisted of two sequentially presented objects – the initial study image was displayed for 250 ms, and the second test image was displayed until the participant responded. In the face condition, the study image showed a face from a frontal viewpoint and the test image showed a face from a 30-45° angle. Faces were drawn from the Cambridge Face Memory Test–Australian (McKone et al., 2011) and the Bosphorous Face Database (Savran et al., 2012), and were edited to remove external features. Houses were created using the software Realtime Landscaping Plus (Idea Spectrum Inc., 2012). Each house contained the same number of features (three sets of windows and a door), placed onto a constant background texture. The shape and location of the features, the luminance of the background texture, and the overall shape of the house varied throughout the set. As in the face condition, the study and test images presented the houses from two different viewpoints (frontal and 15° profile). Hand images were extracted from the Bosphorus Hand Database (Dutağacı, Yörük & Sankur, 2008), and showed the palm and fingers of a hand. Images were chosen to exclude rings, watches, cuffs, or other identifying features. Study and test images showed the hands in two different positions (e.g., fingers splayed and fingers together), with the wrist pointing downwards (upright condition) or upwards (inverted condition). Each category contained 32 pairs of images (16 same identities, 16 different identities). All pairs were presented twice upright and twice inverted. Trials were blocked by stimulus type, with upright and inverted trials presented randomly within each stimulus type. The order of blocks was randomised between participants. The measure *d'* (a bias-free measure of sensitivity; MacMillan & Creelman, 2005) was calculated for category of stimulus, and used in all analyses.

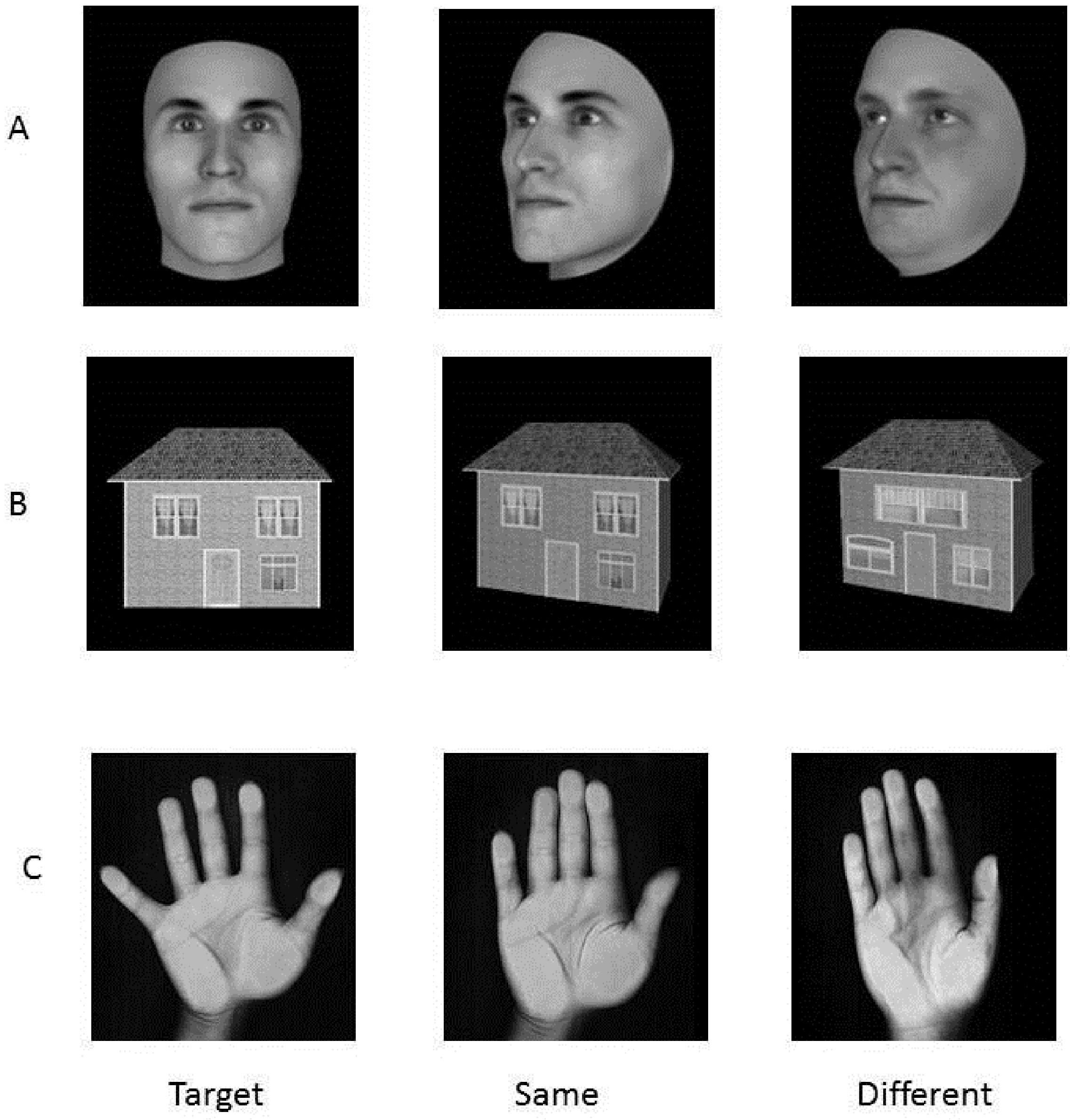

Figure 2. Sample stimuli from the object matching task: (A) faces, (B) houses and (C) hands. In the hands stimuli, finger splay rather than orientation differed between exemplars. Face images shown in this figure are computer-generated and for illustration only. The stimuli that were actually used in the test were of real faces, but publication rights cannot be obtained.

A repeated measures ANOVA on control participants' data revealed main effects of object, $F(2,19) = 26.99$, $p < .0005$, $\eta_p^2 = .74$, and orientation, $F(1,20) = 31.57$, $p < .0005$, $\eta_p^2 = .61$, and a significant interaction between object and orientation, $F(2,19) = 23.40$, $p < .0005$, $\eta_p^2 = .71$. Pairwise comparisons (Bonferroni corrected) confirmed that control participants showed a significant inversion effect for faces ($p < .0005$), but not for hands ($p = .325$) or houses ($p = .072$) (see Table 2).

On an individual level, two SRs (JN and DF) were significantly better at matching upright faces than control participants, JN: $t(20) = 3.22$, $p = .004$, $Z_{CC} = 3.30$ (95% CI: 2.19 - 4.39), estimated % of population below JN's score = 99.78%; DF: $t(20) = 2.80$, $p = .011$, $Z_{CC} = 2.86$ (95% CI: 1.88 – 3.84), estimated % of population below DF's score = 99.44% (see Table 2). TP, CH, GK, and CW performed better than control participants, but these differences did not reach significance (*p*s > .1)[1]. Single case analyses showed no significant differences between controls and SRs when matching upright hands (*p*s > .15) or houses (*p*s > .25), nor any inverted objects (all *p*s > .07), except for GK who was significantly better than the control group at the matching of inverted hands, $t(20) = 2.22$, $p = .042$, $Z_{CC} = 2.19$, 95% CI [1.405, 3.015), estimated % of population below GK's score = 97.88 (see Table 1). However, the same participant was significantly worse than controls at matching inverted houses, $t(20) = -4.26$, $p < .001$, $Z_{CC} = -4.36$, 95% CI [-5.767, -2.951], estimated % of population below GK's score = 0.02. It is of note, though, that negative d' values can suggest that the participant did not correctly follow the instructions, and it is possible that GK misunderstood the response labelling in this part of the task.

RSDT comparing the inversion effect of individual SRs for faces revealed that JN and DF showed a significantly greater effect of inversion than controls for faces, JN: $p = .004$ Z_{DCC}

[1] Discussion of performance on the upright faces condition is expanded below in the consideration of face perception skills (see section 4.2).

Table 2. Results from the object-processing tasks administered in Study 1. All values for SR participants are expressed in the number of SDs away from the control mean.

	Controls			Super-Recognisers					
	Mean	SD	N	CH	DF	JN	GK	CW	TP
Matching test (d'):									
Faces upright	2.00	0.40	21	1.60	2.90*	3.3*	0.10	1.10	1.80
Faces inverted	1.00	0.60	21	0.40	-0.50	-0.70	-0.50	-0.80	-0.60
Face inversion effect	1.04	0.61	21	1.56	2.51*	3.03*	0.56	1.62	1.82
Hands upright	2.00	0.70	21	1.60	-0.40	0.50	1.10	0.10	-0.10
Hands inverted	1.90	0.60	21	0.40	0.70	0.90	2.20*	-1.10	-0.30
Hand inversion effect	0.10	0.46	21	2.67*	-1.35	-0.39	-1.04	1.37	0.17
Houses upright	2.80	0.60	21	-1.20	0.00	0.20	0.70	-1.80	1.00
Houses inverted	2.60	0.70	21	-1.90	0.2	0.30	-4.40*	-0.30	1.00
House inversion effect	0.20	0.51	21	-1.37	-0.31	-0.24	7.14*	-1.71	-0.29
CCMT[a]:									
Females	50.4/72	7.20	93	-	-	0.60	-	-	-
Males	57.4/72	8.30	60	0.20	0.90	-	-0.70	0.40	1.60

* indicates participant significantly differed to controls using Crawford et al.'s (2010) modified *t*-tests for single-case comparisons ($p < .05$)

[a]Cambridge Car Memory Test (test and norms from Dennett et al., 2012) – performance varies according to gender on this test.

= 3.30 (95% CI: 2.18 – 4.57), estimated % of population showing a larger difference than JN = 0.21%; DF: p = .013, Z_{DCC} = 2.75 (95% CI: 1.78 – 3.85), estimated % of population showing a larger difference than DF = 0.68%[2]. TP, GK and CW did not show a disproportionate inversion effect when compared to controls (ps > .07). GK, however showed a significantly greater level of inversion than controls for houses, p < .001, Z_{DCC} = 6.59 (95% CI: 4.54 – 8.96), estimated % of population showing a larger difference than GK = 0.0004%. Moreover, CH showed a larger inversion effect for hands, p = .02, Z_{DCC} = 2.54 (95% CI: 1.67 – 3.52), estimated % of population showing a larger difference than CH = 1.4.

In sum, this investigation presents little evidence that SRs excel at the perception and recognition of objects in a matching task that places no demands on long-term memory. Only GK displayed enhanced processing in one object condition (inverted hands), yet also showed diminished processing of inverted houses. While it is likely that the latter finding represents a misunderstanding of task instructions, further investigation is required with this individual to present convincing evidence of enhanced object-processing capabilities.

3.2. Object memory

Memory for objects was assessed using the Cambridge Car Memory Test (CCMT; Dennett et al., 2012). The CCMT is an object equivalent of the CFMT – like its face counterpart, participants are required to learn six cars, then choose which of three presented cars is one of the learnt set. The CCMT consists of 72 trials across three blocks, which become progressively more difficult. Although single-case analyses indicated that all SRs scored within the normal range (all ps > .05; see Table 2), it should be noted that, for male

[2] Further discussion of inversion effects on this task can be found in section 5.2.

participants, even a perfect score on this test would not be considered significantly greater than controls (p = .088 for 100% accuracy). However, examination of the raw scores on this test indicates that only one individual (TP) approached ceiling on this task, scoring 71/72 (all other participants achieved scores that were within 1 SD of the control mean). Further, TP reported that he does not have a particular interest in cars, raising the possibility that his superior memory skills may generalize beyond faces.

3.3. Summary of Study 1

Four of the six SRs failed to show any evidence of superior processing of objects, on either a matching or a memory task. These findings suggest that, at least in some cases, super recognition is domain-specific. While CW outperformed controls at the matching of inverted hands, it is of note that his performance was not heightened in any other condition, nor on the memory task. Further investigation is required with this individual to convincingly conclude that his object processing skills are also superior to those of typical perceivers. The case of TP is of interest, given his near-ceiling performance on the object memory task. Given he did not outperform controls on the matching task, it is possible that his superior face recognition skills are underpinned by more general enhancements in memory. The next investigation speaks to this issue, examining whether the SRs are proficient at different stages of the face-processing framework.

4. STUDY 2: LOCATING SUPER RECOGNITION WITHIN THEORETICAL MODELS OF FACE-PROCESSING

A second investigation examined processing at theoretically relevant stages of the dominant cognitive model of face-processing proposed by Bruce and Young (1986; see Figure 1). First, we examined facial identity perception skills. Given this process can be differentially affected in prosopagnosia (Darymple, Garrido, & Duchaine, 2014), we sought to further explore this issue. Second, we examined whether non-identity facial perception is also enhanced in super recognition, and, following the theoretical precedent of the prosopagnosia literature (Duchaine, Parker, & Nakayama, 2003; Duchaine, Yovel, Butterworth, & Nakayama, 2006), focus this investigation on the recognition of facial expression. Performance on each process is examined using two different tests, to allow a more conservative assessment of the cognitive presentation of each participant.

4.1. Perception of facial identity

CFPT (Duchaine et al., 2007): This test requires participants to arrange six faces displayed from a frontal viewpoint in order of their similarity to a target face that is presented in a three-quarter viewpoint. The six test faces were created by morphing target faces with distractor faces. Participants complete 16 trials in total: eight with the faces upright and the remainder in an inverted format. Performance on the CFPT is measured as the total number of errors (i.e., how far away the participant is from a perfect arrangement), so that a lower score reflects better performance. Because there is some variability in the scores achieved by typical participants on the CFPT (Bowles et al., 2009), control data was collected from a larger sample of controls (N = 58, see Table 3). Nevertheless, the standard deviation for our sample was still relatively large (as observed in previous work, Russell et al., 2012), preventing any single-case analyses on the upright condition from reaching significance (all *p*s > .17)[3]. It is of note, though, that all participants bar one (CH) outperformed controls by at least one standard deviation. Further,

[3] Analysis of the inverted condition on this test is presented in the discussion of configural/holistic processing below (see section 5.4).

Table 3. Results from the tasks administered in Study 2. All values for SR participants are expressed in the number of SDs away from the control mean.

	Controls			Super-Recognisers					
	Mean	SD	N	CH	DF	JN	GK	CW	TP
Perception of facial identity									
Matching test (upright faces, d'):	2.00	0.40	21	1.60	2.90*	3.30*	0.10	1.10	1.80
CFPT[a]:									
Upright	35.90	15.00	58	-0.70	-1.60	-1.10	-1.30	-1.30	-1.10
Inverted	61.80	11.40	58	0.60	-1.20	0.20	-1.60	-0.30	-1.70
Perception of facial expression									
Ekman 60[b]	51.7/60	4.2	30	-0.60	0.80	0.80	-1.10	0.30	0.80
Mind in the Eyes accuracy[c]	27.6/36	4.0	29	-0.20	0.60	0.10	-0.40	1.40	0.90
Mind in the Eyes RT	7001.19	1872.57	28	-0.69	0.02	0.31	-0.44	-1.00	1.13
General socio-emotional functioning									
EQ[d]:									
Females	50.6/80	9.20	-	-	-	0.70	-	-	-
Males	41.3/80	10.10	-	0.50	1.00	-	2.50*	-0.90	0.50

* indicates participant significantly differed to controls using Crawford et al.'s (2010) modified *t*-tests for single-case comparisons ($p < .05$)

[a]Cambridge Face Perception Test (Duchaine et al., 2007), lower score indicates better performance; [b]Ekman 60 faces test (Young et al., 2002); [c]Reading the Mind in the Eyes (Baron-Cohen et al., 2001); [d]Empathy Quotient (test and norms from Lawrence et al. (2004) – separate norms are provided for male and female participants.

the scores that were achieved are similar to those reported by Russell et al. (2009), which were significantly better than controls in a group-based analysis.

Matching test: Given the statistical difficulties in identifying superior performance on the CFPT, we further assessed face perception skills by considering performance in the "upright face" condition of our matching task described above (see section 3.1 and Table 2). On this task, two of the SRs – JN and DF – showed an exceptional ability to match upright faces compared to controls. Pertinently, DF also achieved the most proficient score on the upright condition of the CFPT.

4.2. Perception of emotional expression

Ekman 60 faces (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002): In this task participants have to label pictures of actors with one of six emotions: anger, happiness, sadness, fear, surprise, or disgust. The emotions are presented by ten actors (four male and six female) and displayed on screen for five seconds. Participants indicate their response by a mouse click on the relevant tab describing the emotion. Single-case comparisons showed no significant differences between any SR and the control group on this test (all *p*s > .1), and no individual SR performed above one standard deviation of the control mean (see Table 3). No reaction time data was available for analysis on this test.

Reading the Mind in the Eyes Test (RMITE: Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). In the RMITE test, participants are presented with 36 images of the eye region of actors (males and females) and provided with a description of four mental states to choose from. Unlike the Ekman 60 faces task which exposes participants to six basic emotional states, the emotion descriptions in this test differ very subtly from one another. Images are presented on screen for an unlimited amount of time and responses are elicited by a key press. Again, single-case comparisons showed no significant differences between any SR and the control group (all *p*s > .1).

Because accuracy was high on this test, we additionally examined the speed of participants' responses. None of the SRs performed significantly faster on this test, and it is of note that the SR (CW) who performed most accurately was one SD slower than controls. However, the RMITE test does not impose a time constraint on participants and it is possible that CW engaged in a more effortful analysis of the stimuli. Nevertheless, there is little evidence to suggest that superior face recognition is also accompanied by enhancements in facial expression recognition.

Empathy Quotient (EQ: Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004): Finally, we also collected a brief self-report measure of socio-emotional functioning in the SR group. Previous work indicates that, in the typical population, higher scores on the EQ are associated with more advanced face recognition skills (Bate et al., 2010), indicating that enhanced socio-emotional rather than visuo-cognitive processes may sometimes contribute to face recognition performance. Importantly, the EQ does not measure socio-emotional functioning using facial stimuli, but consists of 60 self-report questions. Forty items tap into participant's level of empathy, whereas 20 filler items (which are not analysed) are included in the scale to distract participants from a continuous focus on empathy. There are four response options, and the test has a maximum score of 80. TP, CH, JN, CW and DF performed within one standard deviation of the control norm on this test (see Table 3). One SR, GK, achieved a score significantly higher than the gender-relevant control norm, $t(24) = 2.88$, $p = .008$, $Z_{cc} = 2.941$, 95% CI [2.020, 3.849]; estimated % population below his score = 99.59. Interestingly, however, this individual achieved scores that were lower than the control mean on both the Ekman 60 Faces and RMITE tests. Hence, enhanced socio-emotional processing does not seem to boost the recognition of facial emotional expression in GK, although it is possible that it may underpin this individual’s superior facial identity recognition skills.

4.3. Summary of Study 2

This is the first study to examine whether SRs excel at the perception of facial information other than identity. The results suggest that, at least in comparison to the perception of emotional expression, SRs’ abilities are specific to identity-based tasks. These findings argue against the idea that SRs are better at extracting information from faces in general – instead, it appears that SRs are particularly adept at extracting and/or using facial identity information.

5. STUDY 3: PROCESSING STRATEGIES IN SUPER-RECOGNITION

The final investigation considers the actual processing strategies used by SRs. It may be that these individuals use qualitatively different processing strategies to typical perceivers. However, as discussed above, there are a number of perceptual processes that are thought to be particularly important for efficient face-processing: for example, making fine-grained discriminations about the spacing between facial features (configural processing), or integrating information from across the entire face (holistic processing); and it may be that SRs show particularly efficient visual processing of faces as measured by these indices. Here, we present a series of experiments that examine stimulus-general holistic processing (i.e. the tendency to process generic stimuli at a global level, also referred to as global precedence) and more face-specific configural/holistic processing skills.

5.1. The Navon task

Global precedence was examined using a global-local task that requires participants to identify letters at various scales (Navon, 1977). In this test, participants are presented with composite stimuli of small letters making up big letters (e.g., many small “S” letters arranged in the shape of the letter “H”), and asked to identify either the large or small letter. In this version of the test, the stimuli were presented in four different positions, so participants could not focus on any particular part of the screen. The test was divided into four blocks of 48 trials each. In two blocks, volunteers had to respond to the large letter and in the other two blocks, they responded to the small letter. On half of the trials the composite letters were congruent (small and large letters were the same) and on the other half they were incongruent (small and large letters were different).

In order to examine whether the extraordinary performance of SRs in facial identity tasks results from a stronger global bias, an index of global bias was calculated (Duchaine, Germine, & Nakayama, 2007) by dividing average global RT by average local RT ([Global congruent RT + Global incongruent RT] / 2)/([Local congruent RT + Local incongruent RT] / 2). Index values below one indicate a global bias; index values above one indicate a local bias. Comparisons for individual SRs revealed that JN’s global bias index was significantly lower than the control group, JN: $t(27) = -2.95$, $p = .003$, $Z_{CC} = -3.00$ (95% CI: -3.87 - -2.12), estimated % of population below JN’s score = 0.33%, suggesting a particularly strong bias to process stimuli globally (see Table 4). None of the other SRs showed a similar effect (all *p*s > .1).

5.2. Inversion effects

Much previous work has examined face-specific holistic and configural processing by comparing performance on an upright face recognition task with performance on an inverted condition. The two face perception tasks described above (see Section 4.1) contain both upright and inverted conditions, and we revisit the findings of these tasks to evaluate the use of configural processing in super recognition.

Table 4. Results from the configural processing tests described in Study 3. All values for SR participants are expressed in the number of SDs away from the control mean.

	Controls			Super-Recognisers					
	Mean	SD	N	CH	DF	JN	GK	CW	TP
Navon task (global bias index [a])	0.9	0.10	28	0.50	0.60	-2.80*	0.50	0.70	-0.10
CFPT (inversion index [b])	1.0	0.80	58	2.00	3.00*	2.20	1.90	2.70*	1.10
Matching test (faces inversion effect [c])	1.0	0.61	21	1.56	2.51*	3.03*	0.56	1.62	1.82
Composite task (composite effect [d]):									
Faces upright	314.4	368.1	29	-0.70	0.00	0.50	-0.70	-0.70	2.50*
Faces inverted	3.40	213.5	29	-0.20	0.00	0.10	0.60	-2.00	0.10
Dogs upright	-24.00	164.21	29	-0.46	1.89	-0.88	1.58	-0.84	-2.24
Dogs inverted	-38.10	173.83	29	0.94	0.40	-0.81	-0.87	-1.14	2.41

* indicates participant significantly differed to controls using Crawford et al.'s (2010) modified *t*-tests for single-case comparisons ($p < .05$)

[a]Test from Navon (1977), global bias index from Duchaine et al. (2007); [b]Inversion index = (upright-inverted)/upright (calculated using total errors in the upright and inverted condition; Russell et al., 2009) ; [c]Inversion effect = d' (upright) – d' inverted; [d]Composite effect = IE(misaligned) – IE(aligned) (Robbins & McKone, 2007).

CFPT: To examine whether SRs showed a disproportionate inversion effect, we subtracted each participant's score for inverted trials from their score for upright trials, then divided it by their score for upright trials to create an inversion index ([upright-inverted]/[upright]; Russell et al., 2009). The mean inversion effect for the SR group in this study was 2.14 (SD = 0.6), in line with Russell et al. (2009), who reported the inversion effect of SRs to be 2.3 (SD = 0.2). Single case analyses on the inversion index revealed an enhanced effect of inversion for two SRs, CW, $t(57) = 2.15$, $p = .038$, Z_{cc} = 2.171, 95% CI [1.691 – 2.636], % of population below CW's score: 98.09; and DF, $t(61) = 2.59$, $p = .01$, Z_{cc} = 2.611, 95% CI [2.069 – 3.156], % of population below DF's score: 99.35 (see Table 4). Although the inversion indices of the remaining SRs did not significantly differ from the control group (all ps > .12), it should be noted that CH, JN and GK all achieved scores that were approximately two SDs above the control mean. This finding may be interpreted as evidence that all SRs other than TP show evidence of heightened configural processing (i.e. that was approximately or above 2 SDs from the control mean) for upright faces.

Matching test: Analysis of performance on the upright versus inverted "face" conditions of this task provides a further assessment of configural processing with respect to inversion effects. RSDT comparing the inversion effect of individual SRs for faces revealed that JN and DF showed a significantly greater effect of inversion than controls for faces, JN: $p = .004$ $Z_{DCC} = 3.30$ (95% CI: 2.18 – 4.57), estimated % of population showing a larger difference than JN = 0.21%; DF: $p = .013$, $Z_{DCC} = 2.75$ (95% CI: 1.78 – 3.85), estimated % of population showing a larger difference than DF = 0.68%. Three (CH, CW, and TP) of the remaining four SRs performed more than 1.5 SDs above the control mean, with only GK performing in a similar manner to controls (see Tables 2 and 4).

Hence, enhanced inversion effects are most consistently seen across the CFPT and matching task in three of the SRs (CW, DF and JN), with trends also noted consistently in CH. TP and GK only showed a trend towards a heightened inversion effect in one of the two tasks.

5.3. The composite task

While inversion effects have traditionally been used to evaluate configural processing skills, it is generally accepted that they only offer reasonable indicators of the measure and are not directly diagnostic of processing style (e.g., Valentine, 1988). For example, a disproportionate effect of inversion may arise due to difficulties processing local feature information, rather than a more integrative processing style per se (McKone & Yovel, 2009). The composite task is seen as a more direct measure of holistic processing (Rossion, 2013), although performance on this task is variable even in typical perceivers, making it difficult to interpret null results in single case analyses (Konar et al., 2010; Richler et al., 2011). Nevertheless, we administered this task to our SR group.

In the composite task, participants are presented with faces that have been cut in half. The top half of one face is combined with the bottom half of another face, either aligned (i.e., creating the impression of a full face) or misaligned (the two halves are offset). Previous studies have found slower or less accurate performance in face matching tasks when the face halves are aligned than when they are misaligned (e.g., Le Grand, Mondloch, Maurer, & Brent, 2004; Robbins & McKone, 2007, Young et al., 1987). This effect is thought to reflect holistic processing – when the faces are aligned, participants automatically integrate information from the irrelevant bottom halves of the composite faces, which creates the percept of two different faces. When the faces are not aligned, no holistic processing occurs, and participants are able to match the top halves without interference from the irrelevant bottom half (Rossion, 2013). Since holistic processing is thought to be disrupted when faces are presented upside-down (Maurer, et al., 2002), and to be reduced or not be present for objects other than faces (McKone & Robbins, 2011), the same effect does not occur for inverted composite faces or objects other than faces. Thus, if SRs show increased holistic processing of faces but not other objects, we would expect greater interference (i.e., worse performance) than controls for upright aligned faces when compared with upright misaligned faces. If this effect is related to face-specific processing, the same pattern of results would not be present for inverted faces or objects.

In this study, we adapted the composite paradigm used by Robbins and McKone (2007) to examine holistic processing for faces and dogs[4]. Participants were presented with two composite faces or dogs sequentially. The first stimulus appeared for 600 ms, the second stayed onscreen until the participant responded. The stimuli were offset by 25% of the screen size, to prevent matching based on the size or location of the stimuli or features. Participants were asked to indicate as quickly and as accurately as possible whether the top halves of the face or dog (the section with the eyes) were the same or different. The stimuli were identical to those used by Robbins and McKone (2007), except that only 30 stimuli (15 same identity, 15 different identity) were presented in each condition (upright and inverted; aligned and misaligned; faces and dogs). Trials were blocked by object and orientation, with aligned and misaligned trials presented randomly within each condition. As some participants show a composite effect for accuracy, but not reaction time (RT), and other participants show the opposite effect, we used the combined measure inverse

[4] There has been much debate in the literature over the use of this traditional composite task (sometimes referred to as the "partial design") in comparison to a longer version (sometimes referred to as the "complete design") (see Gauthier & Bukach, 2007; McKone & Robbins, 2007; Richler & Gauthier, 2013, 2014; Rossion, 2013 for an overview). We elected to use the current version for several reasons: primarily, the fact that the complete design has been shown to elicit a strong composite effect for inverted faces and objects (e.g., Richler, Mack, Palmeri, & Gauthier, 2011), whereas the stimuli used by Robbins and McKone (2007) show no evidence of a composite effect for either stimulus. Other theoretical justifications for the use of the traditional composite task (e.g., the perceptual and neural locus of the effect) have been comprehensively reviewed by Rossion (2013).

efficiency (IE) ([reaction time]/[accuracy], Townsend & Ashby, 1978;1983) to assess the extent of the composite effect (Rossion, 2013). Follow-up analyses were conducted on the composite effect ([IE aligned] – [IE misaligned]) for each stimulus and orientation.

Control participants showed a typical pattern of results: there was a significant interaction between stimulus (face and dog), orientation (upright and inverted), and alignment (aligned and misaligned), $F(1,28) = 11.48$, $p = .002$, $\eta_p^2 = .29$. Follow-up *t*-tests (Bonferroni corrected) on the composite effect found a significantly greater effect of alignment for upright faces than for inverted faces ($p = .001$) or upright dogs ($p < .0005$), suggesting stronger holistic processing for upright faces than inverted faces or non-face stimuli (see Table 4).

One SR (TP) showed a significantly stronger composite effect for upright faces than controls, $t(29) = 2.16$, $p = .04$, $Z_{CC} = 2.193$ (95% CI: 1.51- 2.86), estimated % of population below TP's score = 98.01%. However, none of the other SRs showed an enhanced composite effect (all *p*'s > .6). To examine whether TP's composite effect was disproportionate for upright faces (i.e., whether this reflects face-specific mechanisms or a more general proficiency at holistic processing), we carried out RSDT comparing the composite effect for upright faces to that for inverted faces and upright dogs. The difference in composite effects for upright and inverted faces was within the normal range compared to control participants ($p = .14$). However, TP showed a significantly stronger composite effect for faces than for dogs when compared to control participants, $p = .001$, $Z_{CC} = 3.59$ (95% CI: 2.63- 4.64), estimated % of population showing a larger difference than TP = 0.06%. This indicates that TP was not showing an increased composite effect for all stimuli – rather, he showed evidence of enhanced holistic processing specifically for faces.

5.4. Summary of Study 3

Study 3 initially examined whether SRs display an enhanced general global processing bias via performance on the Navon task, and this was only observed in one participant (JN). Evidence of enhanced face-specific configural/holistic processing was investigated using face inversion effects, where they were consistently observed in three SRs (CW, DF and JN), and trends were noted across both tasks in CH. However, TP and GK only showed a trend towards a heightened inversion effect in one of the two tasks. Finally, we examined holistic processing using the composite task, where it is more difficult to observe significant differences in single-case comparisons. However, TP demonstrated enhanced holistic processing of faces on this test. In sum, while a more generalised global bias was observed in one participant, evidence of enhanced face-specific processing was observed in all participants but GK.

6. THE COGNITIVE HETEROGENEITY OF SUPER RECOGNITION

Analysis of the performance of each individual SR on the above battery of tests permits insights into the cognitive presentation of super recognition, and its potential underpinnings. We were particularly interested in examining whether the same processes might underpin the superior face recognition abilities in all six participants, or whether the presentation of super recognition is heterogeneous.

Table 5 summarises the pattern of performance observed in each SR. While all SRs were required to demonstrate superior recognition of facial identity in order to be included in this study, it is of note that only two of the six participants outperformed controls on any of the object-processing tests. While GK displayed heightened performance in recognizing inverted hands, TP outperformed controls on a test of object memory. This finding indicates that super recognition is domain-specific in at least some participants.

Notably, evidence of enhancements in facial identity perception was only observed in DF and JN, with a trend also observed in TP. However, no participant displayed enhanced recognition of emotional expression, indicating that when enhancements to face perception do occur, they can be restricted to the processing of facial identity. Further, only one participant (GK) displayed enhanced general socio-emotional processing skills in comparison to the typical population. While this did not improve the participant's ability to recognise facial expression, it may nevertheless contribute to his superior facial identity recognition skills.

Finally, while all participants showed at least a trend towards enhanced face-specific configural/holistic processing (with significant effects noted in four SRs: DF, JN, CW and TP), one participant showed a more generalized bias towards global processing (JN). Taken together, this pattern of findings suggests that (a) not all cases of super face recognition also present with enhancements in facial identity perception, and (b) while there is convincing evidence of domain-specificity in three of the six participants, more generalized visuo-cognitive processes or socio-emotional skills may underpin super recognition in the remaining three participants.

Table 5. The overall pattern of performance noted for each of the six SR participants. A tick refers to cases where a significant enhancement occurred on at least one test, and “T” to a non-significant trend, classed as performance above 1.8 standard deviations from the control mean.

	Super-Recognisers					
	CH	DF	JN	GK	CW	TP
Facial identity recognition	✓	✓	✓	✓	✓	✓
Object-processing				✓		✓
Facial identity perception		✓	✓			T
Facial expression recognition						
General socio-emotional functioning				✓		
General global processing bias			✓			
Face-specific configural/holistic processing	T	✓	✓	T	✓	✓

7. GENERAL DISCUSSION

In this chapter, we report a detailed cognitive assessment of the face- and object-processing skills of six individuals who meet the published diagnostic criteria for super recognition. We specifically addressed four key theoretical issues: (a) the domain-specificity of super recognition, (b) the precise locus of superior processing skills within theoretical models of face recognition, (c) whether super recognition is underpinned by qualitative differences in face-processing strategies, and (d) the cognitive heterogeneity of the phenomenon. Each of these issues is discussed in turn below.

7.1. Domain-specificity of super recognition

Two tests assessed the ability of SRs to process non-facial stimuli: one required the matching of faces compared to houses and hands, and the other test assessed memory for cars (the CCMT) in a paradigm that replicates the standard version of the CFMT. Enhanced performance was only observed in the object conditions of the matching task in one SR participant (GK). However, enhanced performance was only observed in the inverted hands condition in this participant, and not the remaining three object conditions, nor the inverted faces condition. This finding therefore only presents limited evidence regarding the object matching skills of this individual, and it is of note that he did not outperform controls on the CCMT. However, the latter was observed in one other participant (TP). Given the matching task measures different processes to the CCMT, it is possible that TP's enhanced performance on this test results from generalised superior memory skills which would not have aided his performance on the matching task.

Most significantly though, four of the six SRs displayed domain-specificity for faces in the first investigation. Given later investigations indicated one of these participants (JN) also displayed an enhanced general global processing bias, a more conservative conclusion is that domain-specificity for faces was observed in three of the six participants, providing further support for hypotheses that face recognition is a specialised process.

7.2. Locating superior recognition within cognitive models of face-processing

Our second study attempted to locate the underpinnings of super recognition within cognitive models of face-processing. Given that prosopagnosia can broadly be partitioned into two subtypes, one involving deficits in face perception and the other higher-order impairments affecting mnemonic processes (de Renzi et al., 1997), it is possible that a similar pattern may underpin super recognition. That is, the skill may result from an enhancement in face perception (i.e. at the level of structural encoding, see Figure 1), or at later stages of the model (i.e. at the level of the FRUs and the PINs). Further, the hierarchical assumptions of dominant cognitive models of face-processing would suggest that an enhancement at the level of structural encoding would also influence the processing of other aspects of face perception, such as the recognition of emotional expression.

The pattern of findings reported here suggests that only two of the six SRs (DF and JN) present with a facilitation in facial identity perception, with a further trend noted in TP. As stated above, both JN and TP may benefit from more generalised enhancements that result in their superior face recognition skills, but DF presents with domain-specific superior face recognition skills. The remaining three SRs only displayed a facilitation at the level of face memory, suggesting that super recognition may be underpinned by enhancements at latter stages of processing. Given that some SRs may benefit from a facilitation in structural encoding, a second theoretical issue questions whether they are proficient at all aspects of face perception, or if their skills are restricted to the perception of facial identity. We failed to find any evidence of an enhancement in facial expression recognition in any SR, suggesting that any enhancements in face perception are specific to facial identity. However, we did not assess the ability of SRs to perceive other aspects of facial information, such as age or gender, and it is possible that some enhancements also present in these processes. Nevertheless, the data reported here suggest that (a) only some SRs present with an enhancement in facial identity perception, and (b) that this may aid the construction and utilisation of view independent representations that are used specifically in facial identity recognition.

7.3. Processing strategies in super recognition

The third investigation examined whether super recognition is underpinned by specific processing strategies, such as a generalised bias to process visual stimuli globally, or enhancements in face-specific processing styles. Only one SR (JN) displayed a greater generalised bias towards global processing, which is likely to assist with face recognition (Macrae & Lewis, 2002) and may provide an explanation for her skills that is not domain-specific.

Our investigation into the use of face-specific (i.e. configural/holistic) processing strategies focused around face inversion effects in two perceptual tasks, and performance on a composite test that used faces and dogs as stimuli. Enhanced inversion effects were observed in at least one of the two perceptual tasks in three SRs (DF, JN and CW), with trends also observed on at least one test in the other three participants. Inversion effects are generally interpreted as reflecting a disruption of face-specific configural or holistic processing (Maurer et al., 2002). It is possible that SRs in general show particularly strong integration of information across upright faces; or heightened sensitivity to spatial relations between features, and that these skills contribute to their exceptional ability to identify faces. However, it is also possible that SRs are particularly good at extracting facial feature information (which is also affected by inversion; McKone & Yovel, 2009). As we did not manipulate spacing or featural information in any of the tasks, our results cannot speak to SRs' ability to process isolated features; nor can our results discriminate between the two purported face-specific processes (configural or holistic).

It is important to note that an inversion effect (even a disproportionate one) for faces alone does not confirm that SRs show heightened face-specific processing skills. The inversion effect was also examined for two other classes of objects – houses and hands – and four of the six SRs showed typical effects of inversion compared to controls (DF, JN, CW, TP). In other words, for these four cases, the mechanisms underpinning the heightened inversion effect did not generalise to other objects.

Interestingly, while CH's inversion effects for faces in the CFPT and the matching task were on average (albeit non-significantly) greater than those of controls, he displayed an enhanced inversion effect for hands. It is thus possible that his extraordinary face recognition ability is underpinned by more general and object-relevant processing strategies, or a particular proficiency for the discrimination of biological stimuli. This specific finding is of particular relevance to the literature supporting the domain-general organisation of the human brain and the expertise account of face processing (Curby & Gauthier, 2014; McGugin, Van Gulick, & Gauthier, 2015).

The final case, GK, showed a disproportionate inversion effect for houses, although this reflects extremely poor performance (and perhaps misunderstanding of the task) in the inverted houses condition, rather than heightened performance in the upright condition. As such, it is still reasonable to conclude that the somewhat larger inversion effect for faces in the SRs reflects some level of enhanced face-specific processing in at least five out of the six cases.

Only one SR (TP) demonstrated enhanced holistic processing on the composite test. While this finding adds support to the hypothesis that this process may underpin superior face processing skills in this individual, the null effects observed for the other SRs are more difficult to interpret. On one hand, large-scale studies that have examined individual differences in the composite task and face recognition abilities have not always found a significant link between the two measures (e.g., Konar et al., 2010), and fairly low correlations have been reported in studies that have detected an association (r = .13, Wang et al., 2012; r = .40-.48, Richler et al., 2011a). This indicates that holistic processing may only play a small role in determining individual differences in face-processing, which would be entirely in line with the null effects for the SRs in the current study. On the other hand, several researchers have noted that it is difficult to draw conclusions about individual differences from composite effects due to fairly low reliability of the measure (e.g., Richler & Gauthier, 2014; Rossion, 2013). Put simply, a large number of factors could have introduced noise into the composite measure (for both control participants and SRs), which may have obscured potentially significant differences between the groups. This suggests that it may be the measure of holistic processing, rather than the underlying theoretical construct, which led to null results for the majority of the SRs in this study. The fact that the majority of SRs showed a heightened effect of inversion for faces (i.e., some evidence of enhanced face-specific processing) points to the latter explanation.

In sum, the evidence reported here suggests that at least five out of the six SRs display heightened face-specific processing skills – even those who benefit from other facilitations in domain-general processes.

7.4. The cognitive heterogeneity of super recognition

Throughout the battery of tests administered to the SRs, we were particularly interested in examining whether the same processes might underpin the superior face recognition abilities in all six participants, or whether the presentation of super-recognition is heterogeneous. It is of note that single-case analyses revealed a disparate pattern of findings between the six SRs that may account for the superior face recognition skills in some cases. Specifically, enhancements in object processing were tentatively noted in GK and TP; GK also demonstrated enhanced socio-emotional functioning (which may result in an enhanced interest in faces or more frequent social interactions), and JN showed a generalised bias towards global processing. These findings raise the possibility that enhancements in various generalised processes may contribute towards super recognition in some cases.

Further, some disparity was noted in the face-processing profiles observed across the six SRs, and even in the three whose super recognition appears to be underpinned by enhancements in face-specific mechanisms. Specifically, DF presented with enhancements in both the perception and recognition of facial identity, whereas the superior skills of CH and CW were limited to only identity recognition. While this pattern of findings is accommodated by the predictions of dominant cognitive models of face-processing (e.g. Bruce & Young, 1986), it remains to be seen whether some individuals may present with enhanced face perception skills that do not extend to face memory performance. The hierarchical assumptions of theoretical models would make interpretation of such findings problematic, but it is possible that such presentations may arise via repeated rehearsal in some applied settings. For instance, there is growing interest in super recognition in policing and national security settings, with reports of officers who are able to proficiently match faces across a variety of low-quality stimuli. Studies that investigate such self-reported cases and that screen the general population to assess the prevalence of super recognition may therefore bring novel case studies to light that aid the refinement of current theories of face-processing.

One point of interest is that at least five of the six SRs presented with heightened configural/holistic processing strategies, although in two individuals this was reflected by non-significant trends that were at least 1.9 standard deviations above the control mean. This is the most consistent finding across the battery of tests that were administered to the SRs, suggesting that these individuals differ from typical perceivers in the strength or efficiency of their face-specific processing skills. Hence, heightened configural and/or holistic processing may represent a common underpinning mechanism across even heterogeneous cases of super recognition, and therefore may be used as an additional diagnostic indicator to detect super recognition.

It is worth noting that the concept of holistic processing is a topic of significant debate, with many authors offering different and sometimes conflicting definitions of the term (Piepers & Robbins, 2012). Richler and Gauthier (2012) proposed that inversion effects measure one specific aspect or meaning of holistic processing (that most closely aligned with the term configural processing), whereas other face-specific processing styles that also come under the rubric of "holistic processing" may be measured more efficiently through other tasks. For example, the composite effect may be the most appropriate measure of failure of selective attention to face parts; whereas the part-whole task might be the best measure of whether face parts are processed more efficiently in the context of a whole face. If Richler and Gauthier's (2012) proposal is correct, the SRs in this study appear to excel specifically at configurally-based holistic processing – this suggests that they should also show heightened sensitivity to changes in spatial relationships within a face. However, it does not necessarily mean that SRs will show heightened holistic processing according to other definitions, which may explain why only one SR in the current showed a significantly greater composite effect than controls. More thorough investigations of individual and group differences on these varying measures of "holistic processing" are required to parse out whether SRs excel at all aspects of face-specific processing, or whether some meanings or measures of holistic processing are especially predictive of superior face recognition skills.

7.5. Conclusion

In sum, this investigation presents evidence that super face recognition is heterogeneous in its presentation, and in some cases may be underpinned by enhancements in more generalised processes. However, half of our SR sample displayed proficiencies that were face-specific, but nevertheless varied in whether facial identity perception was also enhanced. A facilitation in configural/holistic mechanisms was more consistently noted across the SR group, suggesting SRs have more developed face-specific processing strategies than typical perceivers. Such measures may present an additional indicator of superior face recognition skills.

Chapter 3: Eye-movement strategies in developmental prosopagnosia and "super" face recognition

Chapter published as:

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (In press). Eye-movement strategies in developmental prosopagnosia and "super" face recognition. *The Quarterly Journal of Experimental Psychology*.

1. INTRODUCTION

While Chapter 2 examined the cognitive underpinnings and face-specificity of super recognition, the positioning of these individuals on the face recognition spectrum remains unknown. This is an important theoretical issue given increasing evidence that the face recognition spectrum is wider than previously thought, and hypotheses that individuals with developmental prosopagnosia (DP) are simply those at the "bottom end of normal" as opposed to those with a condition in its own right. This theory has been bolstered by findings that SRs exist, who seem to be as good at face recognition as DPs are bad, presenting evidence for two extremes. However, SRs and DPs have not yet been compared within the same investigation, limiting such claims. While it may be the case that the two simply represent the top and end bottom end of normal, it may be that one or both actually represent a qualitatively different type of processing.

An innovative means of addressing this issue is via the analysis of scanning strategy – a methodology that permits examination of qualitative face-processing strategies. As reviewed in Chapter 1, there are several lines of evidence that suggest acquired prosopagnosia is characterised by reduced attention to the eye region of the face (Ramon & Rossion, 2010; Xivry, Ramon, Lefevre, & Rossion, 2008), and, in some cases, increased attention to the mouth (Xivry et al., 2008; see also Bukach, Le Grand, Kaiser, Bub, & Tanaka, 2008). While little eye movement work has been carried out in DP that directly addresses this issue (Caldara et al., 2005; Schwarzer et al., 2007), it is conceivable that a similar pattern of findings will emerge. Further, one would predict that, if there is a full spectrum of face recognition ability where DPs and SRs represent the bottom and top ends, respectively, patterns of eye movements would vary along the entire spectrum according to the same measure. If this holds true, the proportion dwell time allocated to the eye region is a likely candidate. That is, if DPs spend less time on the eyes than controls, SRs should spend more time on this region, and the measure should also correlate with face recognition ability within typical perceivers.

However, other work using typical participants indicates that the nose may be the critical position for more successful face recognition (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012). Specifically, Hsiao and Cottrell (2008) found that both the preferred landing position (PLP – position where people fixate the most during visual tasks, Rayner, 1979) and the optimal viewing position (OVP – location of the first fixation associated with the best recognition performance, O'Reagan, 1981) are located to the left and to the centre of the nose, respectively. Furthermore, Peterson and Eckstein (2012) reported that the human visual system tends to optimise eye movements depending on the type of task, with identity judgements associates with OVP just below the eye region and close to the nose. If face recognition skills are associated with the proportion of dwell time spent looking at the nose, one would predict that, if DP is simply the tail end of the face-processing spectrum, these individuals would spend less time looking at the nose than controls. Alternatively, if this measure is associated with face recognition ability in both typical participants and SRs, yet DPs mirror the performance of individuals with acquired prosopagnosia (i.e. by spending less time on the eyes and more time on the mouth), this would provide evidence that DP represents a qualitatively distinct group that is independent from the typical population.

The current study addresses the issues discussed above. In Experiment 1, we employed a social scenes eye-tracking paradigm to examine the scanning strategies used by 10 individuals with DP in comparison to age-matched control participants. This paradigm was selected because previous work examining ASD suggests that analysis of the featural distribution of fixations is more fruitful when faces are presented within their natural context rather than as individual static images (see Birmingham, Bischof, &

Kingstone, 2008). This set of analyses specifically looked at more specific patterns of feature exploration, examining the distribution of fixations across the inner versus the outer facial features, and across the eyes, nose and mouth. A second benefit of adopting this paradigm is that it permits novel insights into the salience of faces in super recognition and DP (i.e. by examining the time taken to initially fixate on a face, and the proportion of dwell time allocated to faces versus bodies and background regions). While Chapter 2 explored the cognitive underpinnings of super recognition with reference to the stages of processing posited by Bruce and Young (1986), it is also conceivable that the locus of the phenomenon may be earlier than allowed by the sequential framework. Indeed, speeded face detected and enhanced attention to faces may account for super recognition (or DP) in some individuals.

Experiment 2 used the same paradigm to explore scanning strategies in eight individuals who meet the published diagnostic criteria for super recognition. Given this group significantly differed to the DP group according to age, we conducted this as a separate experiment with an independent age-matched control group for each sample. However, the aims of this experiment were akin to those for Experiment 1: the paradigm allowed us to examine the salience of faces to SRs, and to examine their featural exploration of faces in comparison to matched controls - all within an ecologically valid context. Further analyses examined scanning strategies in the two control groups, to investigate (a) whether any differences between DPs and SRs could simply be attributed to age, and (b) whether attention to the eyes or nose correlated with face recognition skills in typical perceivers. Finally, we conducted a third experiment to attempt to replicate the SR findings using a different paradigm. Indeed, Experiment 2 represents the first eye-movement investigation into superior face recognition, and if its results are robust, we expected them to be reproduced within a more traditional single-face instruction-based encoding task.

2. EXPERIMENT 1

Our first experiment represents the first in-depth analysis of patterns of feature exploration in DP, and adoption of the "social scenes" paradigm also permitted analysis of visual attention to the entire face, and the critical region covering the inner features. Specifically, we replicated the methodology used by Riby and Hancock (2008), and asked 10 participants with DP and 20 matched controls to free-view a set of static images displaying social scenes while their eye movements were monitored.

2.1.Method

2.1.1. Participants

A group of 10 adults with DP took part in this study (7 female, mean age = 57.8 years, SD = 7.1). All participants had undergone neuropsychological testing prior to the investigation to confirm their prosopagnosia. These findings and participants' demographic information are summarised in Table 1. A full description of this battery of tests is reported elsewhere (Bate, Haslam, Tree, & Hodgson, 2008; Duchaine et al., 2007), and these tests are used by several laboratories for background neuropsychological assessments of DP participants (e.g. Bowles et al., 2009; Duchaine et al., 2007; Lee et al., 2010). Critically, all participants performed at least two standard deviations below published control means on the Cambridge Face Memory Test (CFMT: Duchaine & Nakayama, 2006) and a famous faces test that was created and standardized within our laboratory (see Bennetts et al., in press). Some participants were also impaired on the

Cambridge Face Perception Test (CFPT: Duchaine et al., 2007), but it should be noted that impaired performance on this test is not required for a diagnosis of DP. As is noted for acquired prosopagnosia, the condition is heterogeneous in its cognitive presentation, and while some individuals experience deficits in face perception as well as face memory, others only experience difficulties in the latter (see Bate et al., 2014). None of the DPs reported socio-emotional or low-level visual or intellectual difficulties. Indeed, as summarised in Table 1, all of their IQs (estimated using the Wechsler Test of Adult Reading, WTAR: Holdnack, 2001) were high, and no atypical scores were noted on various tests of the Birmingham Object Recognition Battery (Riddoch & Humphreys, 1993). No abnormalities in basic low-level vision were observed using a standard Snellen letter chart (3m) or the Hamilton-Veale contrast sensitivity test.

Twenty (10 female) control participants were also tested, and were matched to the DP group on the basis of age (M = 51.5 years, SD = 6.9) and estimated IQ (using the WTAR). All participants provided written consent and were rewarded with a small monetary payment in exchange for their time. Ethical approval for this experiment was granted by Bournemouth University's Ethics Committee.

Table 1. Demographics of DPs and performance in standard deviation units on tests of face-processing, lower-level vision and object recognition. 'CFMT' refers to the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), 'CFPT' to the Cambridge Face Perception Test (Duchaine et al., 2007), 'Mind in the Eyes' to the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001), and 'BORB' to the Birmingham Object Recognition Battery (Humphreys & Riddoch, 1993). DP scores are compared to published norms for each test (see each paper for control demographics). Note that the CFPT scores represent the number of errors, rather than the number of correct responses.

	Control Mean (SD)	**DP1**	**DP2**	**DP3**	**DP4**	**DP5**	**DP6**	**DP7**	**DP8**	**DP9**	**DP10**
Age		52	49	66	57	64	60	58	59	46	67
Gender		F	M	F	F	M	F	F	M	F	F
Hand		L	R	R	R	R	L	R	R	R	R
IQ		120	117	66	120	119	123	120	120	120	123
Face processing tests:											
CFMT	59.6/72 (7.6)	-2.3*	-2.7*	-2.7*	-3.5*	-4.2*	-4.6*	-2.6*	-4.03*	-3.2*	-2.8*
CFPT	36.7 (12.2)	-0.1	-4.9*	-2.2*	-0.9	-1.3	-5.2*	-4.2*	-1.42	-2.4*	-1.8
Famous faces	90.4% (7.7)	-6.8*	-2.2*	-4.9*	-6.7*	-6.1*	-9.2*	-9.1*	-8.67*	-9.3*	-7.2*
Mind in eyes	26.2 /36(3.6)	0.5	0.5	-1.4	0.2	-1.4	-0.9	0.5	-1.44	-0.3	0.2
Lower-level vision (BORB):											
Length match	26.9/30 (1.6)	-1.2	-1.8	-1.8	0.1	0.1	0.1	-1.2	-0.56	0.1	1.9
Size match	27.3/30 (2.4)	0.7	-1.0	-1.8	0.3	0.7	-1.8	0.3	0.71	0.7	-1.0
Orientation match	24.8/30 (2.6)	0.9	-0.7	0.5	0.1	0.5	-0.3	0.1	0.46	1.2	0.9
Position of gap	35.1/40 (4.0)	0.5	-0.5	1.0	0.5	-0.5	-0.3	1.0	0.48	0.7	1.0
Object decision test	52.4/64 (3.9)	0.2	-0.1	-0.4	0.9	-0.1	-1.1	-0.6	0.15	1.4	1.7

* indicates impaired performance

2.1.2. Materials

Twenty-five colour images were purchased from an online image database for use in this study. Twenty of the images displayed the faces and bodies of people who were engaged in social activities (e.g. a group of friends in a bar, a family having a picnic, and work colleagues meetings in an office; see Figure 1). Between two and six individuals were present in each scene, and their positioning varied across the images. The characters were not facing the camera, and were naturally engaging with each other. The remaining five images depicted natural scenes (e.g. a woodland or coastal image) that did not contain people. These “filler” images were included to keep participants naïve to the aims of the experiment. All images were adjusted to 27.09cm in length and 18.07cm in height, and subtended 20.48 degrees of visual angle when viewed from a distance of 50cm.

Eye-movements were recorded using the Eyelink 1000 system (SR Research Ltd, Canada), a video-based pupil/corneal reflex tracking device sampled monocularly at 2000 Hz with spatial accuracy of between 0.25 and 0.5 degree of visual angle. Head movements were minimized by the requirement that participants placed their head within a chin rest for the duration of the experiment. Eye position was monitored through an infrared CCD video camera that was placed on the desk in front of the participant. In an initial calibration phase and during the actual experiment, eye position on the screen was sent to a Dell host computer, which also collected information about when the stimuli were presented.

2.1.3. Design and Procedure

Participants were seated in a quiet room and were asked to place their head within the chin rest. A nine point calibration of eye fixation position was conducted prior to the experiment. The calibration procedure began with the presentation of a white dot in the centre of a black computer screen.

Figure 1. Example stimuli from Experiments 1 and 2. Black lines represent AOIs.

The dot moved consecutively around the edge of the screen until an adequate corneal lock was achieved in each position. Once each participant had successfully completed the calibration phase they immediately began the experiment. Because the test was administered in one continuous block recalibration was not required.

Participants were informed that they were going to view a set of images and that they should pay attention to each image and allow their eyes to naturally explore the stimuli. They viewed the sequence of 25 images (20 experimental and five filler images) in a random order, with an exposure time of five seconds per image. They were not required to make a response and the visual scanpath was recorded for the entire duration of the experiment. The initial point of retinal attention for each trial was controlled by the presentation of a centrally positioned fixation dot before the stimulus appeared.

2.1.4. Eye Movement Parameters and Statistical Analyses

Eye movements were analysed using Eyelink Data Viewer software (SR Research Ltd), which allows periods of fixation to be identified and user-defined areas of interest to be determined within the images. To investigate visual attention to faces, areas of interest (AOIs) were drawn onto the 20 experimental images using a freehand marquee tool (analyses were not performed on the five filler images). Three sets of AOIs were drawn onto each image (see Figure 1). The first set contained three AOIs: the background of the image (all areas other than the bodies and faces of the characters), the bodies of each character (taken from below the chin), and the faces of each character (including outer features such as the ears and hair). Second, the latter region was further divided into two separate AOIs, in order to investigate attention to the inner (i.e. the area covering the eyes, nose and mouth, and the spaces immediately between them) versus the outer (i.e. all remaining facial areas, including the ears and hair) facial regions. Finally, the "inner" AOI was subdivided into specific features, covering the eyes, nose and mouth.

Data were accordingly entered into three analyses of variance (ANOVAs). To examine visual attention to faces, a 3 (region: faces, bodies, background) x 2 (group: DP, control) mixed factorial ANOVA with repeated measures on the 'region' factor was performed. Second, to examine whether DPs spend less time on the inner features than controls, a 2 (facial region: inner, outer) x 2 (group: DP, control) mixed factorial repeated measures ANOVA was carried out. To examine the featural distribution of fixations, and to investigate previous findings that individuals with prosopagnosia spend less time on the eyes and (in some instances) more time on the mouth in comparison to controls, we carried out a 3 (feature: eyes, nose, mouth) x 2 (group: DP, control) mixed factorial ANOVA. Finally, a univariate ANOVA investigated whether DPs take longer to initially fixate a face than controls.

Effects involving a repeated measures factor are reported with *p* corrected for departures from sphericity using the Huynh-Feldt correction, where appropriate. Effect sizes are calculated using partial eta squared ($\eta\rho^2$). For each variable, participants are also compared to the control group on a single case level, using modified t-tests for single case comparisons (SINGLIMS, Crawford, Garthwaite, & Porter, 2010). Holmes' sequential Bonferroni procedure was used to correct for multiple comparisons where appropriate.

2.2. Results

Analysis of the regional distribution of fixations throughout the entire image revealed that all participants spent longer looking at faces (M = 58.52%, SE = 2.00) than either bodies (M = 24.35%, SE = 1.12) or the background of the images (M = 17.30%, SE = 1.14), $F(2,56)$, = 148.920, $p = .001$, $\eta\rho^2 = .842$. The difference in time spent on the bodies versus the background was also significant, $F(1,28) = 38.223$, $p = .001$, $\eta\rho^2 = .577$. A significant interaction between region and group was also observed, $F(2,56) = 4.754$, $p = .027$. Follow-up analyses indicated that while there was no difference in the time spent studying the background, DPs spent more time looking at bodies and less time looking at faces than controls: $F(1,28) = 2.506$, $p = .125$, $F(1,28) = 5.732$, $p = .024$, $\eta\rho^2 = .170$, and $F(1,28) = 5.170$, $p = .031$, $\eta\rho^2 = .156$, respectively (see Figure 2a). No main effect of participant group was noted in the ANOVA, suggesting DPs spent a similar length of time as controls in fixating on each image, and the findings did not result from a lack of engagement with the task, $F(1,28) = .043$, $p = .837$.

The second set of analyses examined the proportion dwell time spent on the inner versus the outer regions of the face. A main effect of region indicated that all participants spent longer looking at the inner (M = 41.10%, SE = 2.23) versus the outer (M = 17.42%, SE = 1.29) region, and this factor did not interact with participant group: $F(1,28) = 60.523$, $p = .001$, $\eta\rho^2 = .684$ and $F(1,28) = .002$, $p = .965$, respectively (see Figure 2b). A main effect of group was observed, indicating that DPs spent less overall time attending to faces (regardless of region) (M = 26.99%, SE = 1.63) than controls (M = 31.53%, SE = 1.15), $F(1,28) = 5.170$, $p = .031$, $\eta\rho^2 = .156$.

Analyses of patterns of facial feature exploration also indicated differences between DP and control participants. While no main effect of feature was noted, this factor did interact with participant group: $F(2,56) = 2.069$, $p = .143$ and $F(2,56) = 6.208$, $p = .006$, $\eta\rho^2 = .181$, respectively. Follow-up analyses indicated no differences in the proportion dwell time spent on the nose, but DPs spent less time on the eyes and more time on the mouth than controls: $F(1,28) = .323$, $p = .575$, $F(1,28) = 10.334$, $p = .003$, $\eta\rho^2 = .270$ and $F(1,28) = 4.848$, $p = .036$, $\eta\rho^2 = .148$, respectively (see Figure 2c). No main effect of group was noted in this analysis, in line with the above finding that DPs did not spend less time looking at the inner facial features than controls, $F(1,28) = .598$, $p = .446$.

A final analysis indicated that DPs did not take longer (M = 1024.17ms, SE = 62.77) than controls (M = 973.12ms, SE = 36.10) to first fixate upon a face, $F(1,28) = .572$, $p = .456$.

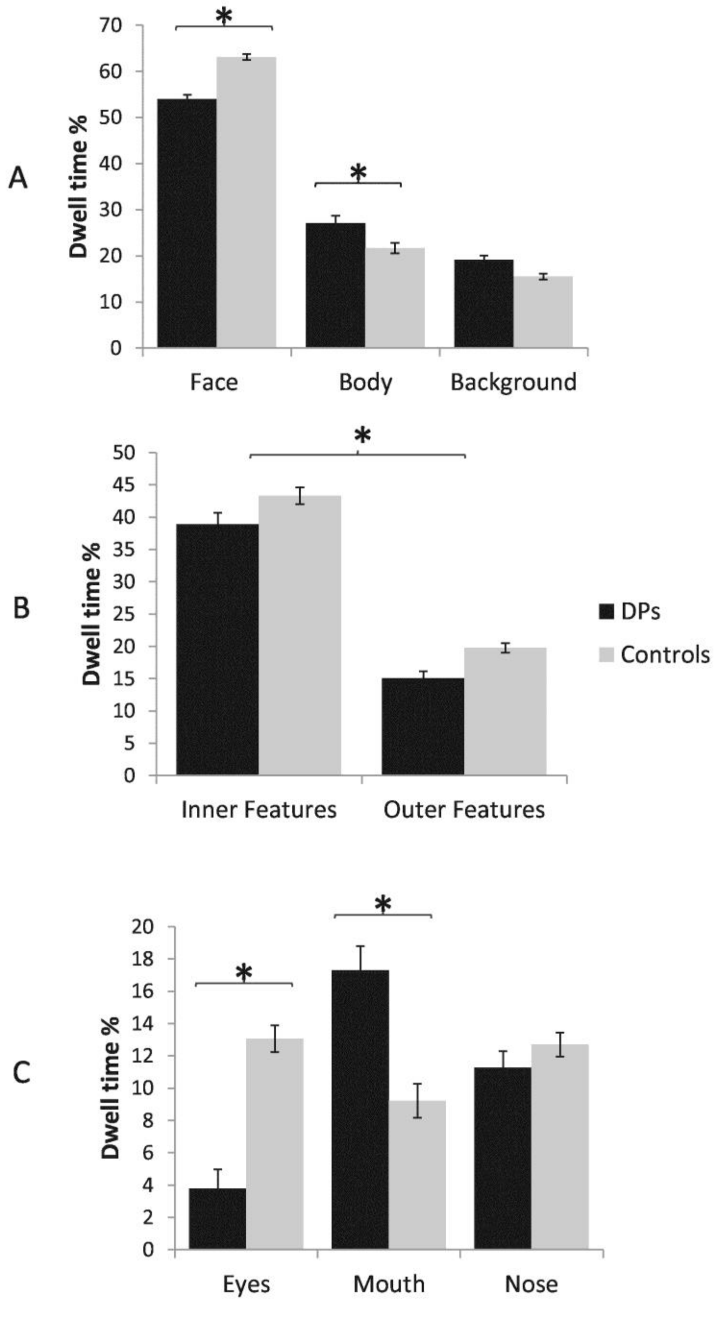

Figure 2. The percentage dwell time spent by DPs and controls on each region in Experiment 1. Error bars represent the standard error of the mean.

Because there is considerable heterogeneity in DP, it is prudent to examine data on a case-by-case as well as at a group-based level. Each individual's score on each measure was therefore compared to the control mean and standard deviation, using modified *t*-tests for single-case comparisons (see Table 2). With reference to the main group-based findings, only two of the 10 DPs spend significantly less time viewing the face than controls (DPs 1 and 10), and four spent a longer time viewing bodies (DPs 1, 2, 8 and 10). Although the proportion dwell time spent viewing the eye region was significant at the group level, no individual DP spent significantly less time than controls fixating this region. The proportion dwell time spent viewing the mouth was also greater in the DPs than controls at the group-level, yet significant case-by-case analyses were only observed in DPs 2 and 4, and a trend towards this pattern was observed in DP6. Finally, only one DP (DP2) took a significantly longer time than controls to initially fixate a face.

One might predict that DPs with impairments in face perception (i.e. those scoring poorly on the CFPT, see Table 1) would be the individuals to show significant abnormalities on the eye-movement measures. However, this pattern did not emerge: only one (DP 2) of the five DPs who were significantly impaired on an eye-tracking measure also presented with a deficit in face perception. This observation is supported by the absence of a correlation between any eye-tracking measure and performance on the CFPT in the DP group. Strikingly, however, DP performance on the CFMT did negatively correlate with the proportion of dwell time directed towards the entire face (r = -.633, p = .050), and specifically the inner features (r = -.678, p = .031). That is, DPs with more severe deficits in face recognition tend to spend less time viewing the inner features of the face.

Table 2. Performance of DPs and controls on each measure in Experiment 1. Performance of the DPs is expressed in the numbers of standard deviations away from the control mean.

	Controls		Individuals with DP									
	Mean	SD	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10
Eye-movement measures												
% dwell time faces	63.1	9.5	-2.7*	-0.8	-0.5	-0.7	-0.9	1.6	-1.5	-1.3	-0.1	-2.7*
% dwell time on bodies	21.7	4.9	2.3*	2.3*	0.6	1.4	0.7	-1.7	0.8	2.1	-0.7	3.1*
% dwell time on background	15.5	5.6	2.1	-0.8	0.3	0.0	1.0	-1.2	2.0	0.3	0.8	2.1
% dwell time on inner features	43.3	11.5	-1.6	-0.6	0.1	-0.3	-0.6	1.8	-1.2	-0.4	0.4	-1.5
% dwell time on outer features	19.8	7.8	-0.8	-0.1	-0.7	-0.4	-0.3	-0.7	-0.1	-1.1	-0.7	-1.1
% dwell time on eyes	13.1	8.6	-0.9	-0.7	-1.4	-1.4	-1.4	-1.2	-1.0	-0.9	-0.8	0.1
% dwell time on nose	12.7	6.3	-1.3	-1.5	1.1	-1.3	-1.2	1.2	-0.2	0.4	1.3	-0.9
% dwell time on mouth	9.2	9.1	0.7	2.2*	0.9	2.6*	1.5	1.9	-0.4	0.4	0.1	-0.9
First fixation to a face (msec)	973.1	161.4	1.8	2.6*	0.9	-0.6	0.8	-1.3	-0.9	0.7	-0.3	-0.5

*significantly different performance to control participants at the .05 level.

2.3. Summary of Experiment 1

Group-based analyses revealed no differences in the time that DPs and controls spent looking at the background of the images, but the DP group as a whole spent more time looking at bodies and less time looking at faces. While no group-based difference was found in the time they spent looking at the inner compared to the outer facial regions, a significant correlation indicated that DPs with more severe face recognition deficits spent less time viewing the inner facial features. At the group-based level, differences in more specific patterns of feature exploration were observed. Specifically, while the DP group spent a similar amount of time looking at the nose as controls, they spent less time fixating on the eyes and more time examining the mouth. However, no individual DP significantly differed in the time spent on the eyes compared to controls, and only two on the time spent on the mouth. The group-based findings reported here are remarkably similar to those previously observed in acquired prosopagnosia, suggesting similar underlying impairments in both forms of the condition.

3. EXPERIMENT 2

In order to examine whether DPs and SRs represent individuals at the extreme ends of the typical face recognition continuum, our second experiment adopted the same paradigm as used in Experiment 1. This time, participants were eight individuals who met the published criteria for super recognition and a new group of 20 controls. The latter group was necessary because the SRs significantly differed in age to the DP group. In addition to performing the same analyses for the SR group as described above for the DPs, the two control groups were also combined for a series of analyses that considered scanning strategies within typical perceivers. The combination of controls across both Experiments provided greater statistical power for these analyses, while permitting examination of the influence of participant age on the key findings. Given the difference in age between the DP and SR groups, this comparison was required in order to infer that any findings related to group membership (i.e. identification of each individual as a DP or a SR) as opposed to participant age.

3.1. Method

Identical protocols were used as described in Experiment 1, but different participants were tested. Following widespread media coverage about super recognition, eight individuals who believed they had extraordinary face recognition ability contacted our laboratory (see Table 3). The same strategies for identifying SR participants were adopted as in Chapter 2, and three of the same participants took part in this study, with four additional individuals. While these diagnostic criteria have been previously explained (see section 2 of Chapter 2), all SR participants reported extraordinary face recognition skills that had been present from an early age and achieved scores that were above 90/102 on the CFMT+ (Russell et al., 2009). Table 3 also reports CFPT data for the SR sample, although as discussed in Chapter 2, significantly enhanced performance on this test is not necessary for categorization as a SR. That is, in line with the predictions of dominant models of face-processing (e.g. Bruce & Young, 1986), superior face memory skills are not necessarily dependent on superior face perception skills. In addition, the large variability in control performance on the CFPT results in a large standard deviation, making significant differences on single-case analyses near impossible to achieve. Hence, in line with the procedure followed by Russell et al. (2009), we performed only a group-based analysis and found that our SR group also significantly outperformed the reported control norm on this test (Duchaine et al., 2007; $F(1, 27) = 15.54$, $p = .001$), $\eta\rho^2 = .575$. To provide persuasive evidence that these individuals are SRs, four were also tested on two alternative face recognition tasks and outperformed controls (GFMT and MFMT: SR2, SR3, SR7 and SR8: see Chapter 5).

A new group of 20 (10 female) control participants also participated in this study, and were matched to the SRs according to age (M = 24.7 years, SD = 5.7) and estimated IQ. All control participants reported typical face recognition skills, and this was confirmed via completion of the CFMT (standard form), where all participants performed within the "typical" range (M =58.0, SD = 8.0). These participants were all Bournemouth University students and staff members who participated in exchange for course credits or a small monetary payment. Ethical approval for this study was granted by Bournemouth University's Ethics Committee.

3.2. Results

SR Participants

As observed in Experiment 1, analysis of the regional distribution of fixations throughout the entire scenes revealed that all participants spent longer looking at faces (M = 67.70%, SE = 2.00) than either bodies (M = 18.80%, SE = 1.12) or the background of the images (M = 12.70%, SE = 1.20), $F(2,52)$, = 262.200, $p = .001$, $\eta\rho^2 = .910$. The difference in time spent on the bodies versus the background was also significant, $F(1,26) = 38.436$, $p = .001$, $\eta\rho^2 = .587$. A significant interaction between region and group was again observed, $F(2,52) = 3.900$, $p = .026$, $\eta\rho^2 = .130$, and follow-up analyses indicated that while there was no difference in the time spent studying the background or faces, controls spent more time than SRs looking at bodies $F(1,26) = 2.597$, $p = .119$, $\eta\rho^2 = .091$, $F(1,26) = 3.959$, $p = .057$, $\eta\rho^2 = .132$, and $F(1,26) = 4.946$, $p = .035$, $\eta\rho^2 = .160$, respectively (see Figure 3a).

However, it should be noted that there was a main effect of participant group, suggesting that SRs (M = 33.30%, SE =0.01) spent more time than controls (M = 32.90%, SE = 0.01) fixating on each image component, $F(1,26) = 6.390$, $p = .018$, $\eta\rho^2 = .197$.

Table 3. Demographical information, CFMT+ scores for the SR participants used in this study and SR and control norms described by Russell et al. (2012).

	Russell et al. (2012) (N = 6)		The current study								
	SRs (N = 6)	Controls (N = 26)	SRs (N = 8)	SR1	SR2	SR3	SR4	SR5	SR6	SR7	SR8
Age	40.7 *(9.9)*	42.2 *(14.1)*	24.5 *(4.2)*	20	29	21	20	33	19	27	27
Gender	-	-	M = 5	F	M	M	M	M	F	M	F
Hand	-	-	R = 8	R	R	R	R	R	R	R	R
CFMT+	95.0 *(1.9)*	75.2 *(11.6)*	95.7 *(1.5)*	96	97	100	95	100	96	101	94
CFPT (upright)	24.7 *(10.3)*	35.4 *(12.9)*	18.5 *(7.2)*	10	26	16	12	16	32	16	24

As in Experiment 1, the second set of analyses examined the proportion dwell time spent on the inner versus the outer regions of the face. A main effect of region indicated that all participants again spent longer looking at the inner (M = 55.70%, SE = 2.50) versus the outer (M = 12.00%, SE = 1.20) region, and this factor interacted with participant group: $F(1,26) = 172.364$, $p = .001$, $\eta\rho^2 = .869$ and $F(1,26) = 9.749$, $p = .004$, $\eta\rho^2 = .273$ respectively. Follow-up analyses indicated that SRs spent more time looking at the inner features and less time looking at the outer features of faces than controls: $F(1,26) = 8.456$, $p = .007$, $\eta\rho^2 = .246$ and $F(1,26) = 6.943$, $p = .014$, $\eta\rho^2 = .211$ respectively (see Figure 3b). The main effect of group approached significance, indicating that SRs spent more time attending to faces (regardless of region) (M = 35.90%, SE = 1.70) than controls (M = 31.80%, SE = 1.11), $F(1,26) = 3.959$, $p = .057$, $\eta\rho^2 = .132$.

Analysis of patterns of facial feature exploration also indicated differences between SRs and control participants. While participants in both groups spent longer looking at the nose (M = 24.20%, SE = 1.60) than either the eyes (M = 12.70%, SE = 1.80) or the mouth (M = 9.40%, SE = 1.40), $F(2,52)$, = 18.899, $p = .001$, $\eta\rho^2 = .421$, the “feature” factor also interacted with participant group $F(2,56) = 5.804$, $p = .005$, $\eta\rho^2 = .182$. Follow-up analyses indicated no differences in the proportion dwell time spent on the eyes and mouth, but SRs spent more time on the nose than controls: $F(1,26) = .557$, $p = .462$, $F(1,26) = .385$, $p = .540$ and $F(1,26) = 17.937$, $p = .001$, $\eta\rho^2 = .408$, respectively (see Figure 3c). There was also a main effect of group, indicating that SRs spend more time looking at the inner features of the face than controls, $F(1,26) = 9.153$, $p = .006$, $\eta\rho^2 = .260$.

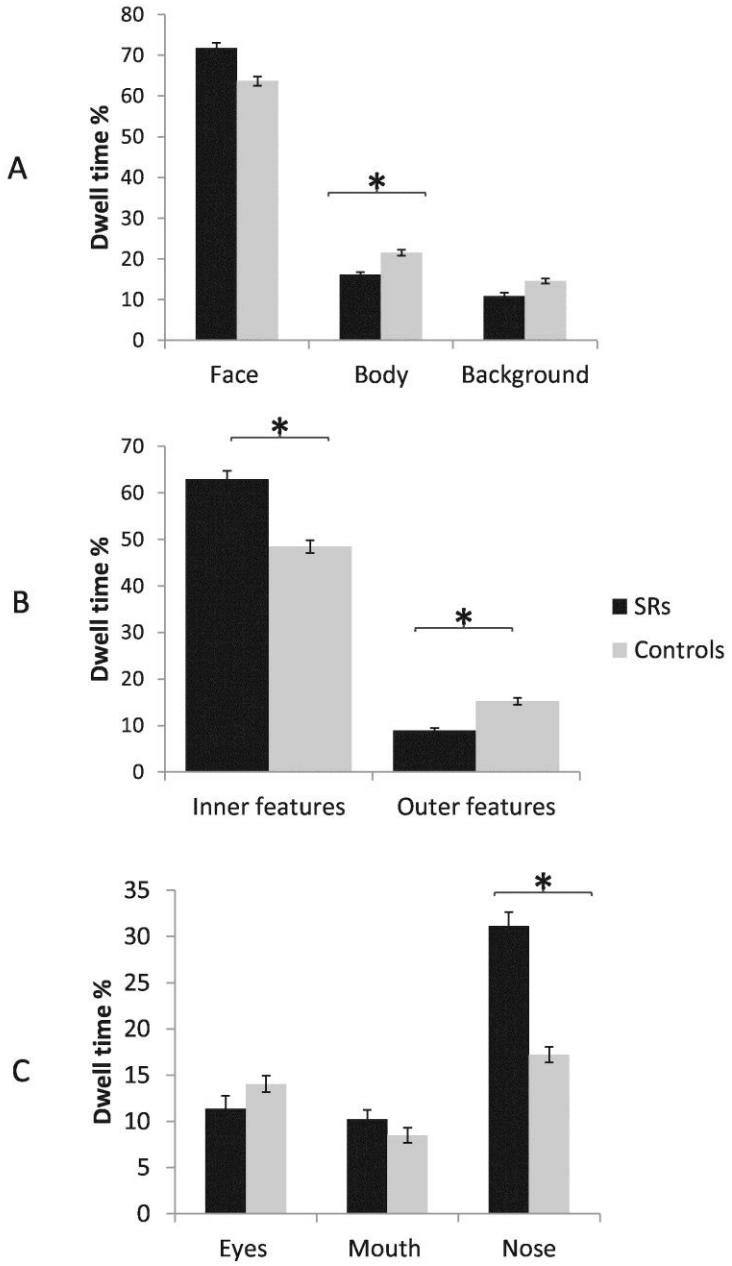

Figure 3. The percentage dwell time spent by SRs and controls on each region in Experiment 2. Error bars represent the standard error of the mean.

A final group analysis indicated that SRs were also faster (M = 794.17ms, SE = 31.68) than controls (M = 934.08ms, SE = 36.38) in first fixating upon a face, $F(1,26) = 5.199$, $p = .031$, $\eta\rho^2 = .167$.

Finally, case-by-case analyses were also performed on the data, given there is reason to suspect that super recognition may also be characterised by cognitive heterogeneity (see Table 4). As in Experiment 1, we performed modified *t*-tests for single-case comparisons for all the eye-tracking measures (Crawford et al., 2010). Although there was a group-based difference in the proportion dwell time spent studying bodies, this measure did not significantly differ in any of the case-by-case comparisons. Likewise, the significant group comparisons noted for the proportion dwell time spent on the inner and outer facial features only resulted in one significant single-case comparison (for SR7). However, four participants (SR1, SR2, SR4, and SR7) spent significantly longer than controls looking at the nose, and an additional participant performed above two standard deviations from the mean on this measure (SR6). Finally, although the SRs as a group elicited their first fixation to a face more rapidly than controls, this finding was not supported in any case-by-case comparisons.

It is more difficult to discern whether the eye-movement measures are indicative of any cognitive variation in the SR group. Indeed, all the SRs achieved high scores on the CFMT+ and the CFPT, and unsurprisingly the cluster of scores at near-ceiling levels and the small sample size prohibited any potential correlations from emerging.

Table 4. Performance of the SRs and controls on each measure in Experiment 2. Performance of the SRs is expressed in the numbers of standard deviations away from the control mean.

	Controls		Super-Recognizers							
	Mean	SD	SR1	SR2	SR3	SR4	SR5	SR6	SR7	SR8
Eye-movement measures										
% dwell time faces	63.7	10.3	0.6	0.7	-0.8	1.29	1.0	1.1	1.8	0.6
% dwell time on bodies	21.5	6.5	-1.2	-0.7	0.2	-1.01	-1.3	-0.9	-1.3	-0.5
% dwell time on background	14.6	5.8	-0.1	-0.6	1.1	-1.15	-0.8	-1.2	-1.7	-0.8
% dwell time on inner features	48.5	12.4	1.2	1.4	-0.5	1.74	0.9	1.4	2.3*	0.9
% dwell time on outer features	15.2	6.3	-1.3	-1.6	-0.4	-1.29	-0.0	-1.0	-1.6	-0.8
% dwell time on eyes	14.1	8.9	-0.8	-1.0	-0.2	-1.14	1.5	-0.9	-0.1	0.3
% dwell time on nose	17.2	7.6	2.6*	2.2*	0.1	3.40*	0.8	2.1	2.6*	0.9
% dwell time on mouth	8.5	7.2	0.6	1.3	-0.4	0.63	-0.9	0.2	0.9	-0.5
First fixation to a face (msec)	934.1	162.7	-1.0	-1.2	-1.0	-1.25	0.2	-0.6	-1.5	-0.5

*significantly different performance to control participants at the .05 level.

Control Participants

In order to examine whether the different patterns of performance observed in the DP versus the SR group could simply be attributed to age (the DP group were significantly older than the SR group), we performed a final set of analyses on the control data to examine whether age interacted with any of the eye-tracking measures or our measure of face recognition ability (i.e. performance on the CFMT). Data for the control groups were combined across Experiment 1 and Experiment 2, and no effect of age was observed on any of the eye-movement variables (all *p*s > .05). Unsurprisingly, performance on the CFMT was lower in the older (M = 49.15, SE = 1.92) compared to the younger (M = 57.95, SE = 1.92) participant group, $F(1,38) = 10.474$, $p = .003$, $\eta\rho^2 = .216$ (see Bowles et al., 2009). Hence, the differing pattern of performance in the DP (i.e. less time spent on the eyes and more on the mouth) versus the SR (i.e. more time spent on the nose) group cannot simply be attributed to participant age.

This pattern of findings raises a final question: which measure reflects the typical face-processing continuum? That is, does the face recognition ability of typical perceivers vary according to the time spent on the eyes and mouth as observed for DPs, or the nose as observed for SRs? To address this question we performed three correlations on the collapsed control data. While no significant correlation was observed between control CFMT performance and the proportion dwell time spent on the eyes ($r = .179$, $p = .268$), there was a marginal negative correlation with the time spent on the mouth ($r = -.309$, $p = .052$) and a stronger positive correlation with the time spent on the nose ($r = .408$, $p = .009$).

3.3. Summary of Experiment 2

This experiment investigated the eye-movement patterns of SRs and typical perceivers in the same eye-movement task as used in Experiment 1. At the group level, SRs spent less time examining bodies and more time examining the inner features of faces than controls. While case-by-case analyses mostly failed to reach significance on these measures, a more consistent pattern emerged for the proportion dwell time spent on the nose. In group and four individual analyses, the SRs spent a significantly longer time looking at the nose. Given this finding does not simply mirror those of Experiment 1 (where DPs were found to spend less time on the eyes and more time on the nose), one could argue that DPs and SRs do not merely represent individuals at the opposite ends of the typical face recognition spectrum. Analysis of the control data revealed a correlation between face recognition ability and the proportion dwell time spent on the nose, indicating that SRs may simply be those at the top end of the spectrum, whereas DPs may reside on a qualitatively different continuum.

Experiment 2 therefore lends support to the findings of Hsiao and Cottrell (2008) and Patterson and Eckstein (2012), indicating that the nose may represent an optimal viewing position in face recognition. However, as this is the first eye movement investigation examining super recognition, we sought to replicate the findings in a final experiment.

4. EXPERIMENT 3

Our last experiment provided an additional examination of eye-movements in SRs, using an alternative paradigm. This time we used single-face stimuli and asked participants to view each face and encode it for a later recognition test (that was never presented).

4.1. Method

Participants

Two of the SRs described in Experiment 2 (SR4 and SR7) agreed to return to the laboratory and took part in this study. Additionally, in accordance with the Russell et al. (2009, 2012) criteria, we recruited two new SRs (SR9 and SR10), a 35 year-old male (92/102 faces identified correctly in the CFMT+ and 20 errors made in the CFPT upright) and a 35 year-old female (97 faces identified correctly in the CFMT+ and 20 errors made in the CFPT upright). Both participants contacted our research group independently and reported instances of extraordinary face recognition since childhood. These individuals also participated in another investigation in our laboratory, where their performance on two further face-processing tests significantly exceeded that of controls (JN and TP, see Chapter 2). A new group of 20 (10 female) control participants also participated, and were matched to the SRs according to age (M = 24.9 years, SD = 4.6) and estimated IQ. All control participants reported typical face recognition skills, and this was again confirmed via completion of the CFMT, where all participants performed within the "typical" range (M = 56.05, SD = 6.75). These participants were all Bournemouth University students and staff members who participated in exchange for course credits or a small monetary payment. Ethical approval for this study was granted by Bournemouth University's Ethics Committee.

Materials

Colour photographs of 24 (12 female) white Caucasian adults were taken from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). In all photographs the person was looking directly at the camera (direct gaze) and had a neutral facial expression. Faces were cropped to remove excess hair, but were not cropped around the hairline. The faces were presented against a white background and measured approximately 10 x 9cm, so that each face subtended 11.42 x 10.28 degrees of visual angle when viewed from a distance of approximately 50 cm. Gaze behaviour was recorded using the same eye-tracker has described in Experiment 1.

Procedure

Participants were seated approximately 50 cm from the screen, and placed their head within the chin rest. The experiment was preceded by a nine-point calibration procedure. Each trial began with the presentation of a central fixation cross, and each face was presented centrally on a single occasion and in a randomized order. Each face was displayed for five seconds, and participants were instructed to memorize the faces for a later recognition test.

Eye Movement Parameters and Statistical Analyses

Scanning behaviour was examined for the entire 5-second period. To investigate fixations to specific regions, a freehand tool was used to draw three AOIs onto each facial image, covering the eyes, nose and mouth. The proportion dwell time elicited to each AOI was calculated for each participant. Due to the very small size of the experimental group, only case-by-case (modified *t*-tests for single-case comparisons: Crawford et al., 2010) statistical procedures were performed on all measures.

4.2. Results

Case-by-case analyses (see Table 5) of SRs and control participants revealed that the difference in scanning strategy between these groups is related to the nose region, where all four SRs spent a significantly longer time examining this area compared to control participants. SR7 spent also more time scanning the mouth region of the studied faces than controls.

Table 5: Performance of the SRs and controls on the eye-tracking task. Performance of the SRs is expressed in the numbers of standard deviations away from the control mean.

	Controls		Super-Recognizers			
	Mean	SD	SR4	SR7	SR9	SR10
Eye-movement measures						
% dwell time on eyes	45.6	17.7	-1.0	-1.9	-0.4	-1.0
% dwell time on nose	21.7	9.2	4.2**	2.7*	2.2*	4.8**
% dwell time on mouth	9.8	5.5	-1.4	2.4*	-1.7	-1.8
% dwell time on inner features	77.1	12.5	1.1	0.4	0.3	1.4

*significantly different performance to control participants at the .05 level.

**significant different performance to control participants at the .001 level

4.3. Summary

The results of Experiment 3 provide further support for the findings observed for the SR participants in Experiment 2, using an alternative paradigm. Specifically, when required to memorise a set of faces, all four SRs spent significantly more time than control participants viewing the nose. Surprisingly, SR7 also spent significantly more time than controls looking at the mouth region, although a similar trend did not emerge in any of the other SR participants, nor for this individual in Experiment 2.

5. DISCUSSION

This investigation monitored the eye-movements of DP, SR and control participants while they viewed images of people engaged in natural social scenes. Findings in the DP group suggest that, in some cases, the condition may be underpinned by reduced attention to faces. Indeed, while the DPs as a group did not take a longer time to initially fixate upon a face than controls, they did spend less overall dwell time examining faces and more time viewing bodies. Further, in contrast to previous work, the DP group did not spend less time on the inner facial features than controls, but this measure did correlate with their score on the CFMT, indicating that individuals with more severe prosopagnosia spent less time examining the inner region of the face. As observed in previous work examining acquired prosopagnosia, the DP group spent less time viewing the eyes and more time viewing the mouth than controls. In contrast, across two experiments, SRs spent more time viewing the nose, suggesting these individuals are not merely the “opposite” of DP. Instead, analysis of control data indicated that the face recognition ability of typical perceivers is also associated with the time spent examining the nose, and therefore that SRs represent individuals at the top end of the typical face-processing system, whereas DPs may be a qualitatively different group.

First, the DP findings will be addressed. One aim of the investigation was to examine whether DP may result from a lack of attention to faces. While the DPs as a group did not take a longer time to initially fixate a face, a disparate pattern of findings emerged in single-case analyses. Indeed, one DP took a significantly longer time than controls to initially fixate on a face (DP2), and another (DP1) performed at 1.77 standard deviations above the control mean. Although three other DPs were slower than controls but within one standard deviation of the control mean, the mean scores of the remaining five DPs were quicker than controls. These findings converge with previous work suggesting that DP is a heterogeneous condition, and that impaired face detection mechanisms may underpin the face recognition difficulties in only a subset of individuals (Garrido, Duchaine, & Nakayama, 2008; Dalrymple & Duchaine, 2015). Indeed, in their recent paper with a group of seven children with DP, Dalrymple and Duchaine (2015) reported that four participants had impaired face detection, but in three children the mechanism was spared and performance remained on par with that of controls. This finding converges with some earlier work with adults (Garrido et al., 2008) where four DPs (out of a group of 14) also had intact face detection mechanisms. The authors attributed this heterogeneity to occurrence of ectopias, regions of cortical disorganisations produced by impaired neural migration (Dalrymple & Duchaine, 2008). Specifically, they suggested that ectopias affecting brain areas responsible for face detection may result in atypical attention to faces and failure to develop a functional face recognition system, leading to DP. On the other hand, ectopias at higher levels of visual system (i.e., occipital and temporal areas of the brain) could result in DPs with impaired face recognition and perception, notwithstanding intact face detection. Finally, a pervasive ectopia of the entire face processing system may lead to face perception and recognition problems that are concomitant but not necessarily resulting from face detection abnormalities. It is thus possible that in the study presented here, DP1 and DP2, but not the other eight DP participants, had partial or widespread ectopias resulting in slower orienting towards faces within social scenes. To assess this heterogeneity fully, future work should endeavour to combine standard behavioural tasks of face detection with more ecological paradigms, such as the social scenes task described in this study.

Further insight into the hypothesis that reduced attention to faces may underpin some cases of DP comes from the measures examining the proportion of dwell time to the background of the images, bodies and faces. It is striking that there was no difference between DPs and controls in the proportion dwell time spent on image background, but group-based analyses indicated that DPs spent more time examining bodies and less time examining faces. Single-case analyses found supporting evidence for this pattern in three DPs (DP1, DP2 and DP10), and similar non-significant trends were noted in all other DPs with the exception of DP6 and DP9. While these findings may indicate that reduced attention to faces underpins DP in some individuals, it is also possible that this measure reflects a social consequence of the disorder. That is, because faces provide little information to people with DP, they rely on alternative sources of information (e.g. bodies or movement, Bennetts et al., in press) to make identity judgments.

The proportion dwell time spent on the inner and outer features has been used in many previous investigations to indicate reduced attention to the core facial features in acquired prosopagnosia (Le et al., 2003), and in one investigation using developmental cases (Schwarzer et al., 2007). The findings reported here are therefore in contrast to previous reports, given we did not find any differences on this measure between the DP and control groups as a whole, or in any single-case comparisons. This finding may result from the different, more ecologically valid paradigm that was used here compared to previous work, where single faces were typically presented against plain backgrounds. However, it is striking that two DPs tended to spend a longer time viewing the inner features: DPs 6 and 9. Without wishing to place too much emphasis on non-significant results, it is nevertheless of interest that these were the same individuals who tended to initially fixate on faces more rapidly than controls. Further, the finding that CFMT scores in the DP group correlated with the time spent on the inner features may account for the pattern of findings reported here: it is possible that DPs with milder prosopagnosia (i.e. those with higher CFMT scores) may represent those at the bottom end of normal rather than membership of a qualitatively separate group. Future work might address this possibility, and the identification of different phenotypes of the condition is likely to have important implications for the development of remediation techniques (see Bate & Bennetts, 2014).

It is also of interest that group analyses revealed that DPs spent less time fixating the eyes and more time fixating the mouth than control participants. Although case-by-case analyses did not reach significance for any individual DP with regard to the "eyes" measure, the trend was in the same direction in all but one participant (DP10). The reduced time spent on the mouth significantly differed for two individual DPs (DP2 and DP4), and the trend was present for all other individuals with the exception of DP10. It is therefore possible that DP10's prosopagnosia has different underpinnings to the rest of the group, but the few significant case-by-case analyses suggests that these patterns of feature exploration may not serve as reliable biobehavioural indicators of the condition.

Notably though, the group-based patterns of feature exploration reported here converge with previous reports of acquired prosopagnosia (Le et al., 2003; Stephan & Caine, 2009), which have typically been attributed to a reduced ability to process faces in a "holistic" or "configural" manner (Stephan & Caine, 2009). Indeed, it is generally accepted that configural processing requires analysis of the particular presentation of the inner features of the face and the spatial relations within them (e.g. Maurer, Le Grand, & Mondloch, 2002), and it is possible that the increased focus on the mouth region may distract attention from the more informative eye region and from employing optimal configural processing mechanisms.

However, this interpretation is challenged by previous work that has identified the nose as an optimal viewing position in face recognition (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012), and provides evidence against a body of other work that suggests the eyes are pivotal in face recognition (Schyns et al., 2002; Sekiguchi, 2011). Both Hsiao and Cottrell (2008) and Peterson and Eckstein (2012) suggested that the nose may be the optimal viewing position because it is the best location for holistic and configural processing of the entire face. Interestingly, the findings in our SR and control sample indicate that the proportion dwell time spent on the nose has a positive association with face recognition ability, although we cannot comment on whether the proportion dwell time spent on the nose is representative of configural or holistic processing skills. A consistent finding in the word reading literature is that there is an optimal viewing position (just to the left of a word's centre) when intentionally processing words (O'Regan & Jacobs, 1992) and when word reading proceeds automatically (Smilek, Solman, Murawski, & Carriere, 2009; Parris, Sharma & Weeks, 2007). In this literature the optimal viewing position is accounted for lexically (Stevens & Grainger, 2003), but has been shown to influence the spatial distribution of attention across non-lexical stimuli (Ducrot & Pynte, 2002). Since better distribution of spatial attention across a face does not necessarily imply better configural processing, it is possible that better face recognition in SRs results from a more efficient spread of spatial attention across faces.

Notably, it seems that this finding is a relatively reliable indicator of super recognition, given it emerged in four of the eight SRs in Experiment 2 (with one other SR exceeding control performance by more than two standard deviations) and all four of the SRs in Experiment 3 (note that one of the latter participants was the same as one who did not significantly differ from controls in Experiment 2). Nevertheless, the finding that it did not emerge in all SR participants leaves open the possibility that superior face recognition may also be characterised by cognitive heterogeneity, and it may have different underpinnings in different individuals. This is perhaps supported by the unexpected finding that one SR (SR7) spent more time examining the mouth than controls in Experiment 3, although the same effect did not emerge for this participant in Experiment 2.

However, it is relevant that our combined analyses of control performance across Experiments 1 and 2 indicates that SRs may simply be those at the top end of the typical face-processing system. Indeed, the proportion dwell time that controls spent on the nose correlated with their face recognition skills, whereas no correlation was noted with the time spent examining the eyes, and only a mild correlation emerged for the time spent on the mouth. Alternatively, given the trends towards a qualitatively different pattern of processing in DPs (where no participant differed from controls on the time spent on the nose, but various effects emerged for the eyes and mouth), our data supports the hypothesis that the condition is comparable to acquired prosopagnosia, and most of these individuals do not simply represent the bottom end of the typical face-processing spectrum.

In sum, this study presents evidence that (a) some cases of DP may be underpinned by reduced attention to faces, and (b) that at least some individuals with the condition represent a qualitatively different group to typical perceivers, rather than simply being the “bottom end” of normal. Conversely, individuals who meet the criteria for super recognition appear to be those at the “top end” of normal, and the work presented here suggests that the nose (as opposed to the eyes) appears to be a critical region involved in successful face recognition. Future work should endeavour to further partition DP, and to establish whether the nose region is also associated with heightened configural or holistic processing skills.

Chapter 4: The identification of super recognition in young British adults

1. INTRODUCTION

In Chapters 2 and 3, identification of SR participants has largely followed the precedent set by Russell et al. (2009, 2012). These authors introduced the CFMT+ - an extended version of the CFMT that is typically used to diagnose prosopagnosia (e.g. Duchaine, Germine, & Nakayama, 2007; Duchaine & Nakayama, 2006). In addition, Russell and colleagues also used the CFPT to demonstrate facilitations in the perception of facial identity, and a "Before they were famous" test, in which participants view faces of celebrities that were taken some time before they became well known in the public domain. These tests were administered to four individuals who self-referred to the authors claiming extraordinary face recognition skills, sharing anecdotal instances to this effect. This diagnostic approach mirrors the technique that is typically used to diagnose prosopagnosia, where participants are required to self-report long-lasting face recognition impairments and instances of inability to recognise faces of their relatives, colleagues and famous people; score within the impaired range on the CFMT (≤ 42/72; Duchaine & Nakayama, 2006) and sometimes on the CFPT (≤ 60 mistakes on the upright trials, Duchaine et al., 2007), as well as show an inability to recognise famous faces on culturally-specific sets of celebrity faces (Bate et al., 2014).

However, it is questionable whether tests born out of the prosopagnosia literature are suitably sensitive to detect superior performance at the top end of the face recognition spectrum, even with the additional 30 "difficult" trials that are used in the CFMT+. The statistical approach used by Russell and colleagues to identify SRs is also questionable, given they employed group-based analyses to compare the performance of four SRs to those of a small number of controls (N = 25). Although the authors were attempting to apply standardised neuropsychological tests of face recognition ability to identify those at the top rather than the bottom end of the spectrum, unfortunately they failed to apply the standard statistical criteria that are typically used in neuropsychological diagnosis. That is, the cut-off of two standard deviations from the control mean is typically calculated to detect impaired or, in this case, superior performance on a case-by-case rather than a group basis (e.g. Schinka et al., 2010). When the sample size of the control group is ample, this technique is deemed appropriate. However, when the control group is small, many researchers use modified *t*-tests for single-case comparisons (Crawford & Howell, 1998) to provide a more conservative estimate of significantly different performance. This individual rather than group approach to diagnosis not only ensures that each individual participant meets the criteria for the condition in question, but is also of key importance when certain conditions are suspected to have cognitive heterogeneity, and the precise presentation of each individual has key theoretical implications. Much published work indicates this is the case in prosopagnosia, and the findings reported in Chapters 2 and 3 raise the possibility that this is also the case in super recognition.

As stated above, application of the standard neuropsychological approach of calculating two standard deviations from the control mean as a cut-off requires a substantial amount of control data to calculate these norms. In addition, there are likely to be different cut-offs on face recognition tasks according to standard demographic variables such as age, gender and even ethnicity (Bowles et al., 2009). This issue particularly applies to the CFPT given it results in much more varied performance in controls, resulting in large standard deviations. In fact, when we applied the published norms for upright CFPT performance (Russell, Chatterje, & Nakayama, 2012) to the scores of our SRs in Chapters 2 and 3, it was near-impossible for these individuals to achieve significantly superior scores than the control group. This may be because the published norms are not suited to the demographics of the SR participants identified in this thesis, but may also represent a more general issue with this test rendering it unsuitable for the detection of super recognition. Nonetheless, the CFPT may be useful for deriving an inversion index, a measure previously used by Russell and colleagues (2009) in their pioneering work on super recognition. Indeed, the effect of inversion is one of the best documented in face recognition research and is widely thought of as the hallmark of configural (or holistic) processing (e.g. Maurer, LeGrand, & Mondloch, 2002; Tanaka & Farah, 1998).

Although the standard form of the CFMT has been shown to have very good psychometric properties making it suitable for the diagnosis of both acquired and developmental prosopagnosia (Bowles et al., 2009; Duchaine & Nakayama, 2006), barely any attention has been directed towards the suitability of the CFMT+ as a diagnostic tool in super recognition. The original paper identified four cases of SRs (Russell et al., 2009), but did not provide the control mean for either of its studies making comparison and direct replication of procedure impossible. In the Russell et al. (2012) study the control mean and standard deviation were provided for the control group (M = 75.2, SD = 11.6) and the SRs (M = 95.0, SD = 1.9), but the individual scores of SRs were not reported, contrary to similar publications in the literature concerning developmental prosopagnosia which usually provide individual face recognition scores, so the reader can easily establish the number of standard deviations below the mean that an individual's score is placed at.

In SR research, diagnosis to date has been based on a loosely defined cut-off point that is approximately two standard deviation above a specific control group's mean. Using different control groups poses a risk in itself due to individual differences in performance between individuals. For instance, the control sample in Study 1 of Russell et al.'s paper appears to contain at least one and potentially three (it is impossible to determine the exact score from the scatterplot) individuals who meet criteria for prosopagnosia. The prevalence of prosopagnosia has been previously reported to be 2-2.9% in an Australian sample (Bowles et al., 2009) and 2.47% in a German young adult sample (Kennerknecht et al., 2006). In the group of 25 control subjects in Russell et al.'s study, a prevalence rate of 3% would corresponds to 0.75 participants. Given that the group potentially contained three individuals with DP, it may be that the control sample was atypical. Although it is possible that these individuals were not impaired in their face recognition performance, it is also plausible that they were affected by other neurodevelopmental disorders that are known to affect face recognition skills, such as autism spectrum disorder (see Weigelt, Koldewyn, & Kanwisher, 2012 for a review).

The initial Russell et al. (2009) publication also provided a test of famous faces where participants are asked to recognise famous people from childhood photographs. The authors reported a strong positive correlation between the "Before They Were Famous" (BTWF) test and the CFMT and CFMT+; $r = .70$, $p < .001$ and $r = .71$, $p < .001$ respectively. These correlations, however, suffer from a sampling error that makes their meaningful interpretation difficult, if not impossible. Namely, within 29 subjects, four SRs make up 13.8% of the sample. This proportion is not representative of the general population where the prevalence of SRs, defined by performance of two standard deviation above the control mean, would not exceed 2.5%. Ultimately the top end of the distribution in the original report on SRs is artificially inflated and potentially confounds the correlational analysis presented in the paper. The conclusion that the BTWF test correlated with the CFMT should therefore be seen as tentative, at least until appropriate control data is published.

Another potential problem with the BTWF test is that it does not control for the amount of exposure to the famous faces. As such, while most people may know the actress Scarlett Johansson (Russell et al., 2009; Figure 1), the number of times they have seen the face would naturally vary with the number of occasions at which her face was seen by a person. Pertinently, recent evidence (Andrews, Jenkins, Cursiter, & Burton, 2015;

Dowsett, Sandford, & Burton, 2015) suggests that people form more stable representations of faces after a multitude of varied exposures. It is plausible that the SRs are more interested in (and thus more highly familiar with) famous people, making them better equipped to recognise them from childhood photographs than controls.

Finally, it is of note that SRs are typically identified via their own self-referral to a laboratory for testing. This tends to occur following media coverage of the phenomenon, where people suspect they may also have extraordinary face recognition skills. The issue of self-report has been contentious in the prosopagnosia literature, where most reports indicate that we have little insight into our face recognition skills (DeHaan, 1999; Palermo et al., In press; but see Kennerknecht et al., 2006; Shah, Gaule, Sowden, Bird, & Cook, 2015). Pertinently, in their recent study, Palermo and colleagues compared the performance of 300 participants on a variety of behavioural measures against their self-reported face recognition ability. The authors argued that while those aware of their profound deficits in face recognition perform poorly on behavioural (objective) tests measuring this ability, typical perceivers have only modest insight into their face recognition skills. These findings are in contrast to the recent study by Shah et al. (2015) who reported strong correlations between their new questionnaire of face recognition ability (PI20) and the CFMT (Duchaine & Nakayama, 2006). Nonetheless, the latter report is a likely result of a statistical omission on the authors' part. Specifically, the correlations presented in the Shah et al. (Study 4, 2015) study were performed using a sample of participants including typical perceivers and developmental prosopagnosics (DP). The DP participants constituted 17.2% of the overall sample – a prevalence that is highly unlikely to occur in any typical population (e.g. Bowles et al., 2009; Kennerknecht et al., 2008). If those with DP do have a greater insight into their abilities than typical perceivers, the inclusion of this special population could have artificially inflated the strength of the reported correlations. It is likely that the same findings apply to the self-report of superior face recognition skills, although no work to date has explored this issue.

The current study adopted the current diagnostic technique that is used to identify SRs, and applied it to a large sample of 254 participants. We used the CFMT+, CFPT and self-report measures to assess the face recognition skills of the participants. The BTWF test was not used given its clear confounds that would be difficult to control within such a large-scale study. This approach allowed us to initially assess the psychometric properties of the CFMT+, and to establish more reliable norms that match the approximate ages and nationality of the individuals tested to date. Given that application of the standard neuropsychological cut-off (i.e. two standard deviations from the control mean) results in the detection of the top or bottom two per cent of the population when a dataset is normally distributed, this approach also allowed us to assess the suitability of the tests for detecting potential SRs. This large dataset also permitted examination of the utility of self-report measures in detecting super recognition.

2. METHOD

2.1. Participants

254 participants (146 females and 108 males) with a mean age of 21.4 years (*SD* = 3.5, range 18-35) were recruited for this study, amongst students and visitors at Bournemouth University. The subjects were enrolled in several ways: through the online sign-up system for Psychology undergraduates, through advertisement at Bournemouth University and social media sites, and via opportunity sampling by the researcher. All volunteers were white Caucasians and were born in the United Kingdom (UK). All have spent the majority of their lives within the UK. No participant reported any known history of brain injury or

neurodevelopmental disorder that is likely to affect their face recognition skills (e.g. ASD, Wigelt, Koldewyn, & Kanwisher, 2012). Most importantly, participants were not selected on the basis of their face recognition ability, but were simply recruited on a voluntary basis from the local population in Bournemouth. Overall, the education level in this sample was high due to the main site of recruitment (a Higher Education facility), reflected by the average number of years spent in full time education M = 15.06, *SD* = 2.26, range 11-26). Ethical approval for this study was granted by the Bournemouth University Ethics Committee.

2.2. Materials

Screening Questionnaire. An initial questionnaire (see Appendix A) enquired about general demographic information (e.g. age, gender, years of education) and any history of developmental disorders (e.g. Autism Spectrum Disorder or Williams syndrome), brain injury and periods of visual deprivation. Participants were also requested to provide self-ratings of their general face recognition ability, including estimates of instances where they have failed to recognise familiar faces, occasions where they have recognising faces only seen briefly or a long time ago, and their ability to follow characters on TV shows (see Appendix A for the scales used to estimate participants in this study).

Cambridge Face Memory Test Long Form (CFMT+, Russell et al., 2009). This is an adapted version of the CFMT (Duchaine & Nakayama, 2006) with an added section of 30 “hard” trials containing heavily degraded images (see Figure 1 Chapter 2). The original CFMT familiarises people with six male faces. In the first section, subjects are presented with each face from three viewpoints and recognition of the same images is tested in a three-alternative forced-choice task. The second section is more challenging and the test images are novel and differ in pose and lighting conditions from the photographs used in the first part of the test. The third section also uses novel images, but with added visual noise which removes fine-grained information from tested. While the CFMT (Duchaine & Nakayama, 2006) is commonly used to identify prosopagnosia and assess face recognition in typical population (e.g. Bowles et al. 2009), it has a ceiling effect that makes it difficult to distinguish individuals with extraordinary face recognition memory. To address this issue, the fourth, additional section in the CFMT+ includes 30 trials of novel images varying in pose, emotional expression and the amount information available (faces are fully cropped from the external features or presented with visible hairstyle and exposed neck). All photographs are heavily degraded using visual noise and the distractor identities recur more frequently than in the first three phases to minimise the difference in familiarity between the studied and distractor faces. These changes are thought to create a level of difficulty that is challenging enough to discriminate between people with typical but high face recognition memory and SRs.

Cambridge Face Perception Test. The Cambridge Face Perception Test (CFPT, Duchaine, Germine & Nakayama, 2007) requires participants to order six male faces in likeness to one target face. Participants complete eight upright and eight inverted trials with a time allowance of one minute per set . In each trial participants are presented with a ¾ view image of the target face. The six test faces have been morphed to contain 88%, 76%, 64%, 52%, 40%, and 28% of the target face. Responses for each trial are calculated based on the sum of deviations from the correct position of each face. For instance, if a face is sorted one position away from where it should be, this constitutes one error. If a face is four positions away, that is four errors. Errors for each trial are then summed to total upright and inverted deviations separately. Chance performance on the upright condition is represented by 93.3 errors.

2.3. Procedure

All participants signed an informed consent prior to the study. They were then invited to sit in testing cubicles and given the screening questionnaire. Subsequently, the CFMT+ and CFPT were administered. The researcher stayed in the room during instructions and practice trials to ensure that participants are comfortable with the tasks and able to ask any questions. Following the testing session, all participants were handed a debrief sheet, or debriefed verbally. The study was granted ethical approval by the departmental Ethics Committee.

3. RESULTS

3.1. CFMT+

The Kolmogorov-Smirnov test was used to test for normality on the CFMT+. The mean correct responses for the male group, $D(108) = 0.08$, $p = .092$, and the mean correct responses for the female group, $D(146) = 0.07$, $p = .200$, were both non-significant, indicating that the data was normally distributed in both gender groups. Analysed together, the distribution of all CFMT+ scores was marginally abnormal, $D(254) = 0.06$, $p = .049$. This small departure from normality likely occurred because the significantly higher mean for female participants (see below) has created a trend towards a bimodal distribution when data is collapsed across genders. Figure 1 illustrates the distribution of CFMT+ scores in males, females, and collapsed for all participants.

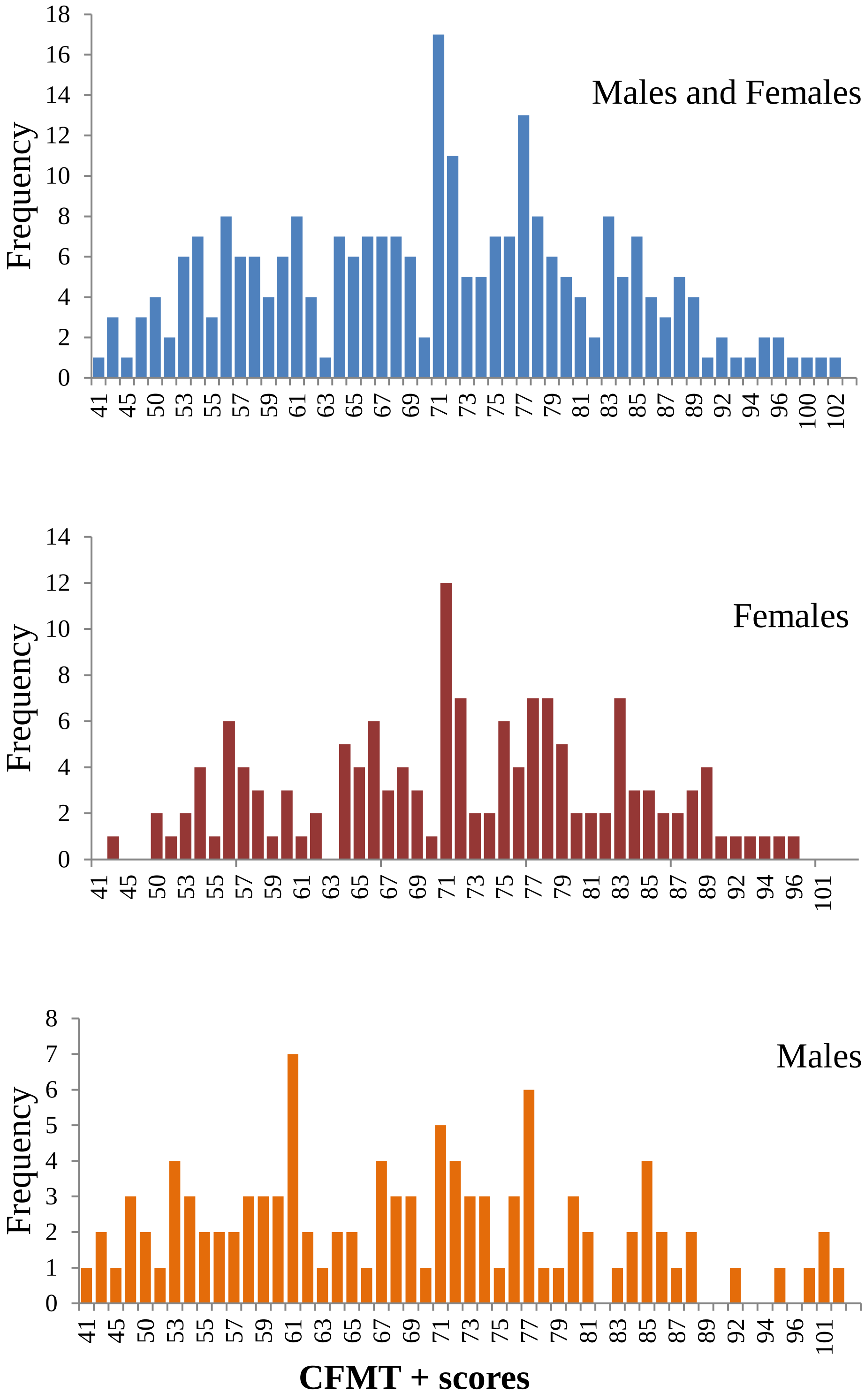

Figure 1. Distribution of CFMT+ scores.

In order to explore potential gender differences in performance, scores on the CFMT+ were further subdivided to represent the number of correctly recognised faces in four sections. It is possible that any differences that may arise between genders, do so at the level where stable representations need to be formed (i.e. after the first familiarisation phase using image, rather than face recognition). This gave scores for the "same images" (section 1) out of 18 (chance performance = 6), for the "novel images" (section 2) out of 30 (chance performance = 10), for the "novel images with .noise" (section 3) out of 24 (chance performance = 8), and for the "difficult images" (section 4) out of 30 (chance performance = 10). For the purpose of analysis, the above sections are labelled as CFMT1 to CFM4. The whole test comprising of 102 trials is referred to as "CFMT+" and the short version of the task including the first three sections (72 trials) is referred to as "$CFMT_{72}$".

First a mixed Analysis of Variance (ANOVA) was conducted on the proportion of correct responses for the four sections of the CFMT+ with test block number as the within–participant factor (CFMT1/CFMT2/CFMT3/CFMT4) and gender as the between-participant factor (females/males). The analysis revealed a main effect of the CFMT test block, $F(3, 756) = 1010.63$, $p < .001$, $\eta_p^2 = .800$; the main effect of gender, $F(1, 253) = 5.41$, $p = .021$, $\eta_p^2 = .021$, and a significant interaction between these two factors $F(3, 756) = 8.67$, $p < .001$, $\eta_p^2 = .033$. Between-group post-hoc analyses revealed that while there were no gender differences in performance on CFMT1 ($p = .136$) and CFMT4 ($p = .569$), but females outperformed males on CFMT2 ($p = .001$) and CFMT3 ($p = .012$).

Follow-up analyses of the within-subjects main effect (Bonferroni corrected) revealed a gradual decline in performance from CFMT1 to CFMT4 for all participants, all *p*s < .001.

Table 1. Gender effects on CFMT+ and CFPT scores for all participants.

	Gender of participants								
	Males			Females			Total		
	N	M	SD	N	M	SD	N	M	SD
CFMT1 (out of 18)	108	17.61	1.01	146	17.76	0.56	254	17.70	0.79
CFMT2 (out of 30)	108	21.76	5.69	146	23.98	4.41	254	23.04	5.11
CFMT3 (out of 24)	108	15.56	4.83	146	17.01	4.35	254	16.39	4.61
CFMT4 (out of 30)	108	13.78	4.60	146	13.47	3.90	254	13.60	4.21
CFMT+ (out of 102)	108	68.71	13.36	146	72.19	11.31	254	70.72	12.32
CFPT Upright	100	40.72	15.65	143	35.17	13.29	251	37.56	14.59
CFPT Inverted	108	69.39	13.50	143	67.07	13.16	251	68.07	13.33
CFPT Inversion effect	108	0.93	0.74	143	1.11	0.72	251	1.03	0.74

Note: CFMT = Cambridge Face Memory Test, CFPT = Cambridge Face Perception Test.

Proposed cut-off points for SRs in young British adults based on the data presented in this study are presented in Table 2.

Table 2. Proposed cut-off points for SR in young British adults.

Neuropsychological Test	Males		Females		Total	
	C*	N**	C	N	C	N
CFMT+ (SR cut off)	95.43	4	94.81	2	95.36	6
CFPT Upright (SR cut-off)	9.42	0	8.59	0	8.38	0

*C- cut-off point on the neuropsychological assessment
**N-number of people in the sample to meet these criteria

3.2. CFPT

A Kolmogorov-Smirnov test assessed normality on the main dependent variable: CFPT error rates. The mean error rates for the upright trials in the male group, $D(108) = 0.13$, $p < .001$, and the mean erroneous responses for the female group, $D(143) = 0.13$, $p < .001$, were both significant, indicating that the data was not normally distributed in both gender groups for the upright trials. The mean error rates for the inverted trials in the male group, $D(108) = 0.07$, $p < .200$, and the mean error number for the female group, $D(143) = 0.07$, $p = .085$, were both non-significant, indicating that the data was normally distributed in both gender groups for the inverted trials. Analysis of collapsed male and female data for the CFPT revealed that in the upright and inverted trials, data was not normally distributed, $D(251) = 0.12$, $p < .001$ and $D(251) = 0.06$, $p = .015$, respectively.

A mixed ANOVA was conducted on the error scores, with trial type as the within–participant factor (upright/inverted) and gender as the between-participant factor (females/males). The analysis revealed a main effect of the trial type, in that participants were overall better at sorting faces in upright than inverted trials, $F(1, 249) = 1080.50$, $p < .001$, $\eta_p^2 = .813$; and a significant difference between groups where females were better than males at sorting faces, regardless of orientation, $F(1, 249) = 6.82$, $p = .010$, $\eta_p^2 = .027$. The interaction between these two factors was non-significant, $F(1, 249) = 3.08$, $p = .080$, $\eta_p^2 = .012$ (see Table 1 for mean and SD performance of the two groups).

A final analysis was conducted on the CFPT inversion index for males and females. Similarly to the method adopted in Chapter 2, each participant's score for inverted trials was subtracted from their score for upright trials, then divided by their score for upright trials to create an inversion index ([upright-inverted]/[upright]; Russell et al., 2009). An independent samples t-test revealed significant differences between the two groups, $t(249) = 2.02$, $p = 0.44$, $d = 0.26$ females emerging as a group with a larger inversion effect (see Table 1).

3.3. Do young adults have insight into their face recognition ability?

Participants were also requested to provide self-ratings of their general face recognition ability, including estimates of instances where they have failed to recognise familiar faces, occasions where they have recognised faces only seen briefly or a long time ago, and their ability to follow characters on TV shows. The responses to these questions were correlated with the CFMT+ scores of all participants (Table 3). These analyses revealed weak to moderate correlations between all variables. Of particular interest is the weak relationship between the CFMT+ scores and self-perceived face recognition ability, no relationship between the CFMT+ and the recognition of faces only seen briefly, and the highly significant relationship between the recognition of faces only seen briefly and self-reported face recognition ability. Additionally, we carried out an independent samples *t-test* to examine any gender differences that may arise in the self-reported face recognition ability. However, there were no group differences in participants' own estimation of their face recognition skills, $t(252) = -.211$, $p = .833$, $d = 0.03$.

Table 3. The correlations between single items questions assessing self-perceived face recognition ability and the accuracy score for CFMT+ ($N = 251$).

Variables	1	2	3	4	5	6
1. CFMT+	–					
2. CFPT upright	-.552**	–				
3. Self-reported face recognition ability	.302**	-.257**	–			
4. Failure to recognise faces	-.190**	.239**	-.379**	–		
5. Following characters in a movie	.161*	-.135*	.420**	-.281**	–	
6. Recognition of faces seen briefly.	.041	-.055	.287**	-.070	.032	–

* Correlation significant at the .05 level
** Correlation significant at the .001 level

4. GENERAL DISCUSSION

This investigation aimed to examine the suitability of the current diagnostic criteria for the detection of super recognition. Using the two main tests that are currently employed for screening (the CFMT+ and the CFPT), together with self-report measures, the psychometric properties of the tests were assessed (for the first time in the CFMT+), and relevant cut-offs were established using a much larger sample than in previous work. The importance of each of these issues is discussed in turn below.

4.1. CFMT+

The CFMT+ emerged as a normally distributed measure without a ceiling effect - appropriate for classification of super recognition in young British adults. The previously published mean for this test was 75.2/102 (*SD* = 11.6) in a sample of 26 participants (gender unknown, Russell et al., 2012). The normative data for the sample of participants tested in this study is substantially lower by 4.48 points (see Table 1). This discrepancy could have occurred for a number of reasons. Firstly, the CFMT+ is comprised largely of a set of face images collected around Boston area of the United States. Bowles and colleagues (2009) reported that the normative data for the short version of the CFMT in an Australian sample is much below the mean performance reported by Duchaine and Nakayama (2006) in their original paper. However, the performance is on par with that of Israeli participants. It is possible that the CFMT and ultimately CFMT+ are better suited to match a Southern European Caucasian sub-type, rather than to be used with participants of British or Northern European ancestry (Bowles et al., 2009). These results, in line with Bowles et al. (2009), further highlight the need for country-specific CFMT+ normative data.

One other explanation for this difference is the age of participants used in both studies. In Russell et al (2012) study, participants mean age was 42.2, whereas this study recruited young university students (mean age = 21.4). There is some evidence to suggest that face recognition ability matures late (Susilo, Germine, & Duchaine, 2013) and it is possible that this difference in the CFMT+ performance reflects the age difference between the two samples. Nonetheless, this highlights the importance of using appropriately matched control samples in face recognition research.

Six individuals (two female) in the sample of young adults met the proposed criteria for SR classification in the CFMT+, amounting to 2.36% of the group. This is unsurprising, given that in normally distributed data, approximately 2.28% of cases fall outside the 2SD cut-off. This limits the case that can be made that these individuals genuinely are SRs, as it simply identifies the top two per cent of the population, drawing into question our definition of super recognition. The obvious answer to this issue is to use multiple tests of face recognition to ensure that the same individuals consistently achieve high performance. While the CFPT is one candidate (and is discussed below), it should be noted that it is not measuring the same process as the CFMT+, and enhanced face perception skills are not necessary for an over-riding classification of SR.

In sum, the CFMT+, an extension of the widely used CFMT (Duchaine & Nakayama, 2006), emerged as a good assessment tool for super recognition, with a normal distribution of scores sufficient for the use in single case approach commonly employed in neuropsychological research. Nonetheless, it is also apparent that well-matched control samples and extensive normative datasets are needed to draw meaningful conclusions from studies using this method of face recognition ability assessment.

4.2. CFPT

The CFPT data in this study was not normally distributed and there was a large variability in performance in both upright and inverted trials. One previous study using the CFPT in typical and SR participants (Russell et al., 2012) reported a similar mean score for the control sample (*Mean* = 35.4) and the mean performance for SRs was 24.7 - a considerably different score from the cut-off point established in this study.

It is important to note, however, that the scoring scheme in the CFPT (the number of deviations from the perfect location of one face is converted to an equivalent number of errors) inevitable means that the other five faces in the array are in incorrect positions.

Given the narrow margin of error, that the CFPT has been developed for studies with DP participants (Duchaine et al., 2007), and that only group statistics only have been reported in two previously published SR studies (Russell et al., 2009; 2012), we do not recommend this test for the screening of SR participants in future studies. Indeed, no participant scored more than 2SDs above the mean in the current study. Instead, this test can serve as a non-binding guidance of participants' face perception ability until more sensitive tests are developed.

4.3. Gender effects

In two previous studies reporting gender effects in face recognition, it was found that females had a 2.5-point (Duchaine & Nakayama, 2006) and 2.7-point (Bowles et al., 2009) advantage in performance over their male counterparts on the CFMT task. In both studies these differences did not reach significance. In the sample reported here, females outperformed males by 3.48 points on their overall score on the CFMT+ (Russell et al., 2009). Pertinently, these differences emerged in the novel stages of the CFMT+ (CFMT2 and CFMT3) but did not extend to the more difficult CFMT4. These discrepancies could have emerged for a number of reasons. Firstly, in the CFMT1 participants are asked to recognise six faces from identical images to those presented in the study phase for each of the identities. This task relies on image matching rather than face recognition ability, using pictorial representations created throughout the study phase. It is likely that this relatively easy section has a ceiling effect for both males and females, obscuring any differences in performance. In contrast, CFMT2 and CFMT3 test actual face memory by presenting novel views of the identities, requiring activation of the view-independent representations created throughout the familiarisation period. It is possible that females are more proficient at creating these view-independent representations following a restricted number of exposure times. What is more, although the test faces in CFMT3 are partially blurred by the overlay of visual noise, the configural and fine-grained details of all faces are largely available. This is not the case in CFMT4, and it is possible that the heavy visual noise is obstructing the detection of face configuration per se and as such makes identity judgements near impossible for the typical perceiver. It is thus likely that the fourth part of the CFMT+ is more suited to assessing the initial level of face detection, rather than higher order relations necessary to extract identity. This process may not be influenced by participant gender. In contrast, it is possible that females process faces more holistically than males, which would have aided performance on CFMT2 and CFMT3. Pertinently, there was a significant difference, albeit with a small effect size, between participants on the CFPT inversion effect, adding support to this hypothesis. Future research may wish to investigate this possible mechanism underpinning the female advantage in face recognition ability.

Importantly, there was also a small, albeit significant advantage of females over males in the CFPT task. Similarly to the Bowles et al. (2009) report, this minor gain was combined with a somewhat smaller variability in the female group. Nonetheless, this produced merely a one point difference in the proposed cut-offs and as such, the strict gender match of control participants to a SR sample is not necessary.

More general explanations for the gender difference may relate to the salience of faces for each group. It is possible that females are more interested in faces due to their saliency and importance for social interaction. Indeed, gregariousness has been identified as a predictor of face recognition ability (Li et al., 2010) and it is possible that the female group was more adept in extracting socially relevant information such as identity. Instead, it is plausible that males have an evolutionary predisposition to attend to other facial characteristics of male faces such as dominance, which may be more relevant in all-male groups. Alternatively, these findings may represent a more general memory advantage for women, in line with previous findings suggesting that females are better at verbal memory and memory for location of objects (Galea & Kimura, 1993) than males. Importantly, females were also previously found to be better at recalling autobiographical events, regardless of their emotional intensity (Seidlitz & Diener, 1998). Given that life events often involve interaction with other people, it is possible that females have more efficient encoding strategies than their male counterparts.

The results from this study are the first to indicate that females are significantly better than males at both face recognition memory and face perception. This finding stands in contrast to previous work reporting an advantage for females in the recognition of female, but not male, faces (Lewin & Herlitz, 2002, McKelvie et al., 1993) and weakens the meaningfulness of the "own-group bias" (Bernstein, Young, & Hugenberg, 2007) explanation for this effect. Indeed, young British females outperformed males at the CFMT+ and the CFPT tasks, which both only use male faces. While there are no equivalents of these two tests utilising female faces as stimuli, it is plausible to assume that females are, in general, better at face-processing that their male counterparts. It is important to mention that while the sample of participants in this study was not selected according to face recognition ability, the participants were recruited opportunistically rather than randomly. It is thus not improbable that the groups differed in their face recognition ability rather than this finding representing a true gender difference. Nonetheless, a seminal study investigating face recognition ability in Australian and Israeli samples (Bowles et al. 2009) showed a similar, albeit non-significant, result, indicating a small advantage in face recognition ability for females. The findings from this study are the first statistically significant evidence to support this previously suggested gender difference in the ability to process unfamiliar faces.

4.4. Self-report

The initial decision to asses an individual for super recognition typically follows a self-report of extraordinary face recognition ability or referral by a member of family or friend (e.g. Russell et al., 2009, 2012). Research has long been interested in the reliability of self-reported face recognition aptitude (Bowles et al., 2009; Kennerknecht et al., 2006; Palermo et al., In press; Shah et al., 2015) and to date all but one study (Shah et al., 2015) has showed that people have limited insight into their face recognition ability. The results from this Chapter further corroborate these reports. Pertinently, self-reported face recognition ability was weakly, albeit significantly, correlated with the CFMT+ and CFPT. The accompanying three questions used to assess participants' perceptions of their face recognition ability in various real-life situations (see Appendix A) also yielded weak or no correlations with the two objective measures of face processing ability used in this study.

In sum, these results show that people have limited insight into their face processing skills and that the self-report should be treated with caution. However, it is also probable that the CFMT+ and the CFPT do not resemble real-life person recognition and the self-report measures would correlate better with applied tasks where there are more cues to recognition than the internal features. This is of particular importance to national security settings and the recognition of people from CCTV where, unless deliberately covered, the external features are typically available in an image. Future studies should endeavour to correlate self-report measures of face recognition with applied face processing tasks to examine this hypothesis.

4.5. Conclusion

Taken together, this study further highlights the importance of establishing the norms in the country of origin for the purpose of neuropsychological research and is the first to show significant gender differences in face recognition memory and face perception in a sample of young adults. The data reported here suggests that face recognition ability is largely normally distributed and that it is plausible to assume that DPs and SRs represent the opposite ends of one continuum. It is thus recommended that future studies in face processing use appropriate ethnic and gender-matched samples in order to draw meaningful conclusions from data. This study also provides converging evidence that we have limited insight into our face recognition ability. Further, it is possible that the CFMT+ and the CFPT are not suitable for assessing real-world face recognition skills at the top end of the spectrum, and more applied tasks may be needed to assess the true extent of people's face recognition ability.

Chapter 5: Applied value of super-recognition: Evidence from line-up and face recognition paradigms

Chapter published as:

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

1. INTRODUCTION

In forensic settings such as border control or crime investigation, successful matching of identities from a photograph to a real person is a matter of national security. Passport checks are now often performed by electronic gates with specialised image processing software, yet when this method fails, or, within the European Union, when an individual is from outside of the European Economic Area, the ID bearer is sent directly to a passport officer. Thus, the identity decision ultimately relies on the ID-to-person comparison skills of a human observer. However, the literature reviewed in Chapter 1 presents a compelling case that most typical perceivers struggle with face-processing tasks involving unfamiliar faces (e.g. Bindemann, Avetiysan, & Rakow, 2012; Bindeman & Sandford, 2011; Burton, Wilson, Cowan, & Bruce, 1999), and even trained passport control officers are not better at matching identities than lay persons, such as university students (White, Kemp, Jenkins, Matheson, & Burton, 2014). Further, few attempts to train face matching skills in typical perceivers have been reported. Two investigations have not been able to improve face matching skills whatsoever (Towler, White, & Kemp, 2014; Woodhead, Baddeley, & Simmonds, 1979), although a recent study suggested that feedback can be a somewhat useful training technique yielding an improvement of approximately 2% following a training period (White, Kemp, Jenkins, & Burton, 2014 but see Alenzi & Bindemann, 2013).

Given that any gains observed in training studies have been minor, an alternative is to employ SRs to carry out these tasks given they naturally have extraordinary face processing skills that are well beyond those of most typical perceivers. These individuals may not only be more adept at making identifications from CCTV footage in a Court of Law, but may also be more proficient at face recognition tasks in national security settings, such as passport control. For identifications in a Court of Law, current practice in the UK is to employ facial image analysts to evaluate the similarity of pictorial evidence to the suspect using a number of techniques - a task at which trained experts appear to be better than the general public (Wilkinson & Evans, 2009). However, despite utilising various techniques in face mapping, the current guidelines of the Forensic Imagery Analysis Group (FIAG) state that a positive identification can only be made on a basis of distinctive features such as marks, scars, or tattoos (The Law Society Gazette, 2004). It should be noted that these employees therefore operate on a very different basis to SRs, whose excellent face recognition skills do not rely upon the use of distinguishing facial characteristics. Indeed, it is possible that if SRs perform well on laboratory-based tests, they would also achieve better results in applied and real-world tasks involving face matching or face recognition. If this is the case, SRs may be an invaluable resource for forensic and national security agencies.

The current investigation examined whether seven SRs are more proficient than control participants at two face processing tasks that resemble real-world scenarios. In Experiment 1, participants had to match still images (taken from video footage) of unfamiliar male faces to arrays containing ten photographs. Similar circumstances can occur in real-world scenarios when the face of a perpetrator is captured on CCTV, and a suspect has been apprehended based on their apparent similarity to this person. In Experiment 2, participants were required to encode a set of faces and subsequently discriminate them from distractor faces when viewed within brief clips of video footage. Such a scenario could occur when missing or wanted persons are spotted by police officers on patrol. No study to date has investigated the ability of SRs to recognise faces from video stills to photographs - a task that typical perceivers are remarkably poor at (Bruce et al., 1999). Similarly, while there is strong evidence that typical observers are poor at remembering unfamiliar faces (Burton et al. 1999), there are no data demonstrating the performance of SRs on this type of task. The main aim of this investigation was therefore to establish whether SRs are better at (a) face matching, particularly when the viewpoint of target and line-up faces differ, and (b) identifying faces from video when a short delay and interference are introduced.

2. EXPERIMENT 1

In an initial experiment we investigated the performance of SRs and typical perceivers on a face matching task that has been well-validated in previous research (Bindemann, Brown, Koyas, & Russ, 2012; Bruce et al., 1999; Megreya et al., 2013). On each trial, participants were required to match a video still of an unfamiliar male face to an array containing 10 faces. Both target-present and target absent-arrays were included, with targets appearing in the arrays on half of the trials. We predicted that SRs would be considerably better at the matching task than typical observers, perhaps due to enhanced encoding strategies and/or more efficient retrieval, and that they would report greater confidence in their responses.

2.1. Method

Participants

Seven SRs participated in this study, and their demographic information is reported in Table 1. Five have already been described in Chapters 2 and/or 3. As previously described, super recognition was confirmed in these participants using anecdotal evidence and performance on the CFMT+ (Russell, Chatterje & Nakayama, 2012). As a group, the seven SRs performed significantly better (M =95.75, SD = 2.28) than the published control norm for this test (M =75.20, SD = 11.60), $t(31) = 4.61$, $p < .001$, d =1.96, 95% CI [.98, 2.92], and on par with SRs from previous studies (Russell et al., 2009; Russell et al., 2012; all *p*s > .05; see Table 1). The individual SRs outperformed the published control mean by 1.4 to 2.1 standard deviations. Although these scores do not meet the criteria of two standard deviations above the control mean for categorization as SRs, all seven individuals were retained in the study as the variability in their performance allowed us to additionally examine whether CFMT+ score correlates with face-processing performance on more real-world tasks. Indeed, it may be the case that some individuals who describe exceptional face recognition skills in everyday life may not necessarily outperform controls on the CFMT+, yet do achieve superior scores on more ecologically valid tasks that represent real-world face recognition tasks.

Control participants (N = 22) were recruited amongst students and visitors to the University of Stirling. All control participants reported typical face recognition skills, and this was confirmed using the CFMT. However, two participants obtained CFMT scores within the impaired range (a score of 42/72 or below; see Duchaine & Nakayama, 2006), and were excluded from all analyses (the CFMT scores of the remaining participants ranged from 44/72 to 69/72). Hence, a total of 20 controls (10 male) were included in this experiment, and their age ranged from 18 to 34 years (M = 25.1, SD = 6.0). All participants had normal or corrected to normal vision, and participated in exchange for a small monetary payment or course credits. Ethical approval for this study was granted by the University of Stirling Psychology Ethics Committee.

Table 1. Demographical information and CFMT+ scores for the SR participants used in this study and those described by Russell et al. (2012). Published norms for typical perceivers on this test are also presented by Russell et al. (2012): *M* = 75.20, *SD* = 11.60.

	Russell et al. (2012) (N = 6)	The current study SRs (N = 7)	SR1	SR2	SR3	SR4	SR5	SR6	SR7
Age	40.7 *(9.9)*	28.7 *(7.0)*	36	19	37	28	21	23	27
Gender	-	M = 4	M	M	F	M	M	F	F
Hand	-	R = 6	R	R	R	R	R	L	R
CFMT+	95.0 *(1.9)*	95.71 *(1.53)*	92	97	97	97	100	93	94
CFPT	24.7 *(10.3)*	18.86 *(5.14)*	20	12	20	26	16	14	24

Materials

This experiment used the same stimuli that were developed by Bruce et al. (1999). Each trial consisted of a still image extracted from video footage, displaying a male face from a 30° viewpoint. The target image was simultaneously presented above an array of 10 male faces depicted from full face viewpoints and arranged in a 5 x 2 line-up. All images were displayed in colour. On half of the trials, the 10 photographs resembled but did not contain the target image (target-absent condition), and on the other half the photograph of the target was amongst the 10 photographs (target-present condition). Photographs in the arrays were numbered 1–10 and the position of the target varied with the constraint that each position was used four times. These photographs were cropped so that no clothing was displayed (but the external facial features including the hair were still visible), and were arranged in arrays stretching 1050 x 700 pixels. Target video-stills were cropped below the shoulders so that some clothing was still visible. These images measured 216 pixels (*W*) x 263 pixels (*H*). An example trial array is presented in Figure 1.

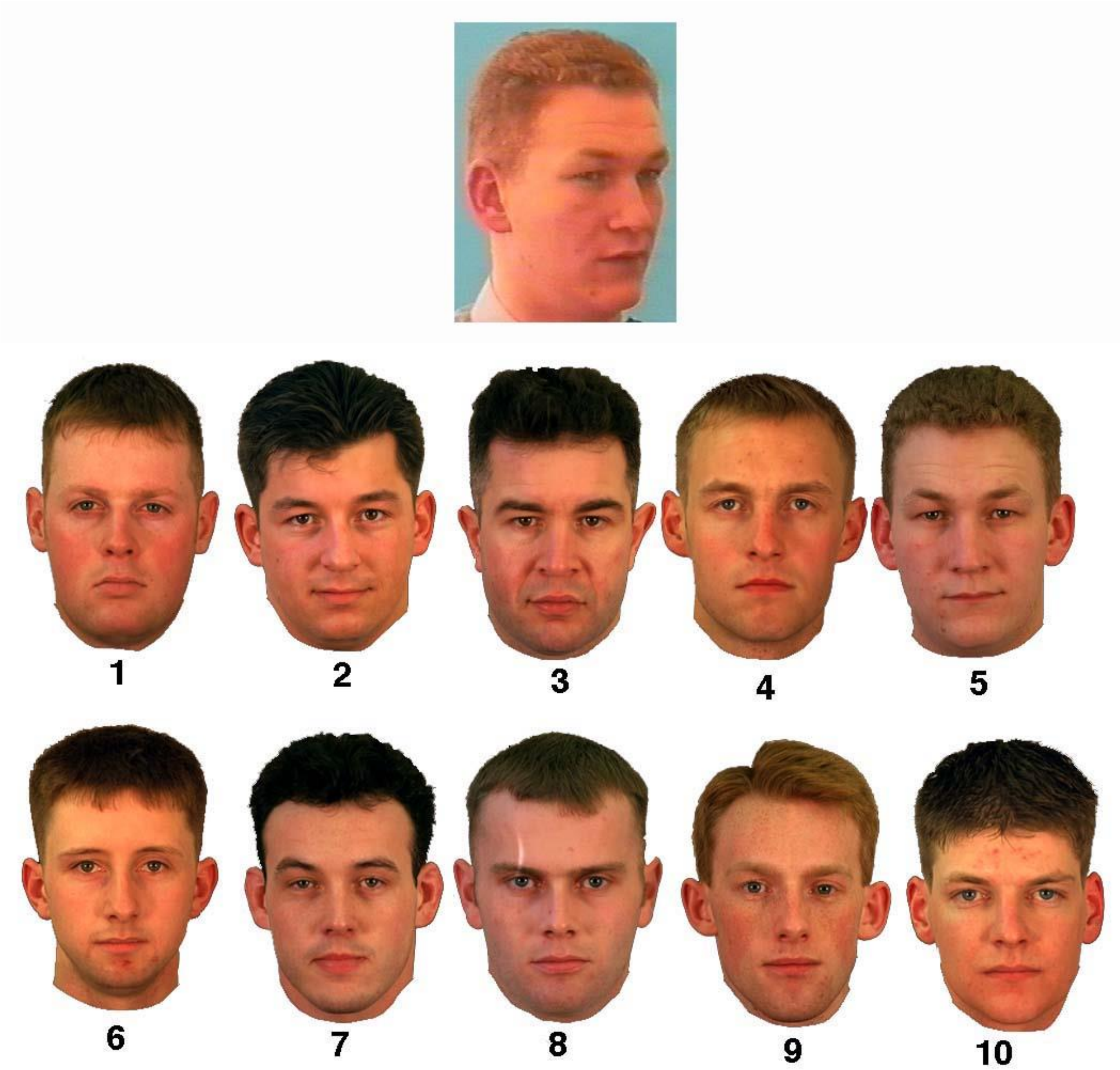

Figure 1. An example trial from a face matching array (Bruce et al., 1999; not drawn to scale). The target or probe is a video still. The images are those paired with the target by Bruce et al. (1999). The target is present in position 5.

Design and Procedure

A 2 x 2 mixed factorial design was employed. The within–subjects factor was target presence (target-present/target-absent) and the between–subjects factor was participant group (controls/SRs). Each individual saw all 80 arrays, with the target present on half of the trials. Trials were presented in a fully randomised order.

All participants were tested individually using E–Prime software (Psychology Software Tools, Sharpsburgh, PA, USA) and a 15.6 inch LCD monitor displayed at a resolution of 1366 x 768 pixels. Participants sat at a comfortable distance from the screen, and elicited their responses using keys on a keyboard under no time constraints. They were instructed that they were going to complete a face matching task, and that the target may or may not be present in each line-up. Targets were identified using the corresponding keyboard number (for targets identified as present in location 10, participants were instructed to press 0). For arrays where no match was detected, participants were told to press the space bar. After each array, participants rated their confidence in their response on a scale from 1 (not at all confident) to 5 (very confident).

Statistical Analyses

The percentage of hits (correct identifications in target-present arrays), misses (no–match decisions in target–present trials), false identifications (selection of a non-match face in target–present trials), correct rejections (correct target-absent responses in target-absent arrays), and false positive responses (false identification of a face in target-absent trials) were calculated for each participant.

While accuracy is a good indicator of the overall patterns of responses between the SR and control groups, additional analyses regarding sensitivity and response criterion permit more in-depth understanding of the differences between them, and are useful for cross-study comparison (McIntyre, Hancock, Kittler, & Langton, 2013). In line-up paradigms with target-present and target-absent arrays, there are five types of response (see above) which is not typical for signal detection analyses. We therefore only used correct responses from target–present trials (hits) and false–positive responses from target-absent trials in these comparisons. Specifically, d prime (*d'*), a measure of sensitivity, was calculated by subtracting the *z* scores for false–positive (F) responses in the target–absent trials from *z* scores calculated from correct identifications (hits, H) in target-present trials [$d' = z(\mathrm{H}) - z(\mathrm{F})$] (see Table 2). Response bias (*criterion c*) was calculated as the negative average sum of *z* scores for the hits and false-positive response $c = -0.5[z(\mathrm{H}) + z(\mathrm{FPs})]$ (MacMillan & Creelman, 2005).

2.2. Results

Accuracy

A 2 x 2 repeated measures analysis of variance (ANOVA) was conducted on accuracy scores, with array type as the within–participant factor (target-present/target-absent) and group as the between-participant factor (SRs/controls). A significant main effect of array type indicated that performance was more accurate for target present (M = 87.20%, SE = 1.80%) than for target absent (M = 78.60%, SE = 4.30) trials, $F(1,25) = 4.54$, $p = .043$, $\eta_p^2 = .154$, 95% CI [.00, .39] There was also a significant main effect of group with a higher percentage of correct responses being made by the SRs (M = 92.90%, SE = 4.60) than controls (M = 73.00%, SE = 2.70), $F(1, 25) = 14.55$, $p = .001$, $\eta_p^2 = .368$, 95% CI [.08, .57]. Although the interaction between array type and participant group was not statistically significant, this may be due to the small number of participants in the SR group, F(1, 25) = 3.16, p = .088, $\eta_p^2 = .112$, 95% CI [.00, .35] (see Table 2). Indeed, independent analyses of group differences on target-present and target-absent trials revealed that SRs performed significantly better than controls regardless of array type, $t(25) = 3.50$, $p = .002$, $d = 1.37$, 95% CI [.57, 2.49] and $t(25) = 3.13$, $p = .004$, $d = 1.37$, 95% CI [.42, 2.30], respectively.

Table 2. Mean (SD) accuracy score for SRs and control participants in Experiment 1. Responses are calculated from the *z* scores of hits and false-positive identifications from target-absent arrays

	Target Present (%)			Target Absent (%)		
	Hits	Miss	False ID	Correct Rejection	*d'*	*c*
SRs	93.57	1.07	5.36	92.14	3.20	-0.01
	(4.75)	(1.33)	(4.61)	(7.69)	(0.59)	(0.37)
Controls	80.87	9.75	9.37	65.12	1.42	-0.25
	(9.07)	(8.26)	(6.17)	(21.98)	(0.90)	(0.30)

Analyses of mistakes on target-present trials (false IDs, i.e. choosing the wrong face in target-present arrays and misses, errors on target–present arrays where participants responded “no match”) were conducted using a 2 x 2 repeated measures ANOVA with mistake type as the within–participant factor (miss/false ID) and group as the between-participant factor (SRs/controls). SRs made fewer mistakes overall on target-present trials ($M = 3.20\%$, $SE = 0.09\%$) than controls ($M = 9.60\%$, $SE = 1.60\%$), $F(1, 25) = 12.28$, $p = .002$, $\eta_p^2 = .329$, 95% CI[.06, .54]. The main effect of mistake type and the interaction did not reach statistical significance, $F(1, 25) = 0.75$, $p = .393$, $\eta_p^2 = .029$, 95% CI [.00, 23] and $F(1, 25) = 1.07$, $p = .310$, $\eta_p^2 = .041$, 95% CI [.00, 25] respectively.

Signal detection analyses

An independent–samples *t*-test on *d'* scores revealed higher sensitivity in SRs than controls, $t(25) = 5.15$, $p <.001$, $d = 2.45$, 95% CI [1.18, 3.30] (see Table 2). A *t*-test on *criterion c* scores revealed a non-significant difference between the groups $t(25) = 1.48$, $p =.151$, $d = 0.64$, 95% CI [-.23, 1.52]. In sum, these findings indicate, in line with accuracy analyses, that SRs are better at unfamiliar face discrimination in a matching task. On an individual level, we performed modified *t*-tests for single-case comparisons (Crawford, Garthwaite, & Porter, 2010) on the *d prime* and *criterion c* data for each of the SRs in comparison to controls. Four participants, SR1, SR3, SR4, and SR7 showed significantly higher sensitivity than controls on the matching task. The response bias of SR7 was also more conservative than that of the control group. Full results are presented in Table 3.

Confidence

Overall SRs were more confident ($M = 4.29$, $SD = 0.40$) in their responses than control participants ($M = 3.69$, $SD = 0.36$), regardless of the trial type or actual accuracy $t(25) = 3.63$, $p = .001$, $d = 1.38$, 95% CI [.63, 2.53]. Mean confidence scores for correct responses were analysed using a 2 (response type: hit/correct rejection) x 2 (groups: SRs/controls) mixed factorial ANOVA. The analysis revealed a significant main effect of response type, $F(1, 25) = 30.73$, $p < .001$, $\eta_p^2 = .551$, 95% CI [.25, .70] with participants reporting higher confidence following correct responses in target–present ($M = 4.36$, $SE = 0.08$) than target–absent ($M = 3.78$, $SE = 0.11$) trials. There was also a significant main effect of group $F(1, 25) = 10.41$, $p = .003$, $\eta_p^2 = .294$, 95% CI [.04, .51] with SRs reporting higher confidence for correct responses ($M = 4.33$, $SE = 0.14$) than controls ($M = 3.80$, $SE = 0.08$). The interaction between response type and group was non-significant, $F(1, 25) = 2.92$, $p = .100$, $\eta_p^2 = .104$, 95% CI [.00, .34]. As some participants did not commit each type of error, confidence for incorrect responses was analysed separately for misses, false IDs and FPs using independent-samples *t*-tests. Participants whose data were missing were not included in analyses involving these mistakes. The differences between groups on confidence ratings following incorrect responses did not reach statistical significance, all *ps* > .05.

CFMT and matching task performance

Because CFMT scores were available for all our control participants (controls were initially screened using the CFMT), we correlated performance on this task with the matching test. Performance on the CFMT and sensitivity on the 1-in-10 matching task is presented in Figure 2. Each marker represents a separate data point. Notably, performance on the CFMT did not predict discriminability on the matching task, $N = 20$, *Spearman's rho* = .238, $p = .311$ (see Figure 2).

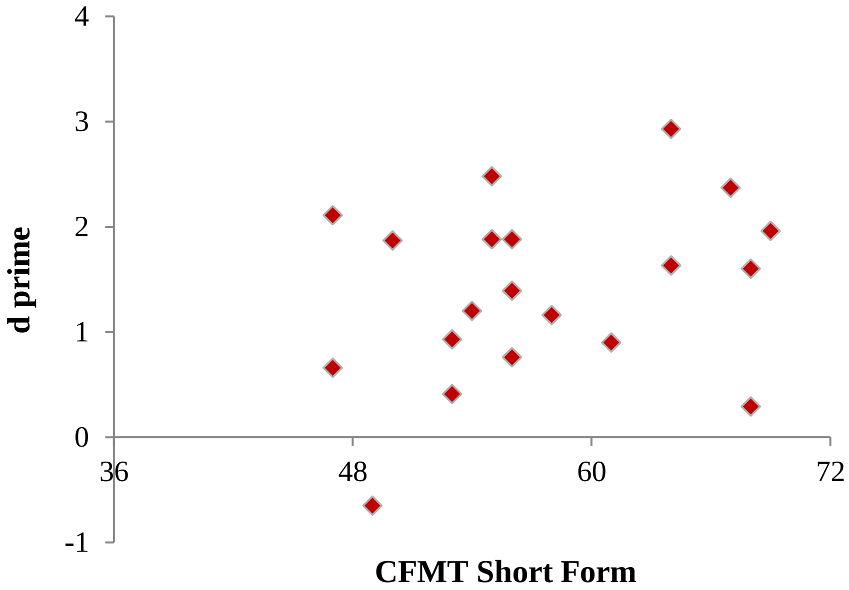

Figure 2. D prime on the 1-in-10 task in Experiment 1 plotted against CFMT score. The dependent variable in the CFMT is the number of correct responses from a maximum score of 72.

Table 3. Individual case analyses of sensitivity of SRs in Experiment1, using modified *t*-tests for single-case comparisons (Crawford et al. 2010)

	Control Mean *(SD)* (N = 20)	SR1	SR2	SR3	SR4	SR5	SR6	SR7
d'	1.39 *(0.85)*	3.92	2.37	3.24	3.92	3.11	2.58	3.28
t(19)	-	2.90	1.12	2.12	2.90	1.97	1.37	2.17
p (two –tailed)	-	.009	.274	.047	.009	.063	.188	.043
Zcc	-	2.98	1.15	2.18	2.98	2.02	1.40	2.22
95% CI for *Zcc*	-	[1.93, 4.00]	[0.57, 1.71]	[1.35, 2.98]	[1.93, 4.00]	[1.24, 2.78]	[0.77, 2.01]	[1.39, 3.04]
% population below individual's score	-	99.54	86.27	97.64	99.54	96.85	90.60	97.85
criterion c	-0.23 *(0.32)*	0.00	-0.25	0.33	0.00	-0.40	-0.35	0.60
t(19)	-	0.69	-0.07	1.71	0.69	-0.53	-0.38	2.50
p (two –tailed)	-	.499	.945	.104	.499	.604	.710	.021
Zcc	-	0.71	-0.01	1.75	0.71	-0.54	-0.39	2.56
95% CI for *Zcc*	-	[0.21, 1.19]	[-0.51, 0.36]	[1.03, 2.44]	[0.21, 1.19]	[-1.00, 0.06]	[-0.84, 0.07]	[1.63, 3.47]
% population below individual's score	-	75.03	47.26	94.79	75.03	30.21	35.49	98.91

2.3. Discussion

This experiment investigated the performance of SRs and typical perceivers on a face matching task consisting of target-present and target-absent line-ups. Our prediction that SRs would outperform controls on this task was supported: SRs were better at the task both in terms of accuracy and perceptual sensitivity measures. SRs were also more confident in their responses than controls in overall analyses, but did not differ from controls when making a “no match response” on target–absent trials. The overall advantage in confidence for SRs may arise from self-awareness of their above–average face-processing skills. On an individual level, four SRs performed significantly better than the control group, and one SR was more conservative in their response bias than controls.

Importantly, the CFMT did not predict accurately the performance on the one-in-10 task (Bruce et al., 1999) in typical observers. This is perhaps unsurprising, because the CFMT (Duchaine & Nakayama, 2006) has a strong memory component whereas the matching task relies merely on perception of the presented faces. A recent study by White et al. (2014) reported that performance on the Glasgow Face Matching Test (GFMT; Burton, White, McNeill, 2010) was a good predictor of performance on a photograph-to-person matching task. The lack of association between accuracy on a memory and a matching paradigm in this study suggests that face perception and face memory are, at least to some extent, dissociable.

3. EXPERIMENT 2

In Experiment 2, we examined face memory in SRs compared to typical perceivers. In the study phase, all participants were presented with 20 good quality stills of male and female faces displayed from a frontal viewpoint. After a letter search filler task, participants viewed 40 video clips where targets were present on half of the trials. In certain situations (such as the 2011 England Riots) multiple people are sought by the Police, increasing memory load for unfamiliar faces. Whilst this task does not directly resemble real-life situations (i.e. where representations of faces are stored incidentally rather than explicitly), it is a convenient way of testing face memory in laboratory settings. We predicted that SRs would perform considerably better on this task and report higher confidence ratings in their responses than control participants.

3.1. Method

Participants

The same seven SRs as described above took part in Experiment 2 (see Table 1). Twenty control participants (10 male) were also recruited amongst students and visitors to the University of Stirling. Their ages ranged from 19 to 33 years (*M* = 24.4, *SD* = 5.7). As for Experiment 1, all control participants were screened using the CFMT (Duchaine & Nakayama, 2006) in order to exclude those potentially affected by prosopagnosia or meeting the criteria for super recognition. No participants were excluded from the control sample, however. The CFMT scores of control participants ranged from 47/72 to 69/72. All controls participated for course credit or a small monetary payment. The study was granted ethical approval by the University of Stirling Psychology Ethics Committee.

Materials

Twenty good quality facial images (10 male) extracted from video footage were taken from the Psychological Image Collection at Stirling (PICS, http://pics.psych.stir.ac.uk/). The stills measured 390 pixels (*W*) x 480 pixels (*H*) and were cropped below the neck.

They were not converted to grayscale to mimic the natural settings when a search for a missing or wanted person would occur. All visible clothing and jewellery was removed using Adobe Photoshop. Forty video clips (20 containing the person depicted in the still images) were extracted from the same database. The clips were sized 640 pixels (*W*) x 480 pixels (*H*) and played at frame rate of 25 frames per second. All videos were adjusted to a duration of 5 seconds using MAGIX Movie Edit Pro. Each clip depicted individuals walking down a corridor in dim lighting, and half were male. Example stimuli from the study and test phases are presented in Figure 3.

Figure 3. Example stimuli from study and test (not drawn to scale) in Experiment 2. The target is present in the video clip on the right.

Design and procedure

A 2 (target presence: present, absent) x 2 (participant group: SRs, controls) mixed factorial design was employed. All participants were tested individually using E–Prime software (Psychology Software Tools, Sharpsburgh, PA, USA) with a 15.6 inch LCD monitor displayed at a resolution of 1366 x 768 pixels. They sat at a comfortable distance from the screen and responded using the keyboard. In the study phase, participants were instructed that they should view the 20 individuals carefully as if they were missing or wanted. The stills were presented in a random order for 5 seconds each. Participants then took part in a simple letter search filler task which was followed by a break, so that the total time between the study and test phases was 20 minutes. At test, participants were instructed to view the 40 clips (presented in a random order) and indicate whether the person in the video was familiar or not using the *M* or *Z* keyboard keys (response mapping was counterbalanced between participants). The duration of all clips was 5 seconds and if participants did not respond during that time, the experiment proceeded to a response screen where participants were prompted to make their response. After each trial, participants were asked to rate their confidence in their response on a scale of 1 (not at all confident) to 5 (very confident).

Statistical Analyses

The percentage of hits (correct identifications in target-present trials) and correct rejections (target-absent trials that were responded to as such) was calculated for each participant. The discrimination of identities appearing in the video clips was also analysed using the signal detection measures of sensitivity and response bias.

3.2. Results

Accuracy

A 2 (target-presence: target-present, target-absent) x 2 (participant group: SRs, controls) mixed factorial ANOVA was conducted on accuracy scores. A significant main effect of group indicated that SRs (*M* = 67.00%, *SE* = 2.80) were better at this task than control participants (*M* = 58.60%, *SE* = 1.80), $F(1, 25) = 5.68$, $p = .025$, $\eta_p^2 = .185$, 95% CI [.02,

.42] (see Table 4). Neither the main effect of clip type (i.e. target-present versus target-absent) nor the interaction with participant group reached significance, $F(1, 25) = 0.57$, $p = .455$, $\eta_p^2 = .022$, 95% CI [.00, .21] and $F(1, 25) = 0.19$, $p = .666$, $\eta_p^2 = .008$, 95% CI [.00, 17] respectively.

Table 4. Performance of SR and control participants in Experiment 2. Sensitivity and response bias are calculated from the *z* scores of hits and false-positive identifications from target-absent trials.

	Hits (%)	Correct Rejections (%)	*d'*	*c*
SRs	64.64 (20.12)	69.29 (9.32)	1.00 (0.59)	0.02 (0.48)
Controls	58.00 (9.65)	59.25 (11.15)	0.45 (0.42)	0.01 (0.18)

Signal detection analyses

An independent-samples *t*-test on *d'* scores indicated that SRs had greater sensitivity than controls, $t(25) = 2.64$, $p = .01$, $d = 1.33$, 95% CI [0.23, 2.06] (see Table 4). However, the equivalent analysis of response bias (criterion *C*) yielded no significant differences between groups, $t(25) = 0.06$, $p = .995$, $d = 0.03$, 95% CI [-0.83, 0.89]. Taken together, signal detection analyses indicate that SRs are better than typical perceivers at discriminating facial identity from poor quality video clips. In line with Experiment 1, we performed case by case analyses on d prime and criterion c for SRs and control participants. Full results are presented in Table 5.

Table 5. Individual case analyses of sensitivity and response bias of SRs in Experiment 2, using modified *t*-tests for single-case comparisons (Crawford et al., 2010)

	Control Mean *(SD)* (N = 20)	SR1	SR2	SR3	SR4	SR5	SR6	SR7
d'	0.45 *(0.45)*	0.64	0.32	0.77	2.09	1.23	0.64	1.35
t(19)	-	0.40	-0.29	0.69	3.55	1.73	0.40	1.95
p (two –tailed)	-	.690	.774	.500	.002	.108	.690	.067
Zcc	-	0.41	-0.30	0.70	3.64	1.73	0.41	1.99
95% CI for *Zcc*	-	[-0.05, 0.86]	[-0.74, 0.15]	[0.20, 1.19]	[1.02, 2.42]	[1.24, 2.78]	[-0.05, 0.86]	[1.22, 2.75]
% population below individual's score	-	65.48	38.74	74.97	99.89	96.85	65.48	96.67
criterion c	0.02 *(0.18)*	0.07	-0.25	0.00	-0.92	0.23	0.07	0.00
t(19)	-	0.29	-0.07	-0.09	-5.12	-0.53	0.29	-0.09
p (two –tailed)	-	.777	.945	.925	<.001	.604	.777	.925
Zcc	-	0.29	-0.01	-0.10	-5.24	-0.54	0.29	-0.09
95% CI for *Zcc*	-	[-0.16, 0.74]	[-0.51, 0.36]	[-0.53, 0.34]	[-6.94, -3.53]	[-1.00, 0.06]	[-0.16, 0.74]	[-0.53, 0.34]
% population below individual's score	-	61.16	47.26	0.003	75.03	30.21	61.16	46.29

Confidence

There was no difference in overall confidence ratings between groups, $t(25) = 0.17$, $p = .861$, $d = 0.07$ 95% CI [-.79, .93]. Mean confidence scores for hits and correct rejections were analysed using a 2 x 2 ANOVA with one between-participants (group: SRs/controls) and one within-participants (response type: hits/correct rejections) factor. The analysis revealed a significant main effect of trial type, with higher confidence reported for correct responses when a target was present in the clip (hits, $M = 3.80$, $SE = 0.11$) than for when participants correctly rejected a video without a target (CRs, $M = 3.15$, $SE = 0.14$), $F(1, 25) = 49.83$, $p < .001$, $\eta_p^2 = .666$, 95% CI [.40, .78]. The main effect was qualified by a significant interaction between the factor trial type and group, $F(1, 25) = 8.96$, $p = .006$, $\eta_p^2 = .264$, 95% CI [.02, .49]. The main effect of group did not reach statistical significance, F(1, 25) = .002, p = .964, $\eta_p^2 = .00008$ 95% CI [.00, .00]. The interaction was explored using independent sample *t*-test for each level of trial type. The analysis, however, revealed no significant differences between groups in confidence ratings for either type of trial (*p*s > .05), reinforcing the non-significant main effect of group in the ANOVA results. For the purpose of analysis of erroneous trials (misses and FPs), one SR who correctly rejected all target absent clips and therefore did not make any FPs was removed from this analysis. A 2 (target- presence: target-present, target-absent) x 2 (participant group: SRs, controls) mixed factorial ANOVA revealed higher confidence levels for FPs ($M = 3.39$, $SE = 0.13$) than misses ($M = 2.90$, $SE = 0.13$), $F(1, 24) = 13.38$, $p = .001$, $\eta_p^2 = .358$, 95% CI [.07, .57]. There were no significant differences between groups and the interaction did not reach statistical significance, $F(1, 24) = 0.65$, $p = .428$, $\eta_p^2 = .026$, 95% CI **[**.00, .23] and $F(1, 24) = 0.31$, $p = .582$, $\eta_p^2 = .013$, 95% CI [.00, 19].

CFMT and face memory scores

As for Experiment 1, we correlated the CFMT scores of all control participants with sensitivity on the video recognition task. The CFMT score was a good predictor of performance N = 20, *Spearman's rho* = .491, p = .028

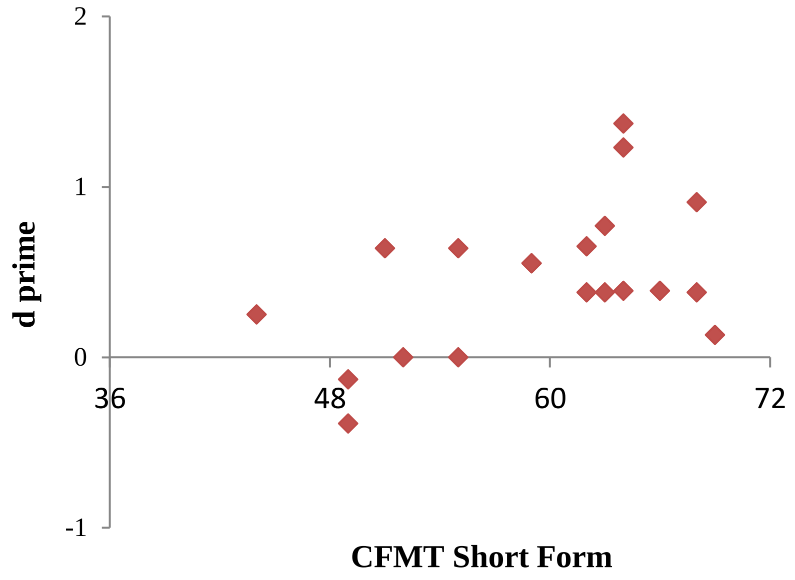

Figure 4. D prime on the face memory task in Experiment 2 plotted against CFMT score. The dependent variable in the CFMT is the number correct from a maximum score of 72.

3.3. Discussion

Experiment 2 investigated whether SRs outperform typical perceivers on an applied test of unfamiliar face memory. It has been previously reported that the performance of the general population on this type of test is remarkably poor and close to chance level (Burton et al., 2009). We predicted that SRs would be better than control participants at recognising unfamiliar faces from poor quality video clips, perhaps due to enhanced encoding and more efficient retrieval upon presentation. Findings from this experiment show that SRs are better than control participants at this task, and suggests they may be useful in applied forensic and security settings. It is noteworthy, however, that the overall accuracy of the SR group was relatively low, with approximately 35% of trials resulting in “misses” and 31% in FPs. On the other hand, control participants performed marginally, but significantly above the chance level on each type of trial, replicating the findings of previous work (e.g. Burton et al., 1999). The low accuracy rate occurred despite the video footage being of relatively high quality. Surveillance systems relying on older technology may produce poorer quality footage recorded from a less favourable viewpoint. The results of this study may therefore be only applicable to higher-quality CCTV systems. Interestingly, while recognition performance was better in the SR group, confidence ratings were not. This may be explained by the high level of difficulty of the task and the brief presentation of stimuli in both the study and test phases. Nevertheless, SRs were still able to outperform typical perceivers, suggesting they may be valuable employees in forensic and security settings. On an individual level, SR4 was the only participant who was significantly better than controls at discriminating between identities in the test phase. Notably, he also showed an increased bias towards making a positive familiarity decision when a target was absent.

Finally, the CFMT was found to be a good predictor of performance on this task, although the correlation was only moderate. It is likely that this is underlined by the relative difficulty of Experiment 2, where many control participants performed at near-chance levels.

4. GENERAL DISCUSSION

This investigation set out to examine the performance of seven SRs on a face matching and a face memory task. In Experiment 1, we compared the performance of SRs to typical perceivers on a well-established 1-in-10 unfamiliar face matching task. In Experiment 2, we tested SRs' memory for faces encoded from high quality CCTV stills that were later presented in poor quality video clips. In both studies we also measured participants' confidence in their responses. Finally, we correlated a standardised test of face memory (the CFMT: Duchaine & Nakayama, 2006) with the overall performance of all participants in Experiments 1 and 2. As predicted, SRs outperformed controls in both experiments, but were only more confident in their performance in the face matching task. Further, CFMT performance was found to be a good predictor of performance on face recognition, but not the face matching task.

There are a number of theoretical and practical implications that arise from this pattern of findings. In terms of theory, research examining unfamiliar face processing has long been concerned with viewpoint and lighting change between study and test images. For instance, Longmore, Liu & Young (2008) reported that change in the direction of illumination and pose between study and test leads to reduced face recognition performance. In matching studies, where there is no memory load in the task or the load is minimal, viewpoint and lighting changes also lead to decrements in performance (Braje, 2003; Bruce et al., 1999). In our study, the arrays used in Experiment 1 were taken from the stimuli set developed by Bruce et al. (1999). We used the most difficult variant of the task, where target and line-up photographs differed according to viewpoint, and different cameras were used to capture each image. According to the Bruce and Young (1986) model of face recognition, a change in viewpoint should activate structural rather than pictorial encoding strategies, so the invariant parts of the facial image are encoded. While participants in the control group performed on par with the original line-up study where the reported error rate was approximately 40% (with the exception of the target-present condition in Experiment 1 where accuracy was higher), SRs were significantly better at face matching with mistakes being made on less than 8% of the trials in both conditions. This finding suggests that SRs may have better mechanisms for the structural encoding of faces than typical perceivers. A more recent cognitive neuroscience model developed by Haxby and colleagues (Haxby, Hoffman, & Gobbini, 2000, 2002) proposes that static components of a facial image, such as a view independent representation, are analysed via a route involving the lateral fusiform gyrus. It is possible that this neural mechanism is better developed in SRs, whereas typical perceivers rely heavily on pictorial representations, prone to disruption by viewpoint changes. Another possibility is that SRs are more proficient at manipulating pictorial codes than typical perceivers. A study using un-textured 3D faces to investigate performance of SRs would elucidate this phenomenon (e.g. Bruce & Langton, 1994).

However, the results are somewhat less clear cut in Experiment 2. While the overall accuracy performance of the SR group was higher than that of the control group, overall performance was generally poor with error rates of 33% and 42% for SR and control participants respectively. It is possible that this is due to the particular paradigm we employed. Participants in our study performed a letter search task for 15 minutes and had a short break of 5 minutes between study and test. This was in order to introduce some distractions that may occur in real life, while not requiring participants to participate on two separate occasions and retaining a degree of experimental control. Yet, the filler task required local letter recognition and was performed directly before recall of the encoded faces. This may have interfered with the newly learned faces by promoting a suboptimal featural processing strategy (Hills & Lewis, 2009; Lewis, Mills, Hills, & Weston, 2009; cf. Farah, 1991). Nonetheless, albeit prone to errors, performance of the SRs was remarkably better than that of control participants, despite different image quality, resolution, size and lighting between the study and test phases. However, it is important to note that while all the above variables were different, viewpoint remained unchanged between study and test aside from individual variations in head tilt in the test phase (the individuals presented in the clips walked for 5 seconds along a dimly lit corridor). It is possible that participants were trying to activate pictorial codes in order to reach a decision, but the video clips were insufficient to match a stored image of a studied face against one presented at test. SRs, as argued before, may have developed a more efficient mechanism for structural encoding and thus are able to form more stable representations of studied faces that facilitate recognition under unfavourable conditions.

The implications for forensic practice are also important to note. Existing evidence suggests that (a) people are poor at unfamiliar face matching (Bruce et al., 1999; Kemp et al., 1997; Megreya et al., 2013) and recognition (Burton et al., 1999), (b) the results of training in face matching are limited (Towler et al., 2014, Woodhead et al., 1979; cf. Kemp et al., 2014), and (c) that the face matching ability of employees in security settings is unrelated to years of service and, ultimately, experience (White et al., 2014). Our study clearly shows that SRs are significantly better at face matching (Experiment 1) and recognition (Experiment 2) than a sample of typical observers. Results from Experiment 2 also show that performance on a standardised test of face recognition (the CFMT: Duchaine & Nakayama, 2006) is related to accuracy in a face recognition task with a strong memory component. Current training studies often use shape classification to improve face recognition, a method which has had little success to date (Towler et al, 2014). A possible way of improving training programmes is to apply findings from detailed investigations of the processes underlying extraordinary face recognition. Another, more readily available, solution would be to select specific personnel for positions requiring excellent face processing skills based on standardised aptitude tests, such as the CFMT+. The findings of this study suggest that the standard version of CFMT

(Duchaine & Nakayama, 2006) commonly used in neuropsychological research is of limited utility in predicting performance on applied tasks. While it moderately predicts discriminability on a simple memory paradigm, the same is not true for face matching. Nonetheless, this study shows that SRs identified by the long version of the CFMT, the CFMT+ (Russell et al., 2009), are significantly better at face matching, and recognition-skills that are essential for police officers or personnel working with CCTV footage. The Metropolitan Police currently employ SR officers, identified within the Force by the means of positive identifications leading to arrests of suspects (Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016). They are often assigned to specific tasks involving CCTV footage, but to the authors' best knowledge, screening of face-processing skills is not a standard practice at the time of recruitment. Such aptitude tests should perhaps become a standard part of the enrolment process for all police officers. This would ensure optimal personnel allocation and, when operated by people with good face processing skills, maximise the utility of CCTV systems.

However, it is important to note that while SRs consistently outperformed control participants as a group, the same was not true in all analyses on an individual level. In Experiment 1, the discriminability of four SRs was significantly better than that of controls, and one SR showed a more conservative response bias. In Experiment 2, one SR performed significantly better than the control group and was also more liberal in their responses. However, another SR performed below the control mean. These single case statistics are a particular strength of the work presented here and show that there is a degree of performance variation within the SR population. These differences may be due to heterogeneity in the cognitive and perceptual processes underpinning superior face recognition, as reflected by the varying performance in Experiments 1 and 2. Previous research with individuals affected by developmental prosopagnosia (face blindness) shows that even within the same family, deficits associated with impaired face processing can vary between relatives (see Susilo & Duchaine, 2013 for a review). Specific processes underlying expert face recognition are still largely unknown, but it is possible that a similar heterogeneity in the SR group is responsible for some, but not all SRs excelling in different types of applied tasks. It is of particular interest, that within the SR group, CFMT+ performance did not appear to correspond directly to participants' discriminability in the applied tasks. For instance, SR1 who obtained the lowest CFMT+ score in the group (92/102), was amongst the highest performers on the face matching task in Experiment 1, and was one of the only participants to significantly outperform controls in case-by-case analyses. Conversely, SR5, who achieved the highest CFMT+ score in the group (100/102), did not discriminate or remember faces significantly better than controls on an individual level. Similarly, SR7 who scored 94/102 on the CFMT+, outperformed the three SRs with the highest CFMT+ scores (SR1, SR2, and SR5) in Experiments 1 and 2. Our data therefore suggests that while SRs as a group are significantly better at face matching and memory than typical perceivers, there is some variability in how well they perform these applied tasks. One possibility is that individual differences in general visual processing could be underlying those differences in performance on applied tasks. Indeed, there is some evidence that developmental prosopagnosia is often underpinned by various deficits in the domain of perception (e.g. De Renzi, Faglioni, Grossi, & Nichelli, 1991) and the findings reported in Chapter 2 suggest that this is also true for SRs. Future work could consider individual differences in general visual processing and performance on applied tasks of face matching and face recognition. Thus, in forensic and national security settings, it would be beneficial to follow up initial screening (i.e. using tests such as the CFMT+) with tests resembling real life scenarios (such as those involving face matching or memory for faces taken from CCTV footage) in order to ensure optimal personnel allocation.

Taken together, the results from this study show that SRs as a group are sizeably better at applied tasks of face matching and recognition. Research in unfamiliar face processing consistently shows that there are large individual differences in face memory and perception (Bowles et al., 2009; Russell et al., 2009), such that some people experience everyday problems with face recognition (for a review see Susilo & Duchaine, 2013) while others are particularly good at face processing and claim to 'never forget a face' (Russell et al, 2009; Russell et al., 2012). The lack of standardised tests of face recognition in security and forensic settings makes these agencies vulnerable to typical error-prone performance, and these agencies may benefit from the employment of SRs. These individuals could be used to offer expert opinion for work assignments involving CCTV footage, and may be a valuable addition to border control and police personnel.

Chapter 6: Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills

Chapter published as:

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

1. INTRODUCTION

In the previous chapter, evidence was presented suggesting that at least some SRs excel at more applied tasks involving face memory, such as recognising previously studied faces from moving video clips or matching faces in line-up arrays. However, very little work has examined more real-world face perception skills in super recognition, and it is currently unknown whether these individuals also excel at simple instantaneous matching tasks resembling ID-to-person matching situations. Although some authors have investigated the performance of SRs on the Cambridge Face Perception Test (CFPT: Russell, Chatterjee, & Nakayama, 2012; Russell et al., 2009), it should be noted that performance was only examined at the group rather than the individual level (where significant differences are impossible to detect due to the high control mean and large standard deviation associated with this task), and the test has little resemblance to critical real-world tasks that require accurate face matching skills (see Chapter 4 for a detailed discussion).

Investigation of the face matching skills of SRs is an important theoretical issue given individuals at the other end of the face recognition spectrum (i.e. those with prosopagnosia: (Bate & Bennetts, 2014; Bate & Cook, 2012; Bennetts, Butcher, Lander, Udale, & Bate, 2015) can present with or without impairments in face perception (Dalrymple, Garrido, & Duchaine, 2014). Such findings have aided the development of dominant models of face-processing (e.g. Bruce & Young, 1986), by indicating that the face-processing pathway can be lesioned at different locations (i.e. at an early stage involving structural encoding or a later stage involving retrieval); yet the presumed hierarchical nature of the framework explains why the hallmark deficit in facial identity recognition presents even in the former group of individuals. Likewise, it follows that super recognition may result from relatively early enhancements affecting facial identity perception, or from later enhancements affecting memory for faces; and the underpinnings of the skills may vary between individuals. In Chapter 2 this issue was addressed using typical laboratory-based tasks of face perception, and it was found that only some SRs show enhancements in face perception as well as face memory. However, as observed in Chapter 5, performance on such tasks do not necessarily mirror performance on more real-world tasks. Examination of the face perception abilities of SRs in such tasks will therefore have important theoretical implications by presenting a novel means to evaluate current theoretical frameworks, but will also have real-world value by assessing the capabilities of these people (both in a group and at individual level) in tasks that are fundamental to national and international security.

The current study addresses this issue by investigating whether seven SRs are better at matching faces (i.e. face perception) than opportunistically chosen samples of typical perceivers. First, participants completed the Glasgow Face Matching Test (GFMT) (Burton et al., 2010) – a task that has been extensively used in previous research (Bindemann et al., 2012; Megreya, Bindemann, & Havard, 2011; Megreya, White, & Burton, 2011). However, because overall accuracy in the GFMT task is high even in typical perceivers, it is hard to detect significant differences in performance on an individual level. With the aim of exacerbating these disparities, we also used a more demanding face matching task, the Models Face Matching Test (MFMT: Dowsett & Burton, 2015). In this test, images are of models who undergo a change in appearance between images. Thus, this task examined whether any superior face matching skills of SRs are also evident when task demands are high, and efficient extraction of information is required in order to elicit an accurate response. Given that these tests resemble tasks that are typically performed by passport control and other security officers, this investigation is of particular interest to those who perform person-to-ID comparisons in an occupational setting.

2. METHOD

Participants

Seven SRs (four male) participated in this task (see Table 1). As in previously Chapters, their superior face recognition skills were confirmed using the CFMT+ (see Table 1) and the published age-appropriate norms that are available for this test (Russell et al., 2012, 2009). Four of the individuals are previously described in Chapters 2, 3, and 5.

In addition, 20 typical perceivers were recruited from students and visitors at the University to act as controls. All control participants were also screened using the CFMT+, to exclude those meeting the criteria for either super recognition or prosopagnosia (Duchaine & Nakayama, 2006). However, no exclusions were necessary. They were matched to the SR group according to age and gender (10 male). All participants had normal or corrected-to-normal vision, and participated on a voluntary basis or in exchange for course credits. Because of the concerns related to how well student participants are motivated to perform tasks in group studies, an additional gender- and age-matched group was recruited offering a financial incentive. Specifically, this group of 20 individuals were informed that they would receive reimbursement for their time in accordance with the departmental policy (£8 per hour of participation), but they would also gain an extra £1 for every 10% increase in accuracy above 50% (chance performance) on the face matching tasks. In sum, this group was aware that the better they perform the tasks, the more money they would get in addition to their statuary reimbursement. These participants are referred to as "motivated controls" in the remainder of this manuscript. Importantly, all seven SR participants recruited for this study outperformed both control groups on the CFMT+, in both group and case-by-case analyses (see Table 1). The latter were performed using modified t-tests for single case comparisons (John R. Crawford, Garthwaite, & Porter, 2010) (all *p*s < .05). Ethical approval for the study was granted by Bournemouth University's Ethics Committee.

Table 1. Demographical information and CFMT+ scores for the SR participants used in this study and those described by Russell, Chatterjee, and Nakayama (Russell et al., 2012). Standard deviations are in parentheses.

	Controls	Motivated Controls	Russell et al.'s SRs	The current study							
	(N = 20)	(N = 20)	(N = 6)	SRs (N = 7)	SR1	SR2	SR3	SR4	SR5	SR6	SR7
Age	25.2 *(5.6)*	24.2 (5.0)	40.7 *(9.9)*	25.0 *(5.4)*	27	29	20	27	33	20	19
Gender	10 F	10 F	-[1]	3 F	M	M	F	F	M	M	F
CFMT+	71.1 *(10.5)*	71.8 (12.7)	95.0 *(1.9)*	97.7 *(3.2)*	101	97	96	94	100	97	96
CFPT	n/a	n/a	24.7 (10.3)	19.43 *(8.06)*	16	26	10	24	16	12	32

[1]Gender data was not available for Russell et al.'s (Russell et al., 2012) participants

Materials and Procedure

GFMT: The original GFMT (long version) is comprised of 168 pairs of male and female faces: half contain faces of the same identity and half do not (Burton et al., 2010). All images were 350 pixels in width (the images were standardised by width and their height varied naturally) and were displayed in greyscale at a resolution of 72ppi without noticeable jewellery or clothing, but hairstyle was visible. The 84 people used in the matched trials were also paired with a distractor on mismatched trials. The distractors were chosen based on their similarity to the target images, adopting a sorting procedure used by Bruce et al. (1999). All participants were tested individually using E–Prime software (Psychology Software Tools, Sharpsburgh, PA, USA) and a 22 inch LCD monitor displayed at a resolution of 1920 x 1080 pixels. Participants sat approximately 60cm from the screen, and made their responses using the *s* and *k* keys on a keyboard under no time constraints. Each individual saw all 168 pairs, and the trials were presented in a randomised order.

MFMT: The MFMT is comprised of 120 pairs of male faces: 60 matched according to identity and 60 mismatched (for further stimuli details see Dowsett & Burton, 2015). All images measured 300 *(W)* x 420 *(H)* pixels and were displayed in colour to mimic natural settings when face matching would occur. The images did not contain any visible jewellery, but were not occluded of clothing and the hair was not cropped. As for the GFMT, similarity ratings for the mismatched trials were gathered using the method devised by Bruce et al. (1999). Each individual saw all 120 pairs of faces, with the two images matching on half of the trials. The trials were presented in a randomised order. The equipment, procedure and instructions were identical to those used in the GFMT.

Statistical Analyses

For each task a 2 x 3 mixed factorial design was used, with a within–subjects factor of trial type (matched/mismatched) and a between–subjects factor of participant group (controls/motivated controls/SRs). The percentage of hits (correct responses in matched trials), misses (no–match decisions in matched trials), correct rejections (correct "mismatched" responses in mismatched trials), and false positive responses (match decisions in mismatched trials) were calculated for each participant.

While accuracy is a good indicator of the overall patterns of responses between the SR and control groups, additional analyses regarding sensitivity and response criterion permit a more in-depth understanding of the differences between them, and are useful for cross-study comparison (McIntyre, Hancock, Kittler, & Langton, 2013). Specifically, *d prime* (*d'*), a measure of sensitivity, was calculated by subtracting the *z* scores for false–positive (F) responses in the mismatched trials from *z* scores calculated from match responses (hits, H) in matched trials [$d' = z(\mathrm{H}) - z(\mathrm{FPs})$] (see Table 2). Response bias (*criterion c*) was calculated as the negative average sum of *z* scores for the hits and false-positive response: $c = -0.5[z(\mathrm{H}) + z(\mathrm{FPs})]$ (Macmillan & Creelman, 2004).

3. RESULTS

Glasgow Face Matching Test

Accuracy: A 2 x 3 mixed analysis of variance (ANOVA) was conducted on accuracy scores, with trial type as the within–participant factor (matched/mismatched) and group as the between-participant factor (SRs/controls/motivated controls). Means and SDs performance of all three groups are displayed in Table 2. There was a significant main effect of group, $F(2, 44) = 6.12$, $p = .005$, $\eta_p^2 = .218$ and a post hoc Tukey test revealed that SRs performed better than controls ($p = .004$) and motivated controls ($p = .008$), and that there were no overall differences in accuracy between the two control groups ($p = .949$). There was also a main effect of trial type whereby all participants were more accurate on matched trials $F(1, 44) = 7.47$, $p = .009$, $\eta_p^2 = .145$. These main effects were qualified by a significant interaction between participant group and trial type, $F(2, 44) = 5.45$, $p = .008$, $\eta_p^2 = .198$.

Table 2. Group accuracy descriptive statistics in GFMT. Standard deviations are in parentheses.

	Matched trials	Mismatched trials	Total Accuracy
	Hits (%)	Correct Rejection (%)	(%)
SRs	97.02	97.7	97.36
	(3.73)	(1.98)	(2.16)
Controls	91.31	88.45	87.43
	(6.81)	(7.5)	(5.26)
Motivated	96.20	84.55	87.85
controls	*(3.62)*	(10.93)	(5.45)

To investigate the interaction, follow-up analyses were conducted for matched and mismatched trials. For matched trials, a one-way ANOVA revealed significant differences between groups, $F(2, 44) = 5.49$, $p = .007$, $\eta_p^2 = .200$. Planned comparisons showed that SRs were more accurate than controls ($p = .049$) but not motivated controls ($p = .932$), who in turn were more accurate than the standard control participants ($p = .004$). There were also significant between-group differences in mismatched trials, $F(2, 44) = 5.89$, $p = .005$, $\eta_p^2 = .211$. SRs were marginally better than controls ($p = .052$), significantly better than motivated controls ($p = .004$), and there were no significant differences between both control groups ($p = .344$).

Signal detection analyses: A one-way ANOVA on *d* prime scores revealed significant differences between the three groups of participants, $F(2, 44) = 10.91$, $p < .001$, $\eta_p^2 = .332$ (see Table 3). Follow-up analyses revealed that SRs were better at discriminating pairs of faces than controls ($p < .001$) and motivated controls ($p = .001$), and there were no differences in sensitivity between the two control groups ($p = .456$). Furthermore, a one-way ANOVA on criterion *c* scores indicated differences between the three groups of participants, $F(2, 44) = 5.42$, $p = .008$, $\eta_p^2 = .198$. There were no differences in response bias between the SRs and controls ($p = .691$), but the motivated control group was more prone to elicit a "yes" response than either SRs (p = .023) and, interestingly, controls (p = .028). On an individual level, modified *t*-tests for single case comparisons were performed (J. R. Crawford & Howell, 1998) on the *d* prime and criterion *c* data for each of the SRs in comparison to the control groups. This method was specifically developed for single case studies in neuropsychology where, instead of a large set of normative data, individuals are compared to small samples (N < 50) of control participants. Three SR participants (SR1, SR2 and SR 5) showed significantly higher sensitivity than controls on the matching task and, importantly, none of the SRs displayed a different response bias to control participants. In contrast to the motivated controls, the same three SR participants were better at discriminating simultaneously presented faces and one was more prone to reject image pairs as mismatched (SR3).

Table 3. Individual case analyses of sensitivity of SRs in GFMT, using modified *t*-tests for single-case comparisons (Crawford et al., 2010)

	Mean *(SD)*	Single-case comparisons SR1	SR2	SR3	SR4	SR5	SR6	SR7
d prime								
SRs	4.23 *(0.74)*	5.03	5.03	3.50	4.32	4.78	3.65	3.34
Controls	2.82 (0.73)	**.008**	**.008**	.374	.059	**.016**	.281	.495
Motivated Controls	3.08 (0.63)	**.007**	**.007**	.523	.069	**.016**	.388	.692
criterion c								
SRs	0.02 *(0.27)*	0	0	0.51	-0.36	0.13	0.16	0
Controls	-0.11 (0.32)	.741	.741	.074	.455	.473	.420	.741
Controls Motivated	-0.39 (0.38)	.329	.329	**.032**	.939	.197	.174	.329

Models Face Matching Test

Accuracy: A 2 x 3 mixed analysis of variance (ANOVA) was conducted on accuracy scores, with trial type as the within–participant factor (matched/mismatched) and group as the between-participant factor (SRs/motivated controls/controls). Means and SDs performance of all three groups are displayed in Table 4. There was a significant main effect of group, $F(2,44) = 16.40$, $p < .001$, $\eta_p^2 = .427$, and post hoc Tukey tests revealed that SRs performed better than controls ($p < .001$) and motivated controls ($p = .001$), and the difference between the two control groups was marginally significant ($p = .059$). There was also a main effect of trial type whereby all participants were more accurate on mismatched trials, $F(1, 44) = 6.45$, $p = .015$, $\eta_p^2 = .128$. The interaction between these two factors was non-significant $F(2, 44) = 1.99$, $p = .049$, $\eta_p^2 = .083$.

Table 4. Group accuracy descriptive statistics in MFMT Standard deviations are in parentheses.

	Matched trials	Mismatched trials	Total Accuracy
	Hits (%)	Correct Rejection (%)	(%)
SRs	71.90	88.19	82.5
	(14.92)	(5.31)	(6.29)
Controls	63.33	65.58	64.46
	(17.66)	(13.74)	(7.46)
Motivated	67.65	71.95	69.80
controls	(12.24)	(9.79)	(7.19)

Signal detection analyses: A one-way ANOVA on *d* prime scores revealed significant differences between the three groups of participants, $F(2, 44) = 22.79$, $p <.001$, $\eta_p^2 = .509$ (see Table 4). Follow-up analyses indicated that SRs were better at discriminating pairs of faces than controls ($p < .001$) and motivated controls ($p < .001$), and there were no differences in sensitivity between the two control groups ($p = .128$). Furthermore, a one-way ANOVA on criterion *c* scores indicated differences between the three groups of participants, $F(2, 44) = 4.4$, $p = .018$, $\eta_p^2 = .167$. SRs were more likely to report a mismatch than both controls ($p = .019$) and motivated controls ($p = .023$), and there were no differences in bias between the two control groups ($p = .986$). On an individual level, modified *t*-tests for single case comparisons were performed on the *d* prime and criterion *c* data for each of the SRs in comparison to the two control groups (see Table 5). All but one SR (SR3) were better at the MFMT than the control group, and four SRs (SR1, SR2, SR4, SR6) were better at discriminating simultaneously presented faces in comparison to the motivated control group. Two SRs (SR2 and SR7) were also more biased towards classifying image pairs as mismatched than both the control and motivated control groups.

Table 5. Individual case analyses of sensitivity of SRs in MFMT using modified *t*-tests for single-case comparisons (Crawford et al., 2010).

	Mean *(SD)*	Single-case comparisons SR1	SR2	SR3	SR4	SR5	SR6	SR7
d prime								
SRs	2.23 *(0.46)*	3.1	2.38	1.68	2.35	2.11	2.22	1.83
Controls	0.82 (0.46)	**<.001**	**.004**	.083	**.004**	**.013**	**.007**	**.045**
Motivated Controls	1.12 (0.5)	**.001**	**.023**	.288	**.027**	.068	**.045**	.182
criterion c								
SRs	0.48 *(0.4)*	0.58	0.94	0.54	-0.21	0.33	0.27	0.92
Controls	0.03 (0.41)	.203	**.042**	.237	.573	.481	.573	**.046**
Controls Motivated	0.05 (0.3)	.101	**.009**	.127	.401	.373	.483	**.011**

GFMT versus MFMT performance

Because there were no demographic or baseline face recognition ability differences (*p*s < .05) between participants in both control groups, performance (*d* prime) of all control participants on the GFMT task was correlated with performance on the new, more challenging MFMT test to examine the stability of matching judgments over the two tasks. Analyses revealed that performance on the two tests was strongly correlated, $N = 40$, *Spearman's rho* = .719, $p < .001$ (see Figure 1).

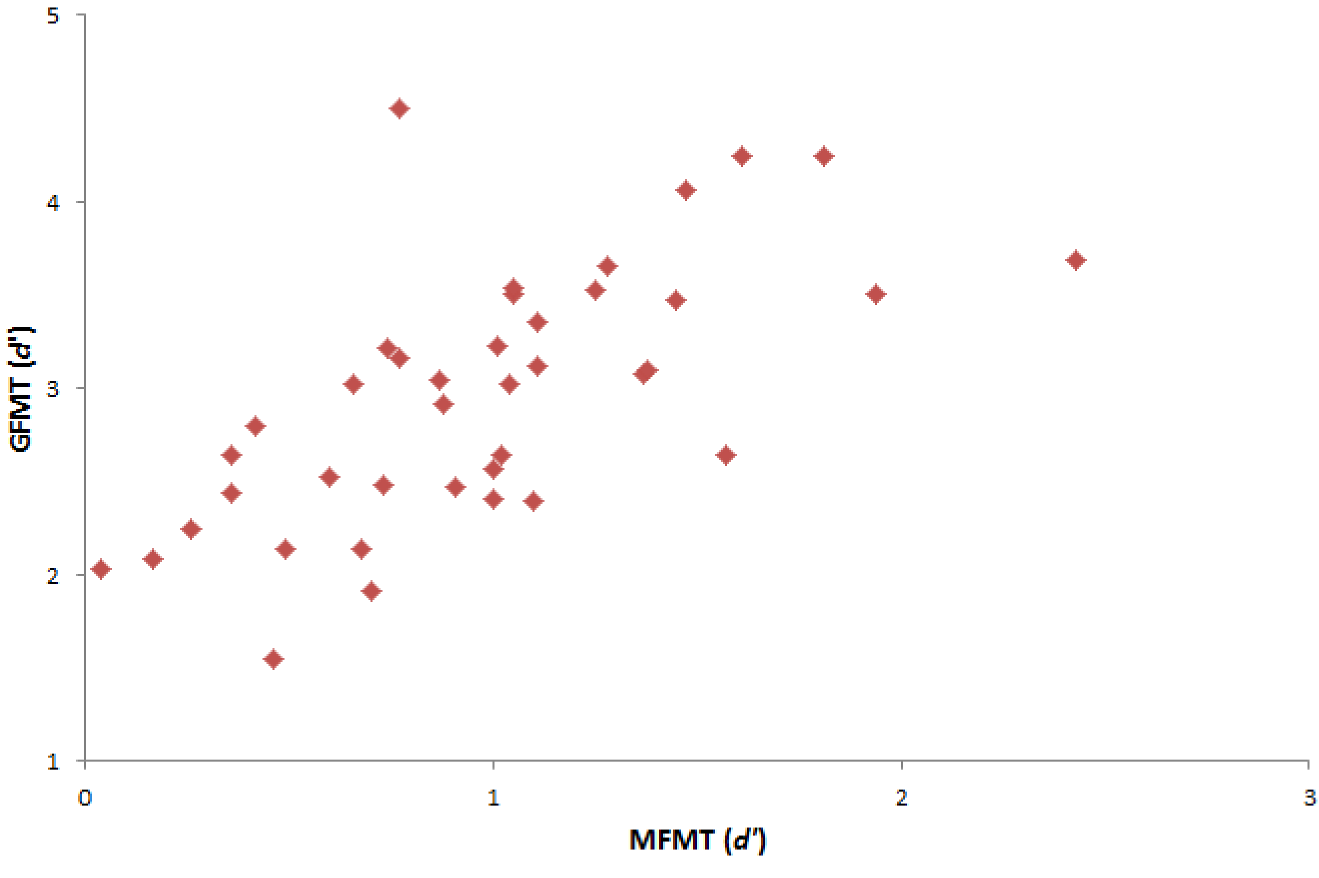

Figure 1. Correlation between GFMT and MFMT performance for all control participants.

4. GENERAL DISCUSSION

In this investigation, the face matching performance of seven SRs was compared to that of two groups of typical perceivers on the GFMT and the more-demanding MFMT. As predicted, group (and most case-by-case) analyses indicated that SRs outperformed controls on both tasks. The SRs also displayed a more conservative response bias (i.e. they were more likely to reject an image pair as mismatched) on the MFMT task at the group and, in two cases, individual level. There was also a strong positive correlation between the GFMT and MFMT scores of all control participants which indicated some stability in face matching performance across the two tasks. These findings have implications for both theory and practice.

In terms of theory, the work presented here is a further testimony to the considerable variation in human face matching ability that has been reported in previous work (Bindemann et al., 2012; Burton et al., 2010; Wilhelm et al., 2010), i.e. while some participants were highly accurate on the matching tasks reported here, others were less so. This range of performance was particularly apparent in the more demanding MFMT (the minimum accuracy was 51% and 53% in control and motivated groups, respectively, while the top performer, SR1, was accurate on 91% of the trials). What is more, the selective enhancement in face matching ability in the SR group provides some insight into the memory-perception dissociation in face processing (Burton et al., 2010; Wilhelm et al. 2010). Specifically, while all the SRs showed enhanced face memory on a standardised laboratory task (the CFMT+), not all of these individuals outperformed controls on the matching tasks. That is, not all of the SRs were "super matchers". This finding converges with previous work by Burton and colleagues (Burton et al., 2010) who reported small correlations between a simple face memory and the GFMT, but a strong correlation between the GFMT and an object memory task. The authors concluded that unfamiliar faces are processed in a similar way to objects. While the two experiments presented here do not have scope to support the latter proposition, they offer some evidence that memory for faces and face perception are not overlapping constructs. Specifically, the relative dissociation between face memory (measured by the CFMT+) and face matching (assessed using the GFMT and MFMT) in the SR group provides some ground for critique of the Bruce and Young's (Bruce & Young, 1986) sequential model of face recognition. The model posits that face perception precedes face recognition and the latter cannot be achieved without the initial perceptual input. Although it is possible that intact, but not enhanced, face perception may be enough for face memory to be enhanced at a later stage, the mechanism behind such augmentation is unclear. The inverse of this effect has been previously observed in the developmental prosopagnosia literature, where all individuals have impaired face memory, but some present with face perception ability on par with that of typical perceivers (Bate & Bennetts, 2014; Dalrymple et al., 2014). A recent study by Bobak and colleagues (Bobak et al., 2016) further speaks to this issue: seven SRs excelled as a group and, mostly on individual level, at face matching in a line-up paradigm, but only one showed exceptional performance on an individual level in a demanding face memory task with a delayed recognition phase (they outperformed the controls in the memory task in group analyses, however).

Without well-defined evidence for direct forward-feeding of perceptual input, the results of this study are better explained by the Interaction Activation and Competition Model (IAC, (Burton, Bruce, & Johnston, 1990)) where pools of units can be activated in parallel. It is possible that this simultaneous activation of perception and memory occurs more strongly in some SRs than others, resulting in enhanced face matching concomitant with excellent recognition memory for faces. In a bid to further clarify this issue, future studies should aim to examine long-term memory for faces in the SR population.

Furthermore, it is also possible that "super matchers" may exist who do not necessarily have superior memory for faces. Indeed, in the current study the range of performance between the control groups and the SRs was somewhat overlapping, i.e. the lowest performing SRs in both the GFMT and MFMT were less accurate than the highest performing control participants. A similar pattern of findings was also reported in Chapter 4, where control participants with "average" face memory displayed a broad range of abilities in their face matching performance. From a theoretical and practical perspective, future work examining individual differences should therefore screen participants on different aspects of face-processing, with a particular emphasis on dissociating face perception and face recognition. One possible way of investigating this phenomenon would be via the use of eye-tracking technology. There is convincing evidence to suggest that eye movements are pivotal in face learning (Henderson, Williams, & Falk, 2005; Sekiguchi, 2011) and it is possible that eye-movement behaviour also underpins skilled face matching. Future studies may wish to investigate this phenomenon in more depth using tasks that are applicable to situations resembling ID-to-person checks.

However, it is important to note that results from two matching tasks described in this report converge with Chapter 4 (Experiment1), where a somewhat inconsistent performance amongst seven SRs was described on another applied task of face perception, the one-in-ten test (Bruce et al., 1999). Pertinently, while all the SRs in Chapter 5 showed enhanced performance on the CFMT+ and outperformed controls in the two experimental tasks on group-based analyses, only four SRs displayed significantly higher performance on a case-by-case basis. The scores of the three remaining participants, although all at least 1SD above the mean of the control group, did not reach statistical significance on a case-by-case basis. The current study presents a broadly similar pattern of findings. While these differences were not pronounced for the SR and control groups on the GFMT, the SRs always outperformed control participants by more than 1SD on in the more challenging MFMT task. The lack of differences on the former test may be accounted for by the high accuracy in the control groups. As it is hard to detect differences on individual level when there is little room for improvement in experimental paradigms, future studies should take heed of our findings and employ carefully constructed paradigms with sufficient accuracy level to allow such enhancements to be detected.

The individual variability reported in this study raises some important questions about conclusions drawn on the performance of groups of experts. In their recent studies, White and colleagues (White, Dunn, et al., 2015; White, Phillips, et al., 2015) argued that forensic analysts are better than control groups (student and "motivated" employees of governmental agencies) at applied matching tasks. However, while the individual data are available, these studies merely presented group statistics and it is unclear whether these group differences may have been driven by particularly exceptional performance of some experts while the accuracy of others may have been on par with the control groups. It is recommended that future studies examining face matching accuracy in applied settings report case-by-case analyses in addition to group performance, taking special care to ensure that the anonymity of participants working in national security settings is protected. Furthermore, the results reported here suggest that motivation to perform well is not a key issue in face matching accuracy. Specifically, monetary payment with an explicit incentive to do well had little effect on the performance of two demographically similar control participant groups. In addition, while the overall accuracy and sensitivity remained at the same level amongst all control participants, motivated controls were more somewhat likely to elicit a "yes" response on all trials. Given that this bias was not replicated in both face matching tasks, it is unclear how this phenomenon may be interpreted. Nonetheless, this finding contributes to the body of literature showing limited effects of enhancement techniques and motivation on face processing ability (Alenezi & Bindemann, 2013; Bate et al., 2014; Kemp et al., 1997).

In terms of practice in the national security settings, given the typically high error rates and limited success of face matching training regimes (Alenezi & Bindemann, 2013; Towler et al., 2014; Woodhead et al., 1979) cf. (Dowsett & Burton, 2015), this study lends further support to the notion that selection of individuals with excellent face matching skills could be a useful recruitment strategy for occupations where ID-to-person matching is a pivotal part of the job. Indeed, SRs identified by self-report and a standardised test of face recognition memory, the CFMT+, were also mostly found to be excellent face matchers. This advantage was particularly evident in the difficult MFMT task, where all but one SR outperformed control participants. It is of note that SRs not only outperformed typical perceivers on both matched and mismatched trials, but they also displayed a more conservative response bias on the more difficult MFMT. This cautious approach to accepting an identity match is a particularly valuable skill in national security settings, and provides further evidence of SRs' utility in these occupations. Accepting a fraudulent ID (akin to eliciting a "yes" response on a mismatched trial) is a potentially costly mistake to make during ID checks and, as such, examination of response bias may also be a valuable procedure during recruitment. While such predisposition may be slowing the processing of applicants on national frontiers, in the wake of calls to tighten national security, this potential deceleration of processing speed is a rather minor expense in comparison to potential threat from individuals carrying fraudulent IDs.

Furthermore, White and colleagues (White et al., reported that performance on the GFMT predicts photo ID-to-person matching performance, in that individuals who are apt at laboratory-based matching tasks are also high performers in real world ID-to-person comparison tasks, and vice versa. However, White et al. produced ID cards from photographs taken only a few days before the testing session. In real life, passports are typically issued for a period of 10 years - a time in which facial appearance can significantly change. These circumstances heighten the task demands faced by passport officers (Megreya et al., 2013). As such, the MFMT used in the current study may be an even better indicator of face matching ability in real–life settings. The stimuli were taken from a database of models where some images are original photographs provided by the models themselves, and some come from later photo shoots where the model has changed appearance (e.g. their hairstyle). These circumstances mimic real-world scenarios to a greater extent than previous laboratory-based face matching tasks. Future studies should examine the predictive value of the MFMT on real-life passport security checks, to establish whether it would be a useful recruitment tool.

In sum, the results presented here show that (1) SRs as a group (and mostly on an individual level) are significantly better than controls at applied tasks of face matching, (2) face matching performance is stable across different tasks, and (3) face matching ability is independent of participants' motivational level. Recent research has repeatedly shown that there are large individual differences in unfamiliar face recognition and facial identity perception (Bowles et al., 2009; Russell et al., 2009), including individuals at one extreme who are particularly skilled at face-processing (SRs) and those at the other extreme who have severe difficulties in facial identity recognition (developmental prosopagnosia). A lack of effective training techniques makes border control agencies vulnerable to costly mistakes in identity judgement, and these agencies would benefit from having SRs in their ranks. While not infallible, SRs are clearly better at face matching, especially when task demands are high.

Chapter 7: Improving face recognition in forensic settings: The effects of oxytocin on face recognition and face memory tasks

1. INTRODUCTION

Chapters 5 and 6 raise the possibility of employing SRs in occupations where excellent face matching or memory skills would benefit national security. While some law enforcement agencies are currently screening their employees in a bid to identify potential SRs, it is likely that smaller agencies will not necessarily have any such individuals within their ranks. This raises an alternative question: Can typical perceivers be trained to become SRs? Or, at least, can their face recognition skills in some way be improved? In Chapter 1, a plethora of literature was reviewed suggesting that typical perceivers vary greatly in their ability to match and recognize unfamiliar faces (e.g. Bindemann, Avetiysan, & Rakow, 2012; Bindemann & Sandford, 2011; Burton, White, McNeil, 2010; Burton, Wilson, Cowan, & Bruce, 1999; Russell, Duchaine, & Nakayama, 2009). This is an important issue given that these skills are fundamental for a variety of tasks implicated in national security (e.g. passport control, CCTV-to-image matching, searches for perpetrators and wanted persons), and the high error rates in both laboratory and real-world implementations of these tasks are striking (e.g. Burton et al., 1999; White, Dunn, Schmid, & Kemp, 2015; White, Kemp, Jenkins, Matheson, & Burton, 2014). Further, there is no evidence to suggest that on-the-job experience with these tasks improves performance, raising the possibility that attempts to train employees in these roles may be fruitless. For instance, in their study with Australian passport officers, White and colleagues (2014) found that years of experience on the job was not correlated to performance on a face matching task resembling passport control.

Such a hypothesis receives some support from the empirical literature. While some studies failed to achieve improvement with feedback (Alenzi & Bindemann, 2013) and shape classification strategy (Towler, White, & Kemp, 2014); others have reported a 2.8% increase in face matching accuracy following feedback (White, Kemp, Jenkins, & Burton, 2014), a 28% improvement in performance on mismatched trials following a motivational incentive (food, Moore & Johnston, 2013), and approximately a 4% improvement following face matching training in pairs (Dowsett & Burton, 2015).

An alternative attempt to improve face-processing skills has adopted a pharmaceutical approach. Instead of trying to train the visuo-perceptual skills of participants, this methodology simply administers an intranasal dose of oxytocin to each individual. Oxytocin is a nonapeptide implicated in social cognition (Heinrichs, von Dawans, & Domes, 2009) and has been found to enhance face processing. For instance, oxytocin has been shown to improve recognition of facial expressions (Van Ijzendorm, Bakermans – Kranenburg, 2012), and more notably the recognition of unfamiliar faces in both typical participants (Rimmele, Hediger, Heinrichs & Klaver, 2009; Savaskan, Ehrhardt, Schulz, Walter & Schächinger, 2007) and individuals with prosopagnosia (Bate et al. 2014). Nonetheless the effects of oxytocin have varied, with some studies finding null or detrimental effects (e.g. Bate et al., 2014; Herzmann, Young, Bird, & Curran, 2012) and some only reporting gains under certain circumstances (e.g. only showing improvements in recognition when faces are presented with specific expressions, Guastella et al., 2008; Saskavan et al., 2008).

Nevertheless, some successful attempts to improve the face-processing skills of typical perceivers using this approach have resulted in relatively large improvements compared to most behavioural attempts. For instance, Rimmele and colleagues (2009) reported a 10% increase in overall recognition of previously unfamiliar faces in a group of men following intranasal inhalation of oxytocin. Similarly, Savaskan et al. (2007) reported a significant improvement of discriminability and a reduction of false alarm rates (responding "familiar" to, in fact, unfamiliar faces) in a group of participants inhaling oxytocin (*d prime* increment of approximately 0.4 thirty minutes post inhalation of the nonapeptide). It is thus important to clarify the potential benefits of using oxytocin in national security settings, initially via laboratory-based paradigms. One such attempt was recently reported by Bate et al. (2014) where participants inhaled a nasal spray with oxytocin or placebo before completing the one-in-ten task (Bruce et al., 1999) - a well-established line-up matching paradigm containing target- present and target-absent trials. While, participants in the oxytocin condition had better accuracy on target-present trials, they were also more prone to make positive identifications in target-absent trials. Bate et al. (2014) suggested that it is possible that the facilitative effect of oxytocin may result from an increased saliency of studied faces, and this effect may become disadvantageous in line-up scenarios. In target absent line-ups, when many faces are presented simultaneously, participants could be mistaking saliency for familiarity and choosing a foil that seems the most similar to the target, even when it is a distractor. This is in contrast to the placebo condition, where without the saliency enhancing effect of oxytocin, participants responded in a more conservative way.

This increased saliency has been also found to be a general drawback of simultaneous line-ups and often, in real life scenarios, sequential line-ups are used to reduce the number of false identifications (Leach, Cutler, & Van Wallendael, 2009). While simultaneous line-ups are generally advantageous when the target is present, in target-absent arrays participants often resort to relative judgements and point to the “next best choice”. However, it is important to note that three recent meta-analytic reviews have arrived at different conclusions with regards to simultaneous and sequential procedures (Clark, 2012; Palmer & Brewer, 2012; Steblay, Dysart, Wells, 2011). Nonetheless, it is possible that in a sequential line-up where only one face is shown at the time, or in a simple matching task, such as the GFMT, oxytocin would have a positive effect on identifications and would not hinder performance in target-absent trials. Such findings would have important consequences for the forensic sector, visa, and immigration agencies where accurate face matching performance is a matter of national security.

The current study set out to examine this issue by investigating the influence of oxytocin on face matching performance in a new version of GFMT, the Models Face Matching Task (MFMT, Dowsett & Burton, 2015) and a face recognition paradigm requiring participants to recognise faces from moving footage.

2. EXPERIMENT 1

2.1. Method

Participants

Forty Caucasian participants (28 female; Mean age = 22.8 years, SD = 4.3) were randomly assigned in a double-blind between-participants procedure to receive either placebo or oxytocin nasal spray. Gender was dispersed equally between the conditions, and there was no significant difference in age between the experimental and placebo groups (oxytocin: Mean: 22.8 years, SD = 4.1; placebo: Mean = 22.8 years, SD = 4.6), $F(1,38) = 0.001$, $p = .971$.

Exclusion criteria were medication (with exception of medical contraceptives), pregnancy, psychiatric or medical illness, personal or family history of epilepsy and substance abuse. Participants were asked to abstain from alcohol, nicotine and caffeine for 24 hours prior to the experiment, and not to consume and food or drink, besides water, for two hours before the testing session. Ethical approval was granted by the Bournemouth University Ethics Committee, and participants received either course credit (psychology students) or were included in a raffle in exchange for their time. Informed consent was obtained from all participants prior to the study. All participants were contacted 24 hours after the study to ensure their well-being. No side effects of the nasal sprays were reported.

Stimuli and Materials

The Matching Task: The face matching task used in this experiment was the MFMT (described in detail in Chapter 6). This comprised of 45 matched and 45 mismatched male face pairs taken from a portfolio website for professional models. The photographs were presented side by side on a computer screen. They were cropped so the external features including the hair were visible, measured 7.5 cm (*W*) x 10.5cm (*H*), and were all presented in colour. On each trial, following a 2 s fixation cross a face pair appeared on the screen and remained there until participants had elicited a response. Accuracy was emphasised to participants and responses were collected using one of two keyboard buttons (S, K). Once a response was made, the experiment proceeded to the next trial. The sequence of trials in the experiment was fully randomised by the experimental software. For matched trials participants could make a correct match response (a “hit”), or a “miss” (i.e. miss – match response when the two images showed in fact the same person). In mismatched trials, the correct responses are referred to as “correct rejections” and incorrect responses as “false – positives”.

The Multidimensional Mood Questionnaire (MMQ): MMQ (Steyer, Schwenkmezger, Notz, & Eid, 1997) was used to measure general affect, to control for the potential mood-altering properties of oxytocin and other non-specific effects of attention and wakefulness. This self-report measure consists of three sub-scales (good-bad, awake-tired, and calm-nervous).

Design and Procedure

A 2 x 2 mixed factorial design was employed. The within-subjects factor was the trial type (matched or mismatched) and the between-subjects factor was spray type (oxytocin or placebo).

All participants initially received a dose of 24 IU of either oxytocin (Syntocinon Spray, Novartis) or placebo (this was identical to the oxytocin spray with the exception of the active substance) spray, equating to three puffs per nostril. Following inhalation, participants sat quietly for 45 minutes before completing the matching task. The resting period was established consistent with previous studies using oxytocin and face processing tasks (e.g. Bate et al., 2014, Herzmann et al., 2012; Rimmele, et al., 2009) and allows sufficient time for the oxytocin levels to plateau (Born et al., 2002; Gossen et al., 2012; Striepens et al., 2013). In line with previous research (Bate et al., 2014a, Bate et al., 2014b) the MMQ was administered at three points in the study (immediately after inhalation, after 45 minutes, and after the matching task had been completed).

2.2. Results

Encoding, accuracy and signal detection theory analyses

Table 1. Performance in oxytocin and control group in Experiment 1. Sensitivity and response bias are calculated from the *z* scores of hits and false-positive identifications from target-absent trials. Standard deviations are in parentheses.

	Hits (%)	Correct Rejections (%)	*d'*	*c*
Oxytocin	68.58 (13.11)	76.33 (14.41)	1.35 (0.59)	0.15 (0.38)
Placebo	65.44 (17.07)	75.89 (10.75)	1.16 (0.51)	0.16 (0.24)

First, the time taken to encode the faces was analysed. A 2 (spray: oxytocin, placebo) x 2 (trial type: matched, mismatched) mixed design analysis of variance (ANOVA) revealed no significant differences between the oxytocin and placebo groups, $F(1,38) = 0.01$, $p = .922$, $\eta_p^2 < .001$; no difference in the length of time taken to encode matched and mismatched pairs, $F(1,38) = 0.01$, $p = .909$, $\eta_p^2 < .001$, and the interaction between the type of spray and the trial type was non-significant, $F(1,38) = 0.001$, $p = .982$, $\eta_p^2 < .001$ (see Table 1). Second, only the correct responses from the matching task (i.e.

the hits and correct rejections) were analysed. Specifically, a 2 (spray: oxytocin, placebo) x 2 (trial type: matched, mismatched) mixed design ANOVA showed no main effect of spray $F(1,38) = 0.52$, $p = .476$, $\eta_p^2 = .013$, but a significant main effect of trial type, $F(1,38) = 8.86$, $p = .005$, $\eta_p^2 = .189$, with higher overall accuracy for mismatched than matched trials . However, the interaction between the spray type and trial type was non-significant, $F(1,38) = 0.23$, $p = .636$, $\eta_p^2 = .006$.

Third, sensitivity (d' prime) and response bias (criterion c) was analysed to examine whether oxytocin changed overall performance, or induced response bias amongst participants. While accuracy is a very good indicator of performance, sensitivity and response bias are useful for a more in-depth understanding of results and allow a better cross-study comparison. D prime (*d'*), a measure of sensitivity, was calculated by subtracting the *z* scores for false–positive (F) responses in the target–absent trials from *z* scores calculated from correct identifications (hits, H) in target-present trials [$d' = z(\text{H}) - z(\text{F})$] (see Table 1). Response bias (*criterion c*) was calculated as the negative average sum of *z* scores for the hits and false-positive response $c = -0.5[z(\text{H}) + z(\text{FPs})]$ (MacMillan & Creelman, 2005). An independent-samples *t*-test on *d'* scores indicated that there were no differences in sensitivity between the oxytocin and placebo groups, $t(38) = 1.05$, $p = .299$, $d = 0.34$ (see Table 1). Additionally, the equivalent analysis of response bias (criterion *C*) yielded no significant differences between groups, $t(38) = -0.82$, $p = .935$, $d = 0.27$. This pattern of results indicates that oxytocin does not improve face matching performance and does not influence the response bias.

MMQ

A mixed multivariate analysis of variance (MANOVA) was conducted on the three sub-scales (Good-Bad, Awake-Tired, Calm-Nervous) with one between-subject factor: spray type (oxytocin or placebo), and one within-subject factor: time (time intervals one, two and three). Due to experimenter error, mood data for two participants was lost. Multivariate analysis on the remaining participants found no significant effect of spray type or time on the sub-scale scores, $F(3, 34)= 0.13$, $p= .942$, $\eta p^2= .011$ and $F(6, 31)=0.955$, $p=.471$, $\eta p^2= .156$ respectively. Furthermore, no significant interaction was observed between spray type and time, $F(6, 31)=1.22$, $p=.332$, $\eta p^2=.191$. These findings suggest that the main analyses cannot be explained by alterations in mood, attention and wakefulness post-oxytocin inhalation.

2.3. Summary

Experiment 1 investigated whether intranasal administration of the hormone oxytocin can increase participants' accuracy in a standardised face matching test, the MFMT. Previous reports suggest that simultaneous matching tasks are not only challenging when photographs are taken some time apart (Megreya, Sandford, & Burton, 2013), but also pose a significant challenge when images are taken on the same day when less change in appearance has occurred (Burton et al., 2010). It was predicted that oxytocin may either improve the performance of individuals and/or, in line with the pattern of findings reported by Bate et al. (2014), shift the response bias towards a "match" response. Findings from this experiment do not provide support for these hypotheses in that there were no differences in encoding time, accuracy, sensitivity, or response bias between the experimental and placebo groups. In sum, the results of this study suggest that oxytocin does not enhance face matching, or change participants' response bias.

3. EXPERIMENT 2

Although numerous studies investigated the effect of oxytocin using static facial stimuli (e.g. Guastella et al, 2008; Savaskan et al., 2008), there are no reports investigating the effect of oxytocin on face recognition from moving footage. Experiment 2 addressed this issue using the paradigm from Chapter 5, where participants first studied twenty faces and were later asked to recognise them amongst distractors in short video clips. This investigation is of particular interest due to its applied value. Specifically, the oxytocin spray was administered after the faces were studied in order to enhance recognition, rather than encoding. If successful, oxytocin could be administered prior to reviewing of CCTV footage in order to facilitate recognition of persons of interest.

3.1. Method

Participants

Forty four Caucasian participants (30 female; Mean age = 20.6 years, SD = 1.4) were randomly assigned in a double-blind between-participants procedure to receive either placebo or oxytocin nasal spray. Gender was divided equally between the conditions, and there was no significant difference in age between the experimental and placebo groups (oxytocin: Mean: 20.73 years, SD = 1.8; placebo: Mean = 20.41 years, SD = 0.96), $F(1,42) = 0.541$, $p = .466$.

Exclusion criteria were identical to those in Experiment 1, and no participant was excluded from the experiment. Similarly to Experiment 1, participants were asked to abstain from alcohol, nicotine and caffeine 24 hours prior to the experiment, and not to consume and food or drink, besides water, for two hours before the testing session. Ethical approval was granted by the Bournemouth University Ethics Committee, and all participants received course credit in exchange for their time. Informed consent was obtained from all participants prior to the study. All participants were contacted 24 hours after the study to ensure their well-being, and no side-effects of the nasal sprays were reported.

Stimuli and Materials

Stimuli in the face recognition paradigm were identical to that in Chapter 5 (Experiment 2) with the exception of the letter search filler task which was not administered in this study. Instead, participants sat through a 45 minute break between the encoding and recognition phase. Similarly to Experiment 1, participants completed the MMQ (Steyer et al., 1997) at four time points. This was due to a more complex protocol requiring cognitive effort to first remember and then recognise faces from moving footage.

Design and Procedure

A 2 x 2 mixed factorial design was employed. The within-subjects factor was the trial type (target present or target absent) and the between subject factor was spray type (oxytocin or placebo).

All participants encoded 20 faces (10 female) initially followed by a dose of 24 IU of either oxytocin (Syntocinon Spray, Novartis) or placebo (this was identical to the oxytocin spray with the exception of the active substance) spray equating to three puffs per nostril. Following inhalation, participants sat quietly for 45 minutes before completing the recognition task. In the test phase, volunteers viewed 40 video clips (each clip was 5 seconds long) of people walking down a dimly lit corridor. Twenty clips contained previously studied faces and 20 were age, gender and hairstyle matched distractors.

In line with previous research (e.g. Bate et al., 2014a, Bate et al., 2014b) and similarly to the procedure employed in Experiment 1, the MMQ was administered at three points in the study (immediately after inhalation, after 45 minutes, and after the matching task had been completed).

3.2. Results

Accuracy

A 2 (clip type: target-present, target-absent) x 2 (participant group: oxytocin, placebo) mixed factorial ANOVA was conducted on accuracy scores. A significant main effect of clip type indicated that participants in both groups were more accurate when targets were absent (M = 61.00%, SE =.019) than when targets were present (M = 53.00%, SE = .016), $F(1,42) = 7.329$, $p = .010$, $\eta_p^2 = .149$ (see Table 2). Neither the main effect of group (i.e. oxytocin versus placebo) nor the interaction with participant group reached significance, $F(1,42) = .003$, $p = .956$, $\eta_p^2 = .001$ and $F(1,42) = 1.676$, $p = .202$, $\eta_p^2 = .038$, respectively.

Table 2. Performance in oxytocin and control group in Experiment 2. Sensitivity and response bias are calculated from the *z* scores of hits and false-positive identifications from target-absent trials. Standard deviations are in parentheses.

	Hits (%)	Correct Rejections (%)	*d'*	*c*
Oxytocin	51.36 (9.41)	62.95 (13.06)	0.39 (0.41)	0.16 (0.24)
Placebo	55.23 (12.01)	59.32 (12.81)	0.38 (0.34)	0.05 (0.27)

Signal detection analyses

An independent-samples *t*-test on *d'* scores indicated that there were no differences in sensitivity between the oxytocin and placebo groups, $t(42) = 0.07$, $p = .941$, $d = 0.02$ (see Table 2). Additionally, the equivalent analysis of response bias (criterion *C*) yielded no significant differences between groups, $t(42) = 1.36$, $p = .180$, $d = 0.42$. Taken together, signal detection analyses further indicate that intranasal inhalation of oxytocin has no effect on discriminating facial identity from poor quality video clips.

Confidence

Mean confidence scores for hits and correct rejections were analysed using a 2 x 2 ANOVA with one between-participants (group: oxytocin, placebo) and one within-participants (clip type: target-present, target-absent) factor. No significant main effect of spray type on confidence rating was found, $F(1,42) = .003$, $p = .957$, $\eta_p^2 = .001$. However, a significant within-group difference was observed between the target present and target absent trials for confidence rating $F(1,42) = 24.264$, $p <.001$, $\eta_p^2 = .366$. Results indicated that participants were significantly more confident in target present trials than in target absent trials (target present: $M = 3.54$, $SE = .090$, target absent: $M = 3.11$, $SE = .086$). No significant interaction was observed between this effect and spray type, $F(1,42) = .848$, $p = .362$, $\eta_p^2 = .020$.

Reaction Time

A final mixed 2 x 2 multifactorial ANOVA indicated that there was no significant main effect of spray-type on median reaction time (ms), $F(1,42)= .749$, $p = .392$, $\eta_p^2 = .018$. However, a significant within-group difference between target present and absent trials was detected $F\ (1,42) = 4.569$, $p = .038$, $\eta_p^2= .098$. Results suggest that participants were significantly quicker to respond in the target present condition than in the target absent trials (target present: $M= 4551.65$, $SE= 221.53$, target absent: $M= 4752.57$, $SE= 228.27$). However, this effect did not interact with spray type, $F(1,42) = .074$, $p = .787$, $\eta_p^2 = .002$.

MMQ

A mixed multivariate analysis of variance (MANOVA) was conducted on the three sub-scales (Good-Bad, Awake-Tired, Calm-Nervous) with one between-subject factor: spray type (oxytocin or placebo), and one within-subject factor: time (time intervals one, two, three and four). Multivariate analysis found no significant effect of spray type or time on the sub-scale scores, $F(3, 40)= 1.333$, $p= .277$, $\eta p^2= .091$ and $F(9, 34)=1.865$, $p=.092$, $\eta p^2= .331$. Furthermore, no significant interaction was observed between spray type and time, $F(9, 34)=1.166$, $p=.347$, $\eta p^2=.236$. These findings suggest that the results in the main analysis cannot be attributed to alterations in mood, attention and wakefulness post-oxytocin inhalation.

3.3. Summary

Experiment 2 investigated whether intranasal administration of the hormone oxytocin can increase participants' accuracy in recognition of previously studied faces from video footage. One previous study using a similar paradigm (Burton et al., 1999) found that typical perceivers tend to perform at chance level and that trained Police officers unselected for their face recognition ability are not better at this task than lay people. It was predicted that oxytocin may either improve the face recognition memory or, in line with findings reported by Bate et al. (2014a), shift the response bias towards a "yes" response. Nonetheless, data from this experiment do not provide support for these hypotheses in that there were no differences in RT, accuracy, sensitivity, or response bias between the placebo and experimental groups. In sum, the results of this study suggest that oxytocin does not enhance face recognition memory, or change participants' response bias.

4. GENERAL DISCUSSION

Previous research has identified face matching (e.g. Burton et al., 2010) and recognition faces from moving footage (Burton et al., 1990) as a highly error prone task, irrespective of whether is performed by lay persons or trained police or passport officers (Burton et al., 1999; White et al., 2014; but see Robertson et al., 2016). This investigation set out to examine the performance of typical observers following inhalation of oxytocin in a face matching and a face recognition task. Experiment 1 investigated the effects of oxytocin on the MFMT In Experiment 2, participants' memory for faces encoded from high quality CCTV stills that were later presented in poor quality video clips was tested. Experiment 2 also measured participants' confidence in their responses to assess whether the nonapeptide may produce a shift in their self-perceived accuracy. Finally, encoding time (Experiment 1) and RT (Experiment 2) were measured to scrutinise any differences in participants' analytical approach. The data from these two experiments stood in contrast to the predictions made based on previous research. Specifically, there were no differences between individuals' accuracy, encoding and reaction time, and confidence on the MFMT, or the face recognition task. Additionally, participants did not exhibit a more liberal response bias following the inhalation of oxytocin.

In contrast to previous findings reporting that participants who inhale oxytocin exhibit enhanced activation in the fusiform gyrus, a brain region within the core face processing network (Domes et al. 2007; Kirsch et al., 2005; Labuschagne et al. 2010; Petrovic, Kalisch, Singer, and Dolan, 2008) and enhancements in face recognition accuracy (Rimmele et al., 2009; Savaskan et al., 2008), this study failed to find evidence of facilitation following oxytocin administration on face matching and face recognition memory. These results could not be attributed to the changes in mood, wakefulness or attention across the task.

One possible explanation is that oxytocin specifically increases attention to the eye region in static face images and this information may be difficult to access in moving footage such as that in Experiment 2. Pertinently, Guastella, Mitchell and Dadds (2008) found that oxytocin increased the number and duration of fixations to the eye region of high quality face photographs. It is possible that, in Experiment 2, due to poor lighting, the relatively small size of the faces in the footage, and movement of the faces in the video footage, this facilitation is not useful in poor quality moving footage. Nonetheless, this explanation does not account for the null findings of Experiment 1, where faces were presented in full frontal, or three-quarter view and the eye region was always visible. It is also possible, given the findings from Chapter 3 highlighting the importance of the nose region in face learning and free viewing of faces in a social context, that the facilitation of attention to the eye region following the inhalation of oxytocin did not result in face matching, or face recognition improvements in both experiments.

Another possibility is that participants in this study did not engage in the tasks. Indeed, in Experiment 2, the overall accuracy was lower than that previously reported for the control participants in this task (see Experiment 2, Chapter 5). However, the same was not true for Experiment 1, where a different group of participants performed roughly in line with previously reported means for the MFMT (Dowsett et al., 2015; Chapter 6 of this thesis, Robertson et al., 2016). Given that both studies recruited participants from the student population at Bournemouth University, it is unlikely that the differences in performance in the memory task can be explained by the motivation level of participants. It is of note that participants in Experiment 2 did not proceed to the recognition phase immediately, but there was a 45 minute waiting period between the study and test (in contrast to the 20 minute filler task and break in Experiment 2 of chapter 5). This waiting period may have contributed to the low performance in this task. It is also not improbable, given the low accuracy rate in general, that the difficulty of this test was not calibrated properly for any significant differences to be detected. Indeed, the currently dominant test of unfamiliar face recognition memory, the CFMT (Duchaine & Nakayama, 2006) requires participants to memorise six faces. Private consultations of the author of this thesis with the local Police Force indicate that, typically, police officers are presented with six to ten missing or wanted identities in their briefings and it is possible that there is a capacity on the number of unfamiliar faces one can simultaneously memorise and later recognise. Future studies may wish to investigate the effect of oxytocin on face recognition memory using a simplified version of this paradigm (e.g. by reducing the number of faces in the encoding stage).

Alternatively, it is possible that the enhancements following oxytocin inhalation are merely applicable to tasks where there is an additional incentive for participants to familiarise themselves with the faces. For instance, Rimmele and colleagues (2009) asked participants to rate faces from 1-7 according to how likely they would be to approach that individual. Future studies may wish to investigate the effects of oxytocin using additional tasks allowing participants to evaluate the viewed faces in more depth. However, while this strategy may be useful in applied settings resembling the matching task in Experiment 1, it would be of limited utility when attempting to recognise faces from video footage or individuals passed on the street. Furthermore, in the instructions for the encoding phase of Experiment 2, participants were asked to imagine that the faces they studied were of people who were missing or wanted, and were told explicitly that it is important that they remember them for later recognition. Such instructions are arguably more likely to occur in real-life scenarios involving policing and border control than the subjective evaluations used in the Rimmele et al. (2009) study.

In addition, oxytocin did not shift the response bias of participants, in contrast to the results reported by Bate and colleagues (2014a). The latter study found that, following the same oxytocin inhalation protocol as in the current study, participants were more likely to elicit a target present response in a 1-in-10 face matching task. Interestingly, in both experiments of the current study, all participants were more accurate in target-absent or mismatched trials, in that they were better at rejecting a pair as mismatched and classify a person as unseen than to accept a match or make a “familiar” response in the memory task. This is most likely due to the methodological differences between the Bate et al. (2014a) study and the current project. Pertinently, false-positive errors in simultaneous line-up paradigms are well-documented in research (Lindsay et al., 1991; Lindsay & Wells, 1985; Sporer, 1993, Steblay, Dysart, Fulero, & Lindsay, 2001, Steblay, Dysart & Wells, 2011) and it is possible that used in a simultaneous paradigm, oxytocin further increased the social saliency of the faces prompting participants to make a “target-present” response. The MFMT employed in this study shows only two faces per trial and it is thus unlikely that such saliency effect would be applicable to this paradigm. Experiment 2, on the other hand, utilised a sequential presentation, a practice where response bias towards identifying non-studied faces are unlikely to arise (see Valentine & Fitzgerald, 2016 for a most recent review of simultaneous and sequential line-up practices).

One possible explanation for the higher accuracy in the mismatched trials in the MFMT and the target absent trials in Experiment 2 is that the faces used in mismatched trials were more distinctive that these used in trials where participants had to elicit a negative response. Pertinently, previous research showed that distinctive faces are not only easier to remember, but are also more likely to be rejected as unseen in memory paradigms (Going & Read, 1974, Valentine & Bruce, 1986). Future research should aim to standardise the stimuli across conditions to account for the pre-rated distinctiveness of faces.

Additionally, Experiment 2 failed to find any evidence for oxytocin-induced changes in confidence levels between participants. Nonetheless, all volunteers were more confident following their responses in the target present trials, which, although consistent with previous research (Bruce et al.,1999), stands in contrast to the accuracy levels whereby participants were more accurate in target-absent trials. This result further demonstrates the rather limited relationship between confidence and accuracy (Deffenbaucher, 1980).

Finally, there were no differences in the encoding and reaction times between oxytocin and placebo groups in the matching and face recognition tasks and it appears that the nonapeptide has no influence over the time of decision making regardless of trial type. Interestingly, in Experiment 2 participants took longer to reject previously unseen faces than to respond to faces as “familiar”. This difference may reflect the strategies adopted by participants and is in line with the confidence levels for target-present trials in this experiment. Specifically, individuals may have felt that they need to see the full clip to make sure that the person present in the video was unfamiliar. Experiment 1 did not impose time limits on decision making time and the images were static, whereas in Experiment 2 the longer the participant viewed the clips the closer the people in them were to the camera.

Collectively, the results of this study show limited application of oxytocin for enhancement of face matching and face recognition ability in applied tasks resembling real-life scenarios. Indeed, it is possible that the positive effects of the nonapeptide are restricted to emotional faces (Savaskan et al., 2008; Guastella et al., 2008) and individuals with deficits in face recognition (Bate et al., 2014b). All faces used in Experiments 1 and 2 displayed neutral facial expressions. It is pertinent that extreme affect is rarely displayed in real life situations such as passport checks or in one's everyday activities that may be recorded by CCTV, and it is thus clear that administration of oxytocin is not useful for enhancing performance on those tasks.

Taken together, this study adds to the mounting evidence that administration of oxytocin is not entirely beneficial for social cognition (Bartz, Zaki, Bolger, Ochsner,2011). Although the nonapeptide may improve face recognition in the instance of emotional faces or in those affected by prosopagnosia, there is no benefit of oxytocin in a face-matching task resembling passport checks and face recognition from video footage. Previous attempts at improving face matching ability have yielded mixed results (White et al., 2014; Alenzi & Bindemann, 2013, Moore & Johnston, 2013) and this study adds to the body of research showing that enhancing one's natural face recognition ability is a difficult task. Future studies should aim to clarify the circumstances under which oxytocin enhances face recognition ability, both by employing various experimental paradigms and controlling for individual differences in the face processing ability of participants. This is an important avenue of research, given that oxytocin has been recently administered in those with profound face recognition difficulties (Bate et al., 2014) and has shown promising improvements in this population. Nonetheless, the data presented here suggests that oxytocin should not be used to enhance face matching and face recognition ability in applied forensic settings.

Chapter 8: General Discussion

1. THE NEUROPSYCHOLOGICAL "DIAGNOSIS" OF SUPER RECOGNITION

1.1. Use of the CFMT+

The Cambridge Face Memory Test Long Form (CFMT+: Russell, Duchaine, & Nakayama, 2009) was born from the developmental prosopagnosia literature (Duchaine & Nakayama, 2006) and has been since employed in a number of investigations using Super-Recognisers (SRs, Chapters 2 to 6 of this thesis, Russell et al., 2009, 2012). Unlike the CFMT, it does not suffer from ceiling effects and has an added difficult section aimed at differentiating SRs from typical perceivers.

However, it remains an open question whether the CFMT+'s highly constrained format and lack of ecological validity make it the best candidate to assess individuals' face recognition ability. Indeed, Chapter 5 clearly showed that the CFMT is of limited utility when assessing the face recognition ability of typical perceivers. In a seminal review paper, Burton (2013) argued that the use of highly constrained stimuli and the conflation of image and person recognition are some of the reasons why face recognition research has progressed rather slowly. Pertinently, tests of face recognition using photographs taken in controlled conditions (i.e. with the same cameras, lighting, and at specific times) greatly obscure the within-person variability that is typically encountered in real life. To avoid the criticism that such tests merely assess image recognition, manipulations in viewpoint or expression are often used to ensure the task actually measures face recognition skills. Nonetheless, given the images are captured on cameras with consistent settings, they likely still share many pictorial properties. The CFMT is an example of such a test, and it is thus unsurprising that its correlations with more applied tests of face recognition are rather unconvincing.

Furthermore, the study reported in Chapter 4 further highlighted the need for country specific norms and cut-offs, an issue previously reported by Bowles and colleagues (2009). The male identities used in the CFMT (and the CFMT+) are of Mediterranean appearance and are not an accurate ethnic match to populations inhabiting Northern Europe.

Finally, this thesis is the first to report significant gender differences in the CFMT+, with women outperforming men on the recognition of male faces. Previous reports have shown a female advantage, but only for recognition of female faces (Lewin & Herlitz, 2002, McKelvie et al., 1993) possibly due to a strong sense of belonging and "own group bias" (Bernstein, Young, & Hugenberg, 2007). Nonetheless, this thesis seems to contradict these findings and provides preliminary evidence that females may be, in general, better at face recognition than their male counterparts (for a detailed discussion of this finding see section 4.3 in Chapter 4).

Taken together, the CFMT+ appears to be of limited utility in assessing the true extent of face recognition skills at the top end of the spectrum, possibly due to its rather constrained and highly controlled format. While the CFMT and CFMT+ have become the "gold standard" in neuropsychological research, there is a clear need for a new, ecologically valid test of face recognition ability using natural images of whole faces, taken on numerous occasions in order to mimic the natural variability occurring within people's faces over time. It is only when such a new assessment tool is developed that researchers will be able to test the true extent of human face recognition ability.

1.2. Use of the CFPT

Similarly to the CFMT+, the CFPT emerged from research with prosopagnosia participants (e.g. Avidan, Tanzer, Behrman, 2009; Bowles et al., 2009; Duchaine, Germine, & Nakayama, 2007) and has been adopted to assess typical and superior face recognition (Russell et al., 2009; 2012). Pertinently, Russell and colleagues (2009) reported that SRs outperform typical perceivers on this test of face perception and, as such, the CFPT has been adopted in this thesis (Chapter 2) to further asses the perceptual underpinnings of super-recognition.

However, the CFPT suffers from a number of shortcomings that make it unsuitable for studies involving experts in face recognition. Firstly, the large standard deviation in the performance of individuals unselected for face recognition ability (see Chapter 4) makes single case comparisons difficult, if not impossible. Specifically, a performance of 2 SDs below the control mean (the CFPT records number of errors away from a perfect configuration, i.e. a lower score signifies fewer mistakes) would necessitate near-perfect performance, which is largely impossible due to the complex scoring system employed in the test. This intricate design of the CFPT makes it impossible for single case comparisons to achieve significance and underlies the inconsistency between Russell et al.'s (2009) study and the results reported in this thesis. Specifically, Russell and colleagues employed group-level statistics to assess the face perception skills of four SRs, while this thesis (Chapter 2) adopts a more conservative case-by-case approach (see the next section for an in-depth discussion of case-by-case comparisons).

Furthermore, the CFPT, similarly to the CMFT+, uses constrained and tightly controlled stimuli. A paradigm requiring participants to sort the similarity of morphed faces is also unlikely to mimic real-life situations where face perception is used to access identity. As such, it is imperative for future research to develop new tests that assess face perception ability within a more ecologically valid paradigm.

Dowsett and Burton (2015) have recently devised one such test. These authors reported results from the Models Face Matching Test (MFMT), mimicking the structure of the seminal Glasgow Face Matching Test, (GFMT, Burton, White, & McNeill, 2010), but with naturalistic stimuli where faces vary according to factors such as pose, appearance, hair style and expression. The MFMT has since been used to assess the face matching ability of civilian SRs (Chapter 6) and the London Metropolitan Police SRs (Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016). Future research should continue to develop new comprehensive assessments of face perception skills, manipulating additional dimensions such as ethnicity and aging.

1.3. Importance of case-by-case statistical comparisons

The conservative case-by-case statistical approach (Crawford & Howell, 1998) that is frequently used in the prosopagnosia literature (e.g. Bate, Haslam, Tree, & Hodgson, 2008; Lee, Duchaine, Wilson, & Nakayama, 2009) is a particular strength of the work presented in this thesis. Pertinently, previous studies with SRs (Russell et al, 2009) and trained experts of face recognition (e.g. White, Dunn, Schmid, & Kemp, 2015; White, Phillips, Hahn, Hill, & O'Toole, 2015) merely reported group-level analyses and it is possible that their findings were exaggerated by the strongest performing individuals – particularly in light of the small sample sizes recruited in these investigations. Notably, while the neuropsychological approach of Chapter 2 did not allow for the reporting of group-level analyses (they were, in fact, strongly discouraged by the reviewers of the published version of Chapter 2), studies reported in Chapters 5 and 6 clearly showed that while the SRs were significantly better at face matching and face recognition as a group, the same was not always true on individual level. Such results inform both psychological theory and practice. In terms of theory, they show that face memory and face perception are, at least to some extent, dissociable. In terms of practice, individuals with extraordinary face processing skills do seem to be valuable personnel for national security and forensic agencies. However, it is important to ensure that the performance of each individual is superior to that of age-matched controls. Facial image comparison assignments such as passport control are jobs that are performed individually and, as such, a group analysis of Border Force recruits would not be informative. That is, it would not allow identification of the highest and lowest performing individuals.

Furthermore, given the heterogeneity of the cognitive and perceptual presentation of SRs that is evident throughout this thesis, it is vital that future studies supplement group statistics with individual analyses. A detailed investigation of Russell et al.'s study (2009) further speaks to this issue, where significant enhancements in face perception and a face inversion effect (a hallmark of typical face processing: Maurer, Le Grand & Mondloch, 2002) were observed for the SR participants. However, these data were only reported using group statistics and it is unclear whether all these participants displayed a significantly greater inversion effect individually. To illustrate this point, we have analysed the inversion effect data from Chapter 2 using group statistics: the SRs as a group showed an enhanced inversion effect on the matching task: $t(25) = 2.70$, $p = .012$. The same pattern was not marked in all case-by-case analyses, and a considerable heterogeneity in performance of the SR participants was evident throughout the remaining chapters of this thesis. Studies examining forensic experts have also typically only reported group analyses (Norell et al., 2014; White et al., 2015b) and what this thesis clearly shows (specifically the findings of Chapters 5 and 6) is that the results of these comparisons can be driven by the best-performing participants. It is thus pivotal for future investigations to examine individual performance in order to gain a full understanding of discrete abilities in unfamiliar face recognition. While at first glance this inconsistency in findings may be perceived as weakness of this work, it is this very point that is important in studies reported in this thesis.

1.4. Identifying cognitive subtypes

Studies reported in this thesis show clear evidence that face perception and face recognition are, at least to some extent, dissociable. All six SR cases examined in Chapter 2 displayed enhancements at the level of facial identity recognition (the CFMT+), but only two participants showed facilitation at the level of facial identity perception (face matching task). However, the same pattern of findings did not always emerge in studies in Chapters 5 and 6. Specifically, SR participants recruited following self-report and screening with the CFMT+ did not always excel at the well-established one-in-ten task (Bruce et al., 1999) and the MFMT (Dowsett & Burton, 2015).

This pattern of findings broadly mirrors the prosopagnosia literature where the deficit can be divided into two subtypes: one affecting face perception and other resulting in impairment of higher-level mnemonic processes (de Renzi et al, 1997). Pertinently, given that some (but not all) SRs show facilitation at the level of face perception, it is possible that some individuals are “super-matchers” and do show a similar enhancement in face memory. In terms of theory, given the lack of direct evidence for forward feeding of perceptual information, the results of this thesis are best explained by the Interaction Activation and Competition Model (IAC, Burton, Bruce, & Johnston, 1990) where in contrast to the sequential model of face recognition (Bruce & Young, 1986), processes involved in face recognition are activated in parallel. It is thus possible that the concurrent activation of face recognition and face perception ensues more strongly in some SRs than others, resulting in only excellent face matching skills in these individuals.

This interpretation of results is further strengthened by the fact that the reverse pattern may also exist. Specifically, in Chapters 5 and 6, control and SR performance on the matching tasks were somewhat overlapping, in that the highest performing controls were more accurate than the least accurate SRs. It is thus possible that “super-matchers” exist who do not have superior memory for faces but are excellent at facial identity perception, providing limited evidence for the sequential structure of the face processing system.

Future studies should take heed of these findings and further investigate the dissociation between face recognition and face perception. From an applied perspective, in industries where superior face recognition skills are pivotal in every-day jobs, comprehensive screening of candidates’ mnemonic and perceptual facets of face processing is recommended for optimal personnel allocation to tasks that best suit individual cognitive profiles.

2. THEORETICAL PERSPECTIVES

2.1. The domain-specificity of super recognition

Several lines of evidence suggest that faces are processed by domain-specific mechanisms. For instance, unlike object processing, face processing is sensitive to the effect of inversion (Yin, 1969) and facial features are better recognised in the context of the whole face, rather than alone (Tanaka & Farah, 1993). Furthermore, cases of patients such as CK (Bodamer, 1947) show that face recognition impairment does not imply an concomitant deficit in object processing. However, there is also a substantial body of research showing that faces are processed similarly to objects of expertise. For instance, Diamond and Carey (1986) showed that dog experts process dog bodies in a similar manner to faces. The domain general account is further bolstered by evidence of patients with acquired prosopagnosia showing concomitant deficits in object processing (Damasio, Damasio, & Van Hoesen, 1982; De Renzi, 1986).

In order to address the issue of domain specificity in face recognition, tests employed in Chapter 2 examined (a) the matching of faces compared to houses and hands, and (b) memory for cars (the CCMT) using a paradigm that replicates the standard version of the CFMT. Only one SR (GK) displayed enhanced matching of inverted hands, but did not excel at the car matching task. Another SR, TP, displayed a trend towards enhanced recognition of car stimuli, a result likely underlined by generally superior memory. Given the other SRs showed enhancements that were restricted to faces, the studies presented here provide novel evidence that supports the hypothesis that face recognition is a highly specialised skill.

2.2. Enhanced configural processing in super recognition

Enhanced configural/holistic processing of face stimuli was the most consistent effect across five of the six SRs described in Chapter 2. Heightened inversion effects were displayed in three SRs, with similar trends observed in the remaining three participants. In four of the six cases, these inversion effects did not generalise to other classes of objects (for a detailed description of these findings see section 7.3. of Chapter 4).

Configural processing, a hallmark of typical face recognition, is thought to be a face-specific mechanism that allows for efficient integration of facial features into a coherent whole. The data reported in Chapter 2 suggests that SRs are particularly proficient at the integration of information in upright faces and this may, in turn, contribute to their extraordinary face recognition ability. Alternatively, it is possible that SRs are particularly sensitive to featural information in upright faces (also affected by inversion, McKone & Yovel, 2009). The latter hypothesis would also explain the increased dwell time on the nose that was reported in Chapter 3. Future studies should manipulate both spacing and featural information to further investigate this issue.

2.3. Regions of facial interest in super recognition

One of the aims of this thesis was to examine face processing strategies in super-recognition. Chapter 3 employed the eye-tracking technology to monitor the eye movements of SRs, participants with developmental prosopagnosia (DP), and typical perceivers. Eye movements are informative of on-line cognitive processing and are known to be functional in face learning (Henderson, Williams, & Falk, 2005) and recognition (Althoff & Cohen, 1999; Luria & Strauss, 1978). Several authors have reported that the eye region is pivotal for successful face recognition (e.g. Bate et al., 2008; Schyns, Bonar, & Gosselin, 2002; Slessor, Riby, & Finnerty, 2013) and that abnormal attention to this region is associated with impaired face recognition in acquired (Caldara et al., 2005; Lê, Raufaste, & Demonet, 2003; Lê, Raufaste, Roussel, Puel, & Démonet, 2003; Stephan & Caine, 2009) and developmental prosopagnosia (Schwarzer et al., 2007) . However, another line of evidence suggests that the region just below the eyes and to the centre of the nose is more functional for identity judgements of face recognition (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012).

The results reported in Chapter 3 seem to bolster the latter hypothesis. Specifically, SRs displayed increased dwell time on the nose region (Experiments 2 and 3), whereas DPs spent less time examining the eye region and more time looking at the mouth (Experiment 1). Pertinently, in control participants, the dwell time on the nose region of the face was positively correlated with their face recognition ability, as measured by the CFMT (Duchaine & Nakayama, 2006).

SRs' reliance on the nose region may play a facilitatory role in their ability to memorise and retrieve identity information. It is possible that more central viewing of facial stimuli enables more efficient encoding (with eyes, mouth and the external features remaining in the periphery), resulting in successful face recognition at a later stage. Whilst the eyes convey considerable information about one's focus of attention and mental state, it is possible that over-reliance on emotional cues is in fact detrimental to the learning of one's identity. However, this hypothesis is contradicted by a number of studies showing that empathy (Bate et al., 2010) and gregariousness (Li et al., 2010) are associated with better face recognition in typical perceivers.

It is also probable that the central fixations are associated with the heightened holistic processing that has been observed in SRs (Chapter 2 of this thesis). Future studies should endeavour to further explore this hypothesis by correlating fixation patterns with measures of holistic processing and face recognition ability.

2.4. Implications for the face recognition spectrum

One of the questions that Chapters 2, 3, and 4 aimed to answer was whether face recognition skills reside upon a spectrum (where super-recognition is at the top end of the distribution), or whether SRs are qualitatively different from the general population. The studies reported here seem to support the former hypothesis, placing SRs at the top end of the face recognition spectrum, rather than identifying a qualitatively different population. Pertinently, super-recognition was found to be underpinned by heightened holistic processing – a skill that has previously been linked to face recognition ability in participants that were unselected for their face recognition skills (Richler et al., 2011; Wang, Li, Fang, Tian, Liu, 2012, c.f. Konar, Bennett, & Sekuler, 2010), and that has been found to be impaired in DP (Avidan et al., 2011; Yovel & Duchaine, 2006; but see LeGrand et al., 2006). Furthermore, results from the eye-tracking investigation in Chapter 3 suggest that SRs do not display a pattern of eye movements that is qualitatively different to typical perceivers, but fixate more on the nose. Importantly, the time spent looking at the nose was also correlated with the face recognition ability of control participants, suggesting that SRs represent merely the top end of this distribution.

This finding has potential implications for the training of face recognition skills in typical perceivers and those affected by DP. Pertinently, if holistic processing and more central fixations do underpin superior face recognition, a training regime might be devised that targets these processes. A recent study by DeGutis, Cohan and Nakayama (2014) reported promising results of such an intervention, whereby a group of individuals with DP showed an improvement in holistic processing and the discrimination of front-view faces. Implementing holistic processing training in typical perceivers would be a natural extension of this research. Additionally, future studies should investigate the effect of instruction on eye-movement patterns and face recognition or matching ability. If successful, this method would provide a potent training tool for forensic and national security agencies.

3. PRACTICAL IMPLICATIONS FOR FORENSIC AND SECURITY SETTINGS

3.1. The utility of super-recognizers in practical settings

The human face is the most common means of accessing one's identity, and yet unfamiliar face recognition and face matching are challenging for typical perceivers. There is a large body of evidence showing that typical perceivers (e.g. Kemp, Towel, and Pike, 1997) and trained passport and Police officers (Burton et al., 1999; White, Kemp, Jenkins, Matheson, & Burton, 2014, but see Robertson et al., 2016) appear to be no better than untrained students or the general public. Indeed, participants tend to make an erroneous response in two out of ten trials in a relatively easy test of face matching, the Glasgow Face Matching Test (GFMT, Burton, White, McNeil, 2010), corresponding to an 80% accuracy rate. This is despite the fact that images of the repeated individuals were taken on the same day, meaning there was no change in the appearance of the target faces.

This work, and other evidence of considerable individual differences in face recognition ability (discussed in depth in Chapters 5 and 6), poses a significant challenge to forensic and national security settings. Specifically, in occupations where the demands of unfamiliar face recognition are high (e.g. in tasks performed by Passport, Police and CCTV officers), the face recognition ability of employees is paramount. A recent study by White and colleagues (2015) showed some evidence that trained facial image analysts are better at face image comparison that those without formal training. Nonetheless, it remains uncertain whether these so-called "experts" have naturally better face recognition skills that draw them to this occupation, or whether their enhanced performance is a result of experience or training. Furthermore, it is unclear whether the performance of these experts was consistent across all individuals, or whether some were better than others (no single-case comparisons were presented).

Chapters 5 and 6 attempted to address these issues by recruiting SRs and control groups to perform applied tasks of face recognition. The findings corroborate evidence that there are large individual differences in our face recognition skills (Bindeman, Aveytisan,

& Rakow, 2012; Burton et al., 2010, Russell et al., 2009), and show that targeted recruitment could be the optimal strategy for occupations requiring a high aptitude of face recognition ability. Indeed, the results from this thesis show that SRs are better at applied face recognition tasks than the general population, and would make valuable employees in border control and national security settings. However, it is important to note that there was considerable variation in the SR group in all four applied experiments reported in Chapters 5 and 6, and that CFMT+ scores did not always correspond to the SRs' scores on the applied tasks. It is thus of pivotal importance that any standard laboratory screening is followed by a battery of applied tasks to assess the true extent of one's face recognition ability.

The London Metropolitan Police has recently established a dedicated SR unit, fully committed to facial image review and comparison. A recent paper by Robertson and colleagues (2016) attempted to evaluate the performance of four SR officers from this task force on a range of familiar and unfamiliar face matching tests. The SR officers consistently outperformed control participants with degraded and high quality images. The London Metropolitan Police SR Unit is the first formation of this kind, putting the extraordinary face recognition skills into good use.

3.2. Identifying super-recognizers for applied face-processing tasks

The identification of SRs for theoretical and applied research, as well as assignments requiring extraordinary face recognition skills is of great interest to both researchers and industry beneficiaries. To date, the CFMT+ (Russell et al., 2009) and the CFPT (Duchaine, et al., 2007) have been utilised for this purpose. However, these tests originate from the DP literature and, as discussed above, may be unsuitable for the identification of SRs.

In Chapter 5, we correlated performance on applied tasks of face matching and face memory with CFMT scores in control participants. The ability to match faces was unrelated to performance on a well-established one-in-ten task (Bruce et al. 1999), and only correlated moderately with an applied task of face recognition from moving footage. What is more, within the SR group, individuals' CFMT+ scores did not reflect their performance on either of the tests. For instance, in Experiment 1 of this chapter, SR1 who obtained the lowest CFMT+ score and was included in Chapters 2 and 5 out of theoretical interest, was also the highest scoring individual in the line-up task. Similarly, the none-to-moderate correlations between laboratory tests and the applied tasks of face recognition in the control group suggest that the CFMT is not a useful screening tool for assignments involving real-life face recognition.

It is important to note that the results from this thesis suggest that perceptual and mnemonic abilities do not always co-occur in all individuals and any new battery of tests should reflect these disparities. Private correspondence of the author of this thesis with the Metropolitan Police Super-Recogniser Unit suggests that such an approach is loosely adopted by the Force. Pertinently, officers deployed in the Unit have been identified via internal league tables that reflect the number of correct person identifications in a variety of assignments, such as the reviewing of CCTV footage or recognition of suspects during patrols. These SR officers have since been successfully involved in a number of high-profile cases both within the UK and abroad. The recruitment and deployment of this small task force is based on their historical performance and has proved to be an invaluable resource for the London Metropolitan Police.

While the retrospective identification of skilled individuals is a worthwhile strategy, it is not necessarily the most effective to adopt. Pertinently, the four officers investigated in Robertson et al.'s (2016) study were aged 33-47 years, and it is plausible that they had at least a decade of service behind them in order to gather the data leading to their deployment in the Unit. However, the Policing College trainees who were recruited as a control sample in this study could potentially include individuals with high recognition skills. An in-depth and sophisticated assessment battery that includes a selection of tests (e.g. assessing face memory, face perception, recognition of other race faces, etc.) and that is administered at recruitment phases would allow for early selection of skilled officers and subsequent deployment to optimal tasks.

3.3. Can super recognition be trained?

Attempts to improve unfamiliar face matching in typical perceivers have yielded mixed results. There is some evidence that feedback prevents decline in performance when tasks are performed over a long period of time (Alenzi & Bindemann, 2013), and also improves face matching (White, Kemp, Jenkins, & Burton, 2014). Other training efforts have included food incentives which were effective at enhancing performance in mismatched trials (Moore & Johnston, 2013), and shape classification (Towler, White, & Kemp, 2014) - a method that is recommended in many training programmes but was found to be largely ineffective in Towler et al.'s paper. Techniques aimed at preventing the temporal decline of performance in face matching, such as desk switching and enforced rest breaks (Alenzi, Bindemann, Fysh, & Johnston, 2015) have also been unsuccessful and further highlight the challenging nature of unfamiliar face matching and the necessity for more effective training techniques.

Interestingly, a recent study by Dowsett & Burton (2015) showed that individuals perform better at face matching when they are allowed to work in pairs and communicate freely. Most importantly, this effect was restricted to the training phase, but showed a promising degree of skill transfer following the training on a new set of faces. The long term effects of this training strategy are unknown, however, and it is possible that this enhancement is restricted to the time immediately following the performing of the task in pairs.

A novel solution, intranasal inhalation of oxytocin, has also shown little promise for improving face recognition. Bate and colleagues (2014) administered the nonapeptide to typical participants and asked them to perform the one-in-ten task following a 45 minute rest break. The intervention did not improve overall accuracy on the task, but did induce a liberal response bias, i.e. individuals were likely to elicit a “yes” response even if the target was absent from the line-up. Such pattern of errors would be particularly detrimental in national security settings where the detection of fraudulent documents is of paramount importance. However, simultaneous line-up procedures are vulnerable to such a shift in response (see Valentine & Fitzgerald, 2016 for a recent review), and it is possible that oxytocin further exacerbated this shortcoming. To address this concern and further explore the potential use of oxytocin in applied tasks of face recognition and face matching, studies reported in Chapter 7 examined the use of the nonapeptide in the Models Face Matching Task (Dowsett & Burton, 2015) and the challenging face memory task previously used in Experiment 2 of Chapter 5 in this thesis. If effective, oxytocin could deliver an instant improvement to one's face recognition ability and thus would be beneficial for use in forensic and national security settings, e.g. during passport checks or other types of facial image comparison. However, the administration of oxytocin did not improve the face recognition and face matching ability of students recruited in these studies.

Overall, there appears to be limited evidence of efficient and cost effective regimes of face recognition and face matching training. While work in pairs and food incentives showed some improvement in participants' performance, these interventions are of limited utility in applied settings for practical reasons, i.e. increased staffing levels and availability of food incentives are unlikely to be implemented in government agencies. Furthermore, the results from Chapter 7 add to the body of evidence that face recognition ability is resistant to improvements elicited by a pharmaceutical intervention. Given the limited indication of effectiveness of training and the studies showing that face recognition ability has a strong genetic component (Kennerknecht, Pluempe, & Welling, 2008; Shakeshaft & Plomin, 2015; Wilmer et al., 2010), optimising staff recruitment and allocation may provide the best solution in forensic and national security settings. While super recognition is likely difficult (or even impossible) to train, Chapters 5 and 6 of this thesis, as well as the evaluation of the Metropolitan Police SR Unit (Robertson et al., 2016) provide strong evidence that allocating the right people to the right jobs would be the most fruitful strategy to employ until more effective training regimes are developed.

4. SUMMARY AND FUTURE DIRECTIONS

This thesis set out to examine the cognitive and perceptual underpinnings of super-recognition, the applied value of extraordinary face recognition skills, the distribution of face recognition ability, and novel ways of improving face recognition and face matching in typical perceivers. In terms of theory, this work has provided further evidence supporting the domain specificity of face recognition and the relationship between holistic processing and face recognition ability. This thesis is also the first to report the central tendency of SRs' eye-movements - a processing strategy that potentially contributes to their extraordinary face recognition skills. Importantly, the reported dissociation between face memory and face perception provided little support for the sequential model of face recognition (Bruce & Young, 1986) and showed that this pattern of findings is better explained within the Interaction Activation and Competition Model framework (IAC, Burton et al, 1990). In terms of practice, SRs have been found to excel at applied face recognition tasks and would make valuable employees within forensic and national security settings.

Future research should aim to develop a comprehensive battery for assessment of super recognition, taking into account the probability that different subtypes exist. It is hoped that the work presented in this thesis will be an important stepping stone in this endeavour.

References

Alenezi, H. M., & Bindemann, M. (2013). The Effect of Feedback on Face-Matching Accuracy. *Applied Cognitive Psychology, 27*(6), 735–753.

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

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 997.

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.

Arnell, K., & Dube, B. (2015). Extraversion predicts superior face-specific recognition ability, but through experience, not positive affect. *Journal of Vision, 15*(12), 175-175.

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia, 49*(9), 2541-2552.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. J*ournal of Child Psychology and Psychiatry, 42*(2), 241-251.

Barton, J. J. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology*, *2*(1), 197-225.

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Sciences*, *15*(7), 301-309.

Bate, S., & Bennetts, R. J. (2014). The rehabilitation of face recognition impairments: a critical review and future directions. *Frontiers in Human Neuroscience*, 8.

Bate, S., & Cook, S. J. (2012). Covert recognition relies on affective valence in developmental prosopagnosia: Evidence from the skin conductance response. *Neuropsychology*, *26*(5), 670-674.

Bate, S., Bennetts, R., Mole, J. A., Ainge, J. A., Gregory, N. J., Bobak, A. K., & Bussunt, A. (2015). Rehabilitation of face-processing skills in an adolescent with prosopagnosia: Evaluation of an online perceptual training programme. *Neuropsychological Rehabilitation*, *25*(5), 733-762.

Bate, S., Bennetts, R., Parris, B. A., Bindemann, M., Udale, R., & Bussunt, A. (2014). Oxytocin increases bias, but not accuracy, in face recognition line-ups. *Social Cognitive and Affective Neuroscience*, nsu150.

Bate, S., Cook, S. J., Duchaine, B., Tree, J. J., Burns, E. J., & Hodgson, T. L. (2014). Intranasal inhalation of oxytocin improves face processing in developmental prosopagnosia. *Cortex*, *50*, 55-63.

Bate, S., Haslam, C., & Hodgson, T. L. (2009). Angry faces are special too: Evidence from the visual scanpath. *Neuropsychology*, *23*(5), 658.

Bate, S., Haslam, C., Jansari, A., & Hodgson, T. L. (2009). Covert face recognition relies on affective valence in congenital prosopagnosia. *Cognitive Neuropsychology, 26*(4), 391-411.

Bate, S., Haslam, C., Tree, J. J., & Hodgson, T. L. (2008). Evidence of an eye movement-based memory effect in congenital prosopagnosia. *Cortex, 44*(7), 806-819.

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.

Bauer, R. M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. *Neuropsychologia, 22*(4), 457-469.

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

Bennetts, R. J., Butcher, N., Lander, K., Udale, R., & Bate, S. (2015). Movement cues aid face recognition in developmental prosopagnosia. *Neuropsychology, 29*(6), 855-860.

Bennetts, R. J., Murray, E., Boyce, T., & Bate, S. (In press). Prevalence of developmental prosopagnosia in middle childhood. *Quarterly Journal of Experimental Psychology.*

Benton, A. L., & Van Allen, M. W. (1972). Prosopagnosia and facial discrimination. *Journal of the Neurological Sciences, 15*(2), 167-172.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18*(8), 706-712.

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., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied, 18*(3), 277-291.

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(2), 96-103.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2), 341-355.

Bishop, D. V. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(4), 899-923.

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

Bodamer, J. (1947). Die prosopagnosie. *Archiv für Psychiatrie und Nervenkrankheiten*, *179*(1), 6-53.

Born, J., Lange, T., Kern, W., McGregor, G. P., Bickel, U., & Fehm, H. L. (2002). Sniffing neuropeptides: a transnasal approach to the human brain. *Nature Neuroscience, 5*(6), 514-516.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., ... & Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology, 26*(5), 423-455. doi: 10.1080/02643290903343149

Braje, W. L., Kersten, D., Tarr, M. J., & Troje, N. F. (2013). Illumination effects in face recognition. *Psychobiology, 26*(4), 371–380.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception, 23*(7), 803-822.

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

Bruce, V., Henderson, Z., Greenwood, K., B, J., Mike, A., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339–360.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339-360.

Brunet, P. M., Mondloch, C. J., & Schmidt, L. A. (2010). Shy children are less sensitive to some cues to facial recognition. *Child Psychiatry & Human Development, 41*(1), 1-14.

Brunsdon, R., Coltheart, M., Nickels, L., & Joy, P. (2006). Developmental prosopagnosia: A case analysis and treatment study. *Cognitive Neuropsychology, 23*(6), 822-840.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., & Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain and Cognition*, *2*(3), 257-284.

Bukach, C. M., Grand, R., Kaiser, M. D., Bub, D. N., & Tanaka, J. W. (2008). Preservation of mouth region processing in two cases of prosopagnosia. *Journal of Neuropsychology*, *2*(1), 227-244.

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., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2015). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science, 40*(1), 202-223.

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

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

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.

Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., & Rossion, B. (2010). Holistic perception of the individual face is specific and necessary: evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia,48*(14), 4057-4092.

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.

Clark, S. E. (2012). Eyewitness Identification Reform Data, Theory, and Due Process. *Perspectives on Psychological Science*, *7*(3), 279-283.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist,12*(4), 482-486.

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology, 27*(3), 245-260.

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition, 111*(2), 219-247.

Curby, K. M., & Gauthier, I. (2014). Interference between face and non-face domains of perceptual expertise: a replication and extension. Frontiers in *Psychology, 5,* 955.

Dalrymple, K. A., & Duchaine, B. (in 2015). Impaired face detection may explain some but not all cases of developmental prosopagnosia. *Developmental Science, 19*(3), 440-451.

Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience, 10*, 10-20.

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia Anatomic basis and behavioral mechanisms. *Neurology*, *32*(4), 331-331.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PloS One*, *6*(12), e28800.

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.

De Haan, E. H. (1999). A familial factor in the development of face recognition deficits. J*ournal of Clinical and Experimental Neuropsychology,21*(3), 312-315.

De Renzi, E. (1986). *Current issues on prosopagnosia*. In Aspects of face processing (pp. 243-252). Springer Netherlands.

De Renzi, E., & Di Pellegrino, G. (1998). Prosopagnosia and alexia without object agnosia. *Cortex*, *34*(3), 403-415.

De Renzi, E., & Saetti, M. C. (1997). Associative agnosia and optic aphasia: qualitative or quantitative difference? *Cortex*, *33*(1), 115-130.

De Renzi, E., Faglioni, P., Grossi, D., & Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. *Cortex, 27*(2), 213-221.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence. *Law and Human Behavior, 4*(4), 243-260.

DeGutis, J., Cohan, S., & Nakayama, K. (2014). Holistic face training enhances face processing in developmental prosopagnosia. *Brain,* awu062.

DeGutis, J., Cohan, S., Mercado, R. J., Wilmer, J., & Nakayama, K. (2012). Holistic processing of the mouth but not the eyes in developmental prosopagnosia. *Cognitive Neuropsychology, 29*(5), 419-446.

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition, 126*(1), 87-100.

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.

Domes, G., Lischke, A., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., & Herpertz, S. C. (2007). Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology, 35*(1), 83-93.

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. (2015). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *The Quarterly Journal of Experimental Psychology*, *69*(1), 1-10.

Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception,32*(7), 827-838.

Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology, 23*(5), 714-747.

Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience, 17*(2), 249-261.

Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *17*(2), 249-261.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia,44*(4), 576-585.

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., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social Cognitive and Affective Neuroscience, 2*(2), 104-113.

Ducrot, S., & Pynte, J. (2002). What determines the eyes' landing position in words? *Perception & Psychophysics, 64*(7), 1130-1144.

Dutağacı, H., Yörük, E., & Sankur, B. (2008). Comparative analysis of global hand appearance-based person recognition. *Journal of Electronic Imaging, 17*(1), 011018-011018.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology, 111*(4), 694-705.

Eimer, M., & McCarthy, R. A. (1999). Prosopagnosia and structural encoding of faces: Evidence from event-related potentials. *Neuroreport, 10*(2), 255-259.

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain, 135*(2), 542-554.

Ellis, H. D., & Young, A. W. (1988). Training in face-processing skills for a child with acquired prosopagnosia. *Developmental Neuropsychology, 4*(4), 283-294.

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

Farah, M. J. (1991). Cognitive Neuropsychology: Patterns of Co-occurrence Among the Associative Agnosias: Implications for Visual Object Representation. *Cognitive Neuropsychology, 8*(1), 1-19.

Fletcher-Watson, S., Leekam, S. R., Findlay, J. M., & Stanton, E. C. (2008). Brief report: Young adults with autism spectrum disorder show normal attention to eye-gaze information—Evidence from a new change blindness paradigm. *Journal of Autism and Developmental Disorders, 38*(9), 1785-1790.

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences, 14*(1), 53-65.

Gao, Z., Flevaris, A. V., Robertson, L. C., & Bentin, S. (2011). Priming global and local processing of composite faces: revisiting the processing-bias effect on face perception. *Attention, Perception, & Psychophysics, 73*(5), 1477-1486.

Garrido, L., Duchaine, B., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology*, *2*(1), 119-140.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition*, *103*(2), 322-330.

Gerrard, G., & Thompson, R. (2011). Two million cameras in the UK. *CCTV Image, 42*, 10-12.

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

Going, M., & Read, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills, 39*(1), 109-110.

Gossen, A., Hahn, A., Westphal, L., Prinz, S., Schultz, R. T., Gründer, G., & Spreckelmeyer, K. N. (2012). Oxytocin plasma concentrations after single intranasal oxytocin administration–a study in healthy men. *Neuropeptides,46*(5), 211-215.

Grueter, M., Grueter, T., Bell, V., Horst, J., Laskowski, W., Sperling, K., ... & Kennerknecht, I. (2007). Hereditary prosopagnosia: The first case series. *Cortex*, *43*(6), 734-749.

Guastella, A. J., Mitchell, P. B., & Dadds, M. R. (2008). Oxytocin increases gaze to the eye region of human faces. *Biological Psychiatry, 63*(1), 3-5.

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

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.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*(1), 98–106.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*(1), 98-106.

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

Hills, P. J., Eaton, E., & Pake, J. M. (2016). Correlations between psychometric schizotypy, scan path length, fixations on the eyes and face recognition. *The Quarterly Journal of Experimental Psychology, 69*(4), 611-625..

Hills, P. J., & Lewis, M. B. (2009). A spatial frequency account of the detriment that local processing of Navon letters has on face recognition. *Journal of Experimental Psychology: Human Perception and Performance,35*(5), 1427-1442.

Holdnack, H. A. (2001). *Wechsler Test of Adult Reading: WTAR*. Psychological Corporation: San Antonio, TX

Hsiao, J. H. W., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science, 19*(10), 998-1006.

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1571), 1671–1683.

Jones, R. D., & Tranel, D. (2001). Severe developmental prosopagnosia in a child with superior intellect. *Journal of Clinical and Experimental Neuropsychology, 23*(3), 265-273

Kane, M. J., & Engle, R. W. (2003). Working-Memory Capacity and the Control of Action: The Contributions of Goal Neglect, Response Competition, and Task Set to Stroop Interference. *Journal of Experimental Psychology: General, 132*(1), 47-70.

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.

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., & Grueter, M. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics Part A, 140*(15), 1617-1622.

Kennerknecht, I., Pluempe, N., & Welling, B. (2007). Congenital prosopagnosia--a common hereditary cognitive dysfunction in humans. *Frontiers in Bioscience: A Journal and Virtual Library, 13,* 3150-3158.

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., & Meyer-Lindenberg, A. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *The Journal of Neuroscience, 25(49),* 11489-11493.

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

Labuschagne, I., Phan, K. L., Wood, A., Angstadt, M., Chua, P., Heinrichs, M., & Nathan, P. J. (2010). Oxytocin attenuates amygdala reactivity to fear in generalized social anxiety disorder. *Neuropsychopharmacology, 35*(12), 2403-2413.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological Medicine, 34*(05), 911-920.

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition, 61*(2), 139-158.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15*(11), 762-768.

Lê, S., Raufaste, E., Roussel, S., Puel, M., & Démonet, J. F. (2003). Implicit face perception in a patient with visual agnosia? Evidence from behavioural and eye-tracking analyses. *Neuropsychologia, 41*(6), 702-712.

Leach, A. M., Cutler, B. L., & Van Wallendael, L. (2009). Lineups and eyewitness identification. *Annual Review of Law and Social Science, 5*, 157-178.

Lee, Y., Duchaine, B., Wilson, H. R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: Detailed neuropsychological and psychophysical investigation of face processing. *Cortex, 46*(8), 949-964.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—Women's faces make the difference. *Brain and Cognition, 50*(1), 121-128.

Lewis, M. B., Mills, C., Hills, P. J., & Weston, N. (2009). Navon letters affect face learning and face retrieval. *Experimental Psychology, 56*(4), 258-264.

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.

Lindholm, T. (2005). Own-age biases in verbal person memory. *Memory, 13*(1), 21-30.

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76*(6), 796-802.

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*(3), 556-564.

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 77-100.

Luria, S. M., & Strauss, M. S. (1978). Comparison of eye movements over faces in photographic positives and negatives. *Perception, 42*(11), 1134-1143.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide* (2nd edition). Mahwah, N.J: Psychology Press.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science, 13*(2), 194-196.

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.

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

McGugin, R. W., Van Gulick, A. E., & Gauthier, I. (in press). Cortical Thickness in Fusiform Face Area Predicts Face and Object Recognition Performance. *Journal of Cognitive Neuroscience.*

McIntyre, A. H., Hancock, P. J. B., Kittler, J., & Langton, S. R. H. (2013). Improving Discrimination and Face Matching with Caricature. *Applied Cognitive Psychology, 27*(6), 725–734.

McKelvie, S. J., Standing, L., Jean, D. S., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society, 31*(5), 447-448.

McKone, E., & Robbins, R. (2007). The evidence rejects the expertise hypothesis: Reply to Gauthier & Bukach. *Cognition, 103(*2), 331-336.

McKone, E., & Robbins, R. (2011). *Are faces special*. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), The Oxford handbook of face perception (pp. 149-176). Oxford, UK: Oxford University Press.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review, 16*(5), 778-797.

McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., ... & O'Connor, K. B. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test–Australian. *Cognitive Neuropsychology*, *28*(2), 109-146.

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

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., Bindemann, M., & Havard, C. (2011). Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica, 137*(1), 83–89.

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

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

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

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*(3), 353-383.

Norell, K., Läthén, K. B., Bergström, P., Rice, A., Natu, V., & O'Toole, A. (2015). The effect of image quality and forensic expertise in facial image comparisons. *Journal of Forensic Sciences, 60*(2), 331-340.

O'Regan, J. K. (1981). *The convenient viewing position hypothesis*. Eye movements: Cognition and Visual Perception, 289-298.

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 185-197.

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., ... & Al-Janabi, S. (In press). Do people have insight into their face recognition abilities? *The Quarterly Journal of Experimental Psychology.*

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia, 49*(5), 1226-1235.

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36*(3), 247.

Parris, B.A., Sharma, D., & Weekes, B.S. (2007). An Optimal Viewing Position Effect in the Stroop Task When Only One Letter Is The Colour Carrier. *Experimental Psychology, 54*(4), 273-280.

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, 109*(48), E3314-E3323.

Petrovic, P., Kalisch, R., Singer, T., & Dolan, R. J. (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *The Journal of Neuroscience, 28*(26), 6607-6615.

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic,""configural," and "relational" in the face perception literature. *Frontiers in Psychology, 3*, 559.

Plews, S., Bowie, J., & Bromby, M. (2004). When Evidence is a Question of Image. *Law Gazette, 34.* Retrieved 08/28, 2014, from http://www.lawgazette.co.uk/news/when-evidence-is-a-question-of-image/42014.article.

Ramon, M., & Rossion, B. (2010). Impaired processing of relative distances between features and of the eye region in acquired prosopagnosia—Two sides of the same holistic coin? *Cortex, 46*(3), 374-389.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8*(1), 21-30.

Rezlescu, C., Pitcher, D., & Duchaine, B. (2012). Acquired prosopagnosia with spared within-class object recognition but impaired recognition of degraded basic-level objects. *Cognitive Neuropsychology, 29*(4), 325-347.

Riby, D. M., & Hancock, P. J. (2008). Viewing it differently: Social scene perception in Williams syndrome and autism. *Neuropsychologia*, *46*(11), 2855-2860.

Richler, J. J., & Gauthier, I. (2013). When intuition fails to align with data: A reply to Rossion (2013). *Visual Cognition, 21*(2), 254-276.

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin, 140*(5), 1281.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011a). Holistic processing predicts face recognition. *Psychological Science, 22*, 464-471.

Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision, 15*(9), 15.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011b). Inverted faces are (eventually) processed holistically. *Vision Research, 51*(3), 333-342.

Riddoch, M. J., & Humphreys, G. W. (1993). BORB: Birmingham object recognition battery. LEA.

Rimmele, U., Hediger, K., Heinrichs, M., & Klaver, P. (2009). Oxytocin makes a face in memory familiar. *The Journal of Neuroscience, 29*(1), 38-42.

Ritchie, K., & Burton, A. M. (In press). Learning faces from variability. *Quarterly Journal of Experimental Psychology.*

Robbins, R. A., Shergill, Y., Maurer, D., & Lewis, T. L. (2011). Development of sensitivity to spacing versus feature changes in pictures of houses: Evidence for slow development of a general spacing detection mechanism? *Journal of Experimental Child Psychology, 109*(3), 371-382.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103*(1), 34-79.

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

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*(2), 139-253.

Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: No special role for surface reflectance processing. *Neuropsychologia*, *50*(2), 334-340.

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

Savaskan, E., Ehrhardt, R., Schulz, A., Walter, M., & Schächinger, H. (2008). Post-learning intranasal oxytocin modulates human memory for facial identity. *Psychoneuroendocrinology, 33*(3), 368-374.

Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., & Akarun, L. (2008). *Bosphorus database for 3D face analysis*. In Biometrics and Identity Management (pp. 47-56). Springer Berlin Heidelberg.

Schinka, J. A., Loewenstein, D. A., Raj, A., Schoenberg, M. R., Banko, J. L., Potter, H., & Duara, R. (2010). Defining mild cognitive impairment: impact of varying decision criteria on neuropsychological diagnostic frequencies and correlates. *The American Journal of Geriatric Psychiatry, 18*(8), 684-691.

Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology, 2*(1), 99-117.

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology, 25*(5), 704-729.

Schwarzer, G., Huber, S., Grüter, M., Grüter, T., Groß, C., Hipfel, M., & Kennerknecht, I. (2007). Gaze behaviour in hereditary prosopagnosia. *Psychological Research, 71*(5), 583-590.

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.

Seidlitz, L., & Diener, E. (1998). Sex differences in the recall of affective experiences. *Journal of Personality and Social Psychology, 74*(1), 262.

Sekiguchi, T. (2011). Individual differences in face memory and eye fixation patterns during face learning. *Acta Psychologica*, *137*(1), 1-9.

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science, 2*(6), 140343.

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

Slessor, G., Riby, D. M., & Finnerty, A. N. (2013). Age-related differences in processing face configuration: The importance of the eye region. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68*(2), 228-231.

Smilek, D., Solman, G.J.F., Murawski, P., & Carriere, J.S.A. (2009). The eyes fixate the optimal viewing position of task-irrelevant words. *Psychonomic Bulletin and Review, 16*(1), 57-61.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17*(1), 99-139.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: a meta-analytic comparison. *Law and Human Behavior, 25*(5), 459-473.

Stephan, B. C. M., & Caine, D. (2009). Aberrant pattern of scanning in prosopagnosia reflects impaired face processing. *Brain and Cognition, 69*(2), 262-268.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF) Hogrefe*. Göttingen, Germany.

Striepens, N., Kendrick, K. M., Hanking, V., Landgraf, R., Wüllner, U., Maier, W., & Hurlemann, R. (2013). Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. *Scientific Reports, 3*, 3440.

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

Susilo, T., Germine, L., & Duchaine, B. (2013). Face recognition ability matures late: Evidence from individual differences in young adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39(*5), 1212-2017.

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

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46*(2), 225-245.

Thomas, A. L., Lawler, K., Olson, I. R., & Aguirre, G. K. (2008). The Philadelphia face perception battery. *Archives of Clinical Neuropsychology*,*23*(2), 175-187.

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

Townsend, J.T., & Ashby, F.G. (1978). *Methods of modeling capacity in simple processing systems.* In J. Castellan & F. Restle (Eds.), Cognitive theory. Vol. 3. (pp. 200-239). Hillsdale, N.J.: Erlbaum.

Townsend, J.T., & Ashby, F.G. (1983). *Stochastic modeling of elementary psychological processes.* Cambridge: Cambridge University Press.

Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology, 38*(5), 690-690.

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., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception, 15*(5), 525-535.

Valentine, T., & Fitzgerald, R. J. (2016). Identifying the culprit: an international perspective on the National Academy of Sciences Report on eyewitness identification evidence. *Applied Cognitive Psychology, 30*(1), 135-138.

Van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2012). A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology, 37*(3), 438-443.

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.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II).* San Antonio, TX: NCS Pearson.

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.

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015a). Error Rates in Users of Automatic Face Recognition Software. *PloS One, 10*(10), e0139827.

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

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS One, 9*(8), e103510.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015b). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B, 282*(1814), 20151292.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of Royal Society B, 282*(1814), 20151292.

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., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology, 29*(5), 360-392.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., ... & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences, 107*(11), 5238-5241.

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. V. (1979). On training people to recognize faces. *Ergonomics, 22*(3), 333-343.

Xivry, J. J. O., Ramon, M., Lefevre, P., & Rossion, B. (2008). Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *Journal of Neuropsychology, 2*(1), 245-268.

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.

Young, A. W., Newcombe, F., de Haan, E. H., & Hay, D. C. (1993). Face perception after brain injury. *Brain*, *116*(4), 941-959.

Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., and Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests (FEEST).*Bury St. Edmunds, U.K., Thames Valley Test Company. Ref Type: Computer Program.

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.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., ... & Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology*,*20*(2), 137-142.

Appendix A: Participant questionnaire used in Chapter 4

1. Participant reference…………………………………………………………………………
2. Date………………………………………………………………………………………
…
3. Age (years)…………………………………………………………………………………
4. Full years of <u>university education</u>………… **and** years of <u>formal education</u> (total from year 1)………………………………………………………………………………………..
5. Gender (**Male /Female**)………………………………………...
6. Handedness (**Right/Left/Ambidextrous**)…………………………………………………
7. To the best of your knowledge, have you had any periods of visual deprivation, especially during childhood? (e.g. eye surgery, temporary blindness, or any other condition that affected your sight for a prolonged period of time) **Yes/No**. If yes, please specify
………………………………………………………………………………………………
…

8. To the best of your knowledge, have you had any brain trauma in the past (e.g. encephalitis, stroke, physical injury, carbon monoxide poisoning, or others) **Yes/No**. If yes, please specify.
………………………………………………………………………………………………
………………………………………………………………………………………………
……
9. How would you rate your face recognition ability, that is how well do you remember and recognise faces as seen before, not how well you remember peoples' names 1 = much worse than average; **2** = slightly worse than average; **3** = average; **4** = slightly better than average; **5** = much better than average)…………………………………..

10. Do you fail to recognize faces you should (e.g. co-workers, family, friends)? (**1** = all the time; **2** = often; **3** = sometimes; **4** = rarely; **5** = never)…………………………..

11. How would you rate your ability to follow characters in a movie? (**1** = very difficult; **2** = sometimes difficult ; **3** = neither difficult or easy; **4** = mostly easy; **5** = very easy)
……………………………………………………………………………………..

12. Do people sometimes find it surprising that you recognized their face (e.g. because you only met briefly, and/or long time ago) ? (**1** = all the time; **2** = often; **3** = sometimes; **4** = rarely; **5** = never)……………………………………………………

13. To the best of your knowledge, do you suffer from any developmental disorders (e.g. Autism Spectrum Disorder, Moebius Syndrome, Attention Deficit Hyperactivity Disorder, Attention Deficit Disorder, or others) **Yes/No.** If yes, please specify.