The Performance of Untrained Humans Verifying Children and Adults in a Face Matching Task

Thomas Pearce

School of Psychology

The University of Adelaide

October 2018

This thesis is submitted in partial fulfilment of the Honours Degree of Bachelor of Psychological Science

Word Count: 8,967

TABLE OF CONTENTS

LIST OF FIGURES V

ABSTRACT VI

DECLARATION VII

ACKNOWLEDGEMENTS VIII

CHAPTER 1: INTRODUCTION 1

1.1 Rationale 1

1.2 Overview of Face Matching 2

1.2.1 One-to-One and One-to-Many Face Matching Tasks 3

1.2.2 The Use of Automated Systems in Face Matching 3

1.2.3 Familiar and Unfamiliar Face Matching 4

1.3 Challenges in Conducting Unfamiliar Face Matching Tasks 6

1.3.1 The Impact of Image Pair Type on Performance 7

1.4 General Face Matching Performance of Novices Compared to Facial Practitioners 8

1.5 Performance Matching Child Faces 11

1.6 The Present Study 14

1.6.1 Research Question 15

CHAPTER 2: METHOD 16

2.1 Ethics Statement 16

2.2 Participants 16

2.3 Design and Measures 17

2.4 Materials 17

2.4.1 Experimental Application 17

2.4.2 The Image Database 18

2.5 Procedure 19

CHAPTER 3: RESULTS 21

3.1 Data Screening, Assumptions, and Test Selection 21

3.2 Performance with Images of Children and Adults 22

3.2.1 Signal Detection Analysis 25

3.3 Performance with Images of Children and Adults on Mated and Non-Mated Pairs 26

CHAPTER 4: DISCUSSION 30

4.1 Overview of Performance on Child and Adult Conditions 30

4.2 Overview of Performance on Mated and Non-Mated Pairs 31

4.3 Limitations 32

4.4 Strengths 33

4.5 Implications 34

4.6 Suggestions For Future Research 36

4.7 Conclusion.....37

REFERENCES 39

APPENDICES 47

Appendix A: University of Adelaide Information Sheet.....47

Appendix B: Defence Science and Technology Information Sheet 49

Appendix C: Interface Consent 50

Appendix D: Consent Form 51

Appendix E: Poster.....53

Appendix F: Notched boxplots explanation 54

Appendix G: Signal detection table.....55

Appendix H: ROC outputs 56

List of Figures

Figure 2.1. Experimental Interface. 18

Figure 3.1. Overall accuracy for the Child and Adult image groups. 22

Figure 3.2. Confidence levels for the Child and Adult image groups. 23

Figure 3.3. Median response time for the Child and Adult image groups. 24

Figure 3.4. Empirical ROC for performance using Child and Adult images (dashed line shows chance performance). 25

Figure 3.5. Overall accuracy for mated and non-mated Child and Adult image groups. 26

Figure 3.6. Confidence levels for mated and non-mated Child and Adult image groups. 27

Figure 3.7. Median response time for mated and non-mated Child and Adult image groups...... 28

Abstract

Identifying children is a high priority in numerous government agencies not only in border checks, but to aid against child exploitation. However, research suggests children are hard to identify due to childhood facial development. Face matching is a common form of identification in areas such as border checks and investigative applications. A recent study demonstrated that trained facial practitioners found it more difficult to verify child identities compared to adult identities in a one-to-one unfamiliar face matching task, but there is a significant gap in research on whether it is naturally challenging to verify child identities. Thus, the present study primarily aimed to determine whether people with no training or experience find it naturally harder to match child compared to adult faces. The study secondly aimed to find if these people performed better determining whether two faces belonged to the same person, or different people. Students (N = 35) were asked to perform 200 one-to-one face matching trials, determining whether pairs of faces belonged to the same or different people and rate their confidence. The results demonstrated novices were significantly less accurate, confident and slower comparing child images compared to adult images, although performance when image pairs were the same versus different people had mixed results. The findings indicate people are naturally worse at verifying children, providing an argument for child specific training for human facial practitioners. Future research should compare people with and without training and experience, for an extensive analysis on whether this effects human face matching performance.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

October, 2018

Acknowledgements

Firstly, I would like to give a huge thankyou to my supervisors Dr Dana Michalski and Dr Carolyn Semmler, and to Defence Science and Technology group external researcher Veneta Macleod. Their endless support and direction, as well as numerous hours taken out of their days helping construct the thesis was crucial in helping me achieve the best I could.

To my friends and family, thankyou for your endless support throughout the year, and for allowing me to spend most of my time at home sitting in front of a laptop but also taking me out when I needed it.

And lastly to my friends in the Honours cohort, you are all amazing and I'm so glad to have shared this experience with you all, I wish you all the best.

Chapter 1: Introduction

1.1 Rationale

The ability to verify the identity of children is highly important for both national security agencies and local law enforcement. Not only are children part of normal identification checks (e.g., passport control), they are subject to exploitation (Kramer, Mulgrew, & Reynolds, 2018). It is estimated that up to 400,000 children are involved in being trafficked across international borders each year (U.S. Department of State, 2007), there are illegal adoption practices taking place (UNODC, 2016), and there is an emerging risk of child radicalisation (Sewell & Hulusi, 2016). Numerous local crimes such as kidnapping may also require local law enforcement to identify children. Therefore, the wellbeing of children may be dependent upon correctly determining their identity in investigative applications and operational contexts. Research has suggested children might be naturally harder to identify than adults, as they have less discriminating facial features which make them more difficult to distinguish from each other (Wilkinson, 2012). Children also experience a significant amount of craniofacial growth and development (Kozak, Ospina, & Cardenas, 2015). Considering the implications, it is important to explore the differences in identifying children compared to adults.

A common form of identification is face matching. This involves comparing two or more faces and deciding whether they belong to the same person or different people (FISWG, 2012). A recent comprehensive study found trained facial practitioners perform with lower accuracy, confidence, and response time when making decisions on images of children compared to adults on a face matching task using faces unfamiliar to the observers (Michalski, 2017)**.** However, a significant gap in the literature is the lack of empirical research on individuals with no training or experience on whether they find it difficult to verify the identities of children. Steps have been taken to aid in the identification of children, for example requiring children to update their passports every five years in Australia (Department of Foreign Affairs and Trade, n.d). However, during facial practitioner training there may be no focus on the specific difficulties comparing children's faces (e.g., FISWG, 2010).

Considering practitioners have been found to have lower performance verifying child identities compared to adults, assessing how people perform with no training or experience would provide insight into whether this is a natural challenge. Consequently, if novices have lower performance comparing images of children, this implies that children are naturally harder to verify, and facial practitioners training should reflect this.[1]

1.2 Overview of Face Matching

Face matching is the task of matching identities by comparing two or more faces or facial features and deciding whether they are the same person or not, by making a match or non-match decision (FISWG, 2012). Face matching is commonly undertaken by trained facial practitioners, notably in passport checks. Whilst research has found experience helps develop face matching skills (Park, Newman, & Polk, 2009), there are significant individual differences (Megreya & Burton, 2008; Royer, Blais, Barnabe-Lortie, Carree, Leclerc, & Fiset, 2016; White, Kemp, Jenkins, Matheson, & Burton, 2014), and a genetic predisposition in ability to process faces (Johnson, Dzirawiec, Ellis, & Morton, 1991). Interestingly, people who self-identify as super recognisers also perform better matching unfamiliar faces than the average person (Bobak, Hancock, & Bate, 2015; Bobak, Dowsett, & Bate, 2016). Due to face matching being

[1] People with no training or experience in face matching will be referred to as novices within the thesis.

implemented in a range of contexts, it is important to understand how agencies can utilise recruitment and training to accurately identify people.

1.2.1 One-to-One and One-to-Many Face Matching Tasks

A common form of face matching is the one-to-one task. This involves comparing pairs of faces as either photo-photo, video-photo, video-person, or person-photo, where the aim is to verify whether the faces belong to the same individual (Spaun, 2007; Wechsler & Li, 2014). There is a process of individualisation, where all other people are excluded, or the possibility that two photos come from one person is eliminated (Spaun, 2007). One-to-many tasks aim to identify whether a person is within a group of several images. This is known as a candidates list (Spaun, 2007). More simply, one-to-one tasks are used to verify an individual's identity, whereas one-to-many tasks are used to identify an individual from a range of possibilities. Candidate lists can vary in the number of images presented, whereas one-to-one tasks remove the effect multiple possibilities might have on the observer. Consequently, research from one task should not be generalised to make inferences about the other. The one-to-one task will be the focus of the present study due to a lack of research in this area and its common use in operational contexts.

1.2.2 The Use of Automated Systems in Face Matching

Despite technological developments leading to the use of automated facial identification and verification systems (algorithms), a trained facial practitioner is often required to make the final decision (Graves et al., 2011; White, Dunn, Schmid, & Kemp, 2015; Phillips et al., 2018). Whilst machines have been found to perform with higher accuracy than humans on face matching tasks using unfamiliar faces (Lanitis, 2008), a recent study by Phillips et al. (2018), found the best performance was from a combination of one practitioner and one algorithm. In a face matching task using images unfamiliar to the practitioners and algorithms, the highest performing algorithm had a median accuracy of 96%, whereas the highest trained practitioner's median accuracy was reported at 93%. Consequently, due to practitioners still being required it is important to determine how to improve the performance of humans.

1.2.3 Familiar and Unfamiliar Face Matching

Face matching can be split into two broad categories; familiar and unfamiliar. Familiar faces are those known to an individual, whereas unfamiliar faces are defined as those an individual has either never seen before or has seen only a few brief times (Mandal et al., 2016). Findings show that people perform better when matching familiar faces (Ritchie et al., 2015; Kramer, Young, & Burton, 2018). Research has also found that increasing the number of images of a target in an unfamiliar candidates list, converted an unfamiliar face to a familiar one through learning, which increased accuracy (Dowsett, Sandford, & Burton. 2015). Therefore, researching familiar face matching cannot be generalised to unfamiliar face matching.

Unfamiliar face matching is used in numerous contexts. Because people are generally poor at matching unfamiliar faces, this has broad implications for national security (White, Kemp, Jenkins, & Burton 2014), including crime prevention and counter-terrorism (Kemp, Caon, Howard, & Brooks, 2016). Additionally, unfamiliar face matching is undertaken by law enforcement agencies to identify criminals, such as using video footage to identify or verify suspects (e.g., closed-circuit television) (Bruce, Henderson, Newman, & Burton, 2001). Demonstrating that matching unfamiliar faces is hard even on a local level; Kemp, Towell, and Pike (1997), found retail assistants accepted fraudulent credit cards that included facial identification 50% of the time even though they knew they were being watched. However, one limitation was that the targets also provided a signature, so it is difficult to know whether participants based their decision on their facial identification, or their signature matching that on the card. Considering the wide use of unfamiliar face matching tasks, understanding how people perform is important.

1.2.3.1 Human Difficulties Processing Unfamiliar faces

Research into why unfamiliar face matching is difficult has investigated how facial features are processed by people. The face contains two distinct types of information, featural and configural (Rhodes, 1988). The featural information are the elements of the face like eyes, nose and mouth (Carey & Diamonds, 1977), and configural information is the relationship between these features (Bruce, 1988). When it comes to features being useful in making judgements in unfamiliar face matching, many studies have found differences. Royer et al. (2016), found greater use of the nose and mouth, whereas other studies found the eyes to be more important (Caldara et al., 2005; Gosselin & Schyns, 2001; Schyns et al., 2002). Lastly, Towler, White, and Kemp (2017), found ears to improve performance, although Megreya and Bindemann (2018) did not support this finding. Generally, in training guidelines knowledge of the ear is recommended (e.g., FISWG, 2010).

It has been hypothesised that familiar and unfamiliar faces are processed differently (Bruce, Henderson, Greenwood, Hancock, Burton, & Miller, 1999; Hancock, Bruce, & Burton, 2000). Burton and Megreya (2006), found a high association between matching unfamiliar faces and completing the same task with inverted images. The researchers concluded unfamiliar face matching may be more like simple image matching (Burton & Megreya, 2006; Bruce et al., 1999; Hancock et al., 2000). Therefore, unfamiliar faces may be processed like images, not faces. Additionally, Kramer et al. (2018), found when viewing internal features (i.e., eyes and nose), rather than external features (i.e., hairstyle), face matching relied on the image similarity, hence why unfamiliar face matching is hard. These findings restate that performance on familiar face matching cannot be generalised to unfamiliar face matching, because the tasks involve different processes.

1.3 Challenges in Conducting Unfamiliar Face Matching Tasks

It has been established that humans are naturally poor at matching unfamiliar faces (White, Burton, Kemp, & Jenkins, 2013). Unfamiliar face matching is affected by lots of variables such as occlusion, pose, lighting (Burton, Miller, Bruce, Hancock, & Henderson, 2001), appearance of the face and time taken between the photos (i.e., ageing). Occlusion occurs when objects (i.e., accessories) block or disrupt view of the face (Mandal et al., 2016). Kramer and Ritchie (2016) found that wearing glasses lowered accuracy when only one of the pair of images contained glasses (74%), compared to no glasses in both (80.9%), and glasses in both (79.6%), indicating accessories could be an effective disguise. Pose refers to the perspectives of the face (Mandal et al., 2016). Estudillo and Bindemann (2014), found that university students performed with lower accuracy in one-to-one trials that had two different views (i.e., frontal and profile view), demonstrating consistency is important in these tasks. Appearance of the face may refer to what features are visible. Estudillo and Bindemann (2014), secondly found accuracy was higher when the full face was shown rather than just internal features using greyscale (i.e., black and white) images. However, Kemp et al. (2016) with similar methodology, found university students performed more accurately and faster when only shown internal features (84.5%), compared to a full face (82.1%), in trials that were deemed harder, using coloured images. Harder trials were those that had a deliberate change in appearance (e.g., hairstyle). In contrast, reported accuracies overall showed higher accuracy observing whole faces (87.4%), than internal features (83.6%), supporting Estudillo and Bindemann (2014). Lastly, many ageing processes such as wrinkles and scars impact human ability to compare faces (FISWG, 2010). Due to these variables being known to affect performance, many of them are controlled in usual unfamiliar face matching environments such as passport screening; although, variables like ageing cannot be controlled.

Research has investigated how working with other people improves performance. White et al. (2014), found that trial by trial feedback improved performance, and that performance generalised to other unfamiliar face matching tasks without feedback. Secondly, Dowsett and Burton (2014), found participants performing the Glasgow Face Matching Task (GFMT) were significantly better at conducting an image-image task in pairs. This indicates that whilst matching unfamiliar faces can be affected by many variables, having a mentor can improve performance.

1.3.1 The Impact of Image Pair Type on Performance

In the context of one-to-one face matching, the type of image pair displayed impacts performance. Research has found higher accuracy in mated pairs in some studies (Michalski, 2017; Burton and Megreya, 2006), and non-mated pairs in others (White, Kemp, Jenkins, Matheson et al., 2014). Mated pairs refer to pairs of faces of the same person, whereas non-mated refers to pairs of different people. A common example of mated pairs being challenging is that people are incorrectly not verified to match their passport. Conversely, an implication of non-mated pairs being challenging is that people are incorrectly verified as matching a passport that is not theirs. Ideally, there would be no discrepancy between pair types and people would be verified with similarly high accuracies. Differences in methodology and participant samples may be the cause of the opposing findings, so there is a need to study performance in a way that reflects usual work tasks. A review of the current research on novices and practitioners is outlined in Sections 1.4 and 1.5.

1.4 General Face Matching Performance of Novices Compared to Facial Practitioners

Of interest in the field of face matching is whether training can improve performance. Most organisations that utilise face matching implement training for their practitioners. Considering the extensive use of unfamiliar face matching to identify or verify individuals, facial practitioners must be reliable (White et al., 2013). A small amount of research has evaluated differences between novices and practitioners (e.g., White, Kemp, Jenkins, Matheson et al., 2014; Wirth & Carbon, 2017; White, Phillips, Hahn, Hill, & Toole 2015, White, Dunn, et al., 2015, Ferguson, 2015); however, the findings do not consistently show that practitioners perform better presumably due to methodological differences and limitations within the studies. Comparing performance is important as discrepancies found between novices and practitioners might show areas where training and experience has not helped practitioners improve. There is research on both novices and practitioners, yet few compare the groups under the same conditions.

White, Kemp, Jenkins, Matheson, et al. (2014), conducted a study with passport officers who had experience in face matching ranging from zero to 20 years. In a person-photo task, participants made comparisons to people and their photos in a condition resembling passport checks. Passport officers wrongly rejected 6% of the photos, and wrongly accepted 14%. There was no significant relationship between employment duration and face matching accuracy, although the researchers acknowledged that the experiment was easier than real life due to the nature of the photos used (White, Kemp, Jenkins, Matheson, et al., 2014). Within this study the participant sample size was small (N = 30), and the images were taken just a few days apart. The image set was severely restricted, with only 34 people of ethnically diverse backgrounds used to develop the conditions for 84 trials.

White, Kemp, Jenkins, Matheson, et al. (2014), also conducted a photo-photo comparison between students and passport officers. Overall performance on mated pairs was 70.9% and 89.4% on non-mated pairs, with no significant difference found between passport officers and students. However, although there was a larger sample size (22 passport officers and 38 students), this study shares the same limitations of the previous regarding the image set. Contrastingly, Wirth and Carbon (2017), also compared novices to passport officers on a similar task but found passport officers outperformed novices.

White, Phillips, et al. (2015), administered three one-to-one tests assessing expertise using an international cohort of forensic facial examiners (i.e., highly trained facial practitioners). Participants from the Facial Identification Scientific Working Group (FISWG) were tested on the GFMT, showing a significant difference between facial examiners, the control group (FISWG meeting attendees with no training in facial comparisons), and the student group. However, the GFMT has some limitations. The source of images were taken from a single university in Glasgow on the same day, cropped, and were presented in greyscale (Burton, White, & McNeill 2010). In the second test developed by White, Phillips, et al. (2015), where the facial stimuli were rendered so that leading algorithms made 100% errors on the set, the experts outperformed students. In the last test, using both upright and inverted faces, experts outperformed students. The researchers also found limiting the exposure time to 30 seconds showed higher accuracy by practitioners compared to students. The findings indicate learned facial examination techniques to observe faces utilised by practitioners help them outperform novices. However, an issue with this study is that dead-lining (i.e., giving participants a time limit to make their decision) is not typically how practitioners conduct facial matching in their usual work flow.

White, Dunn, et al. (2015), conducted a one-to-many study comparing facial examiners, facial reviewers, and a control group of novices, aiming to find if experience and training improved performance. Facial reviewers are practitioners with a lower level of training than facial examiners (FISWG, 2010). This study used images ranging from ages six to 47 years old on a one-to-many task. The results reported facial examiners to perform with 69.1% accuracy, facial review staff 48.1%, and the novices 47.6%. Examiners performed significantly better than the reviewers and the novices, but there was no difference between reviewers and novices. The researchers concluded that experience conducting face matching and being surrounded by other facial examination specialists helped them develop effective strategies for the task. However, this study has a significant limitation in that participants were dead-lined to 18 seconds per trial. It is possible that facial reviewers would perform better if they undertook the task with no time constraints. Additionally, no statistical analyses for the group differences in accuracies for child, adolescent and adult conditions was done. Thus, the researchers missed a significant opportunity to compare the participant groups on whether they differ matching child compared to adolescent and adult faces.

A study by Ferguson (2015), compared practitioners to untrained participants specifically on images of children in a one-to-one study and found there was no difference in ability when matching child faces, with an average accuracy across groups of 61%. However, this study had 20 uncontrolled pairs of images which is a very restricted dataset. Therefore, as stated by Ferguson (2015), there is a need to conduct research using child images with a controlled dataset.

This body of research suggests that practitioners generally outperform novices on face matching tasks, presumably because they have more training and experience and motivation to perform well. Although, not every study produces significant differences supporting these conclusions. The overall research on performance when matching faces lacks consistency, either from differing methodologies or limitations within the study (e.g., Kemp et al., 1997; White, Kemp, Jenkins, Matheson, et al., 2014; White, Phillips, et al., 2015; Ferguson, 2015). One major flaw with this body of research is that dead-lining does not represent tasks that practitioners usually undertake (e.g., White, Phillips, et al., 2015; White, Dunn, et al., 2015), and studies have used images unrealistic to operational settings (e.g., Estudillo & Bindemann 2014; White, Phillips, et al., 2015; Ferguson, 2015). The inconsistency in these results suggest that more research needs to be undertaken on human face matching ability because there are implications for training and recruitment if novices and practitioners perform similarly. An area of current interest that is largely overlooked are the performance differences when comparing faces of children opposed to adults.

1.5 Performance Matching Child Faces

As part of their business processes, many organisations are required to conduct face matching on children. Aside from usual verification checks (i.e., border control), images or videos of faces may need to be compared in forensic situations including child exploitation cases (Kramer, Mulgrew, et al., 2018). Consequently, being able to accurately verify or identify children is important. However, research indicates child faces are distinctly hard to match. There are significant facial changes related to age that impact the shape and texture of the face (Wei & Li, 2017), and changes to the face in childhood are quantitatively and qualitatively different to changes occurring in adulthood (White, Dunn, et al., 2015). Additionally, younger children have less discriminating facial features making them more difficult to distinguish from each other (Wilkinson, 2012), and the level of craniofacial growth and development during the childhood years is significant, leading to distortion of the face (Kozak, et al., 2015; White, Dunn, et al., 2015). Research has also found that girls demonstrate fewer changes between the ages of 12-18 than boys (Chakravarty, Aleong, Leonard, and Perron, 2011). Considering these factors, it is significant that performance comparing children's faces has been largely overlooked in the research on face matching.

Supporting the notion that facial development in childhood makes children specifically hard to verify, Kramer, Mulgrew, et al. (2018), conducted a one-to-one task using faces of one infant (age < 1) and one child (age 4-5). The results showed that university students had low accuracy (64%). In comparison, the researchers found in a photo-photo task with two infants, accuracy was 72%. However, the infant-infant task used a highly uncontrolled dataset using images taken by parents with no record of how long there was between images, and the infant-child task used greyscale images of celebrity's children found through an internet search. Due to these limitations, a more controlled study should be conducted to make conclusions.

Several studies have indicated that humans find matching the faces of children much more difficult than that of adults. Zeng et al. (2012), implemented a one-to-one unfamiliar face matching task with students. The accuracies ranged from 77.5-78.9% for children images under 18, 74.3-82.1% for images of ages 30-39 years, and 75% for image ages 40-68. Whilst the accuracies were higher in the under 18 condition than the 40-68 condition, a limitation with this study is that no statistical analyses were undertaken between the groups, and there was a wider range of ages in the aged 40-68 condition. Thus, no significant differences can be concluded.

White, Dunn, et al. (2015), conducted a one-to-many face matching task with university students comparing accuracy variance due to the age of the target face. Overall there was poor accuracy with only 45.1% correct in the adult condition (ages 40 to 47), 41.1% correct in the adolescent condition (ages 14 to 22), and 39% in the child condition (ages six to 13), where the main effect of age on overall accuracy was significant. Participants made significantly fewer correct matches for children compared to adolescents, and adolescents compared to adults. Children and adolescents were also misidentified more often than adult targets. However, there was not a significant difference in correct rejections for children compared to adults, or in incorrect matches for adolescents compared to adults. From these results the researchers concluded that error rates being poor for images of children and adolescents was indicative of the fast rate of growth of facial structures and features in childhood (White, Dunn, et al., 2015). This study has limitations in that the task was dead-lined to one hour, and six people did not complete the experiment, so performance was affected by the time restraints.

Practitioners have also been found to have worse performance when matching children faces than adult faces in a photo-photo task (Michalski, 2017). In one of the largest studies evaluating performance of reviewers on child and adult images using a controlled operational image dataset, Michalski (2017) found lower accuracy performance in relation to child images (73.9%), than adult images (92.1%). Secondly, confidence was significantly lower and response times significantly higher (there was no dead-lining) in relation to children’s images. Michalski (2017), also found that performance was significantly poorer in the non-mated image groups. This does not support White, Kemp, Jenkins, Matheson, et al. (2014), who found performance was better in non-mated conditions. This indicates experimental conditions or sample population may affect performance depending on the pair type.

This body of research suggests that children's faces are harder to match than adult's, even for practitioners. Therefore, accuracy rates for adults should not be generalised to children. For some agencies, currently there may be no specialised training for practitioners that addresses the significant facial change of children (e.g., FISWG, 2010). There is a significant lack of empirical studies evaluating the performance of novices matching images of children compared to adults that represents usual work tasks. Critical to improving face matching performance is identifying areas that are naturally harder and may require specialised, targeted training. Reviewing the research also shows that the general limitations are extended to age-specific face matching studies (Zeng et al., 2012; White, Dunn, et al., 2015; Ferguson, 2015; Kramer, Mulgrew, et al., 2018). Considering the limitations in current research, there is a need to evaluate people's natural ability to verify the identity of children in unfamiliar face matching tasks.

1.6 The Present Study

Due to the finding that conducting facial comparisons on children is harder than on adults for practitioners, yet a significant lack of empirical research on novices, the present study will have a primary aim to assess university students on an unfamiliar face matching task. The study will use a similar methodology to Michalski (2017), which previously examined the face matching performance of facial reviewers on controlled operational images of children and adults. Doing so will broaden knowledge of unfamiliar face matching, by using the same images with a different participant population. This also provides the possibility of statistical comparison between the datasets in the future. Because facial practitioners are usually trained with minimal regard to different age categories, the performance of novices has potential implications in a real-world setting. If novices perform worse comparing images of children than images of adults, this would demonstrate performance differences may stem from natural challenges in human ability.

Additionally, this would indicate there is a need to implement specific training for comparing children's faces. Due to research suggesting performance in mated conditions has been found to be better in some studies (e.g., Michalski, 2017; Burton & Megreya, 2006), but worse in others (e.g., White, Kemp, Jenkins, Matheson, et al., 2014), the present study will also assess the performance differences between mated and non-mated pairs.

1.6.1 Research Question

To what extent is there a difference in novice performance in a face matching task involving both mated and non-mated images of adults and mated and non-mated images of children?

1.6.1.1 Hypotheses

Due to previous research on face matching involving images of adults and children, and research on mated and non-mated conditions, the following hypotheses were tested:

1) Students will perform better (i.e., more accurately, with higher confidence and faster) on a face matching task on images of adults than images of children.
2) Students will perform better (i.e., more accurately, with higher confidence and faster) in mated trials than non-mated trials.

Chapter 2: Method

2.1 Ethics Statement

This study was approved by the Defence Science and Technology Group (DST) Human Research Ethics Committee (NSID 03) and the University of Adelaide Human Research Ethics Sub-Committee (18/57). Participants were provided with a University of Adelaide Participant Information Sheet (see Appendix A), a DST Participant Information Sheet (see Appendix B) and provided their consent (see Appendix C and D) before commencing the study.

2.2 Participants

The participants (*N* = 35) were students from Australian Universities. The sample included 11 males and 24 females, ranging from 18-28 years old (*M* = 21.3). The majority identified as Caucasian (86%), with a minority identifying Asian (11%) and Middle Eastern (3%). Participants had no training or experience conducting facial comparisons. To participate in the study participants had to be undergraduate students, proficient in English, over 18, and have normal or corrected-to-normal vision.

Seven first year University of Adelaide Psychology students were recruited through the Research Participation System at the University of Adelaide and received one course credit for participating. Further students were recruited with posters (see Appendix E) around the North Terrace campus at the University of Adelaide, social media posts, and through word of mouth. Participation was voluntary, and participants could withdraw at any time.

2.3 Design and Measures

A within-subjects repeated measures design was used in the study. This allowed the performance measures to be split into different conditions of the independent variable. A repeated measures design is where the same participants are tested in each condition of the independent variable. The independent variable was the type of image supplied, where the conditions were whether the images were of children or adults, and whether the images were mated or non-mated. Mated images were defined as a pair of faces that belonged to the same person, whereas non-mated images were a pair of faces that belonged to different people. The dependent variables were the performance measures of the participants; these were accuracy, confidence and response time. Signal detection measures were used to analyse correct hits (correct match decision) against false alarms (incorrect match decision) for the child and adult conditions. To combat order effects, the conditions in the study were randomised.

2.4 Materials

2.4.1 Experimental Application

The computer-based application was written by DST, designed to replicate the application used in Michalski (2017), simulating computer screen layouts facial practitioners might be accustomed to. There were 200 pairs of images, divided equally into trial types based on the conditions of the independent variable. The trial types were; mated pairs of adults (50 images), non-mated pairs of adults (50 images), mated pairs of children (50 images) and non-mated pairs of children (50 images). The image pairs contained an age range of up to 10 years. These trials were randomised for each participant. For an example of the layout of the experimental interface during the task see Figure 2.1. Before the experiment began participants were asked several demographic questions (gender, age, ethnicity), as well as their experience with face matching and their exposure to children, followed by two practice examples before the experiment began. At the end of the experiment, participants were asked general questions about decision making, and what aspects of the task they found challenging.

Figure 2.1 provides an example of experimental interface during a practice trial. These are not the operational images from the experiment.

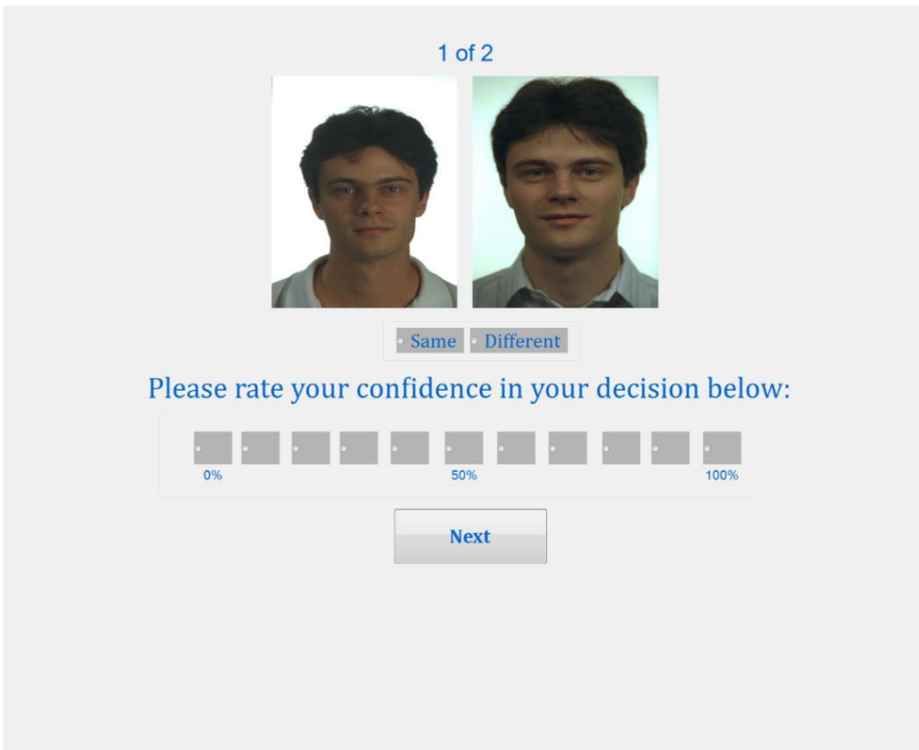


Figure 2.1. Experimental Interface.

2.4.2 The Image Database

The same controlled operational facial image dataset that was used in Michalski (2017), was used in this study. The images in this database were passport style images controlled in terms of being frontal face images of a consistent size with even lighting and neutral expressions. The age of children was defined as being of ages 0-17 years old, whereas adults were defined as

18 and above. Pairs in this dataset were taken up to ten years apart, addressing limitations of previous research that collected datasets specifically for that research, meaning there tended to be image pairs with minimal elapsed time between photos. The average age variation between image pairs was similar for each group (mated adult pairs = 5.34 years, non-mated adult pairs = 5.42 years, mated child pairs = 5.54 years, non-mated child pairs = 5.78 years,). Due to the test images being operational, examples of each pair type from the experiment cannot be provided. Using an operational dataset is beneficial in that the results are more representative of real-life performance.

2.5 Procedure

The study was undertaken in the controlled laboratory room 219 of the Hughes Building at the University of Adelaide. The experimental application was viewed on computer monitors with a resolution of 1920x1080. Each participant was assigned a unique identification number at sign up to ensure anonymity. Participants were given an information sheet (see Appendix A), along with the DST Guidelines for Volunteers (Appendix B) to read before the experiment began. Before proceeding, the information sheet was verbally explained to participants so that they understood what was expected of them and how their data would be managed following the experiment. Participants then indicated their consent and willingness to participate by clicking so on the first screen of the experimental application (See Appendix C).

The following two screens were basic demographic questions on gender, age, ethnicity and experience with children; as well as asking participants to insert their identification number. Written instructions for the task were then presented:

"You will be presented with pairs of facial images. Your role is to look at the pairs of images from the perspective of an employee processing passport applications and decide whether the images are of the SAME person or DIFFERENT people. Please work as quickly and accurately as possible. Once you have made your decision, you will not be able to change it. You will then be required to rate your confidence in your decision.".

These instructions were presented to ensure that each participant conducted the task with the assumption that they were in an operational environment, and generally acted in the same way. Following this statement, two practice examples were provided to ensure that participants were familiar with the requirements before the experiment began. This involved a pair of images being presented for the participants to decide whether the images were of the same person or different people. Once deciding, participants were asked to rate their confidence from 0 to 100 percent, in 10 percent increments. The response time for decision making was also collected as both a measure of performance, and data checking. Participants then completed 200 image pair comparisons.

After the experimental stimuli had been presented, participants were able to provide feedback on the study and describe their decision-making process. Participants were then offered the chance to to receive their results by providing an email address. Each session took approximately 40 minutes.

Chapter 3: Results

3.1 Data Screening, Assumptions, and Test Selection

Prior to data analysis the data were screened to assess normality and check for missing data. Statistical analyses were run on the proportion of correct responses (accuracy), the proportion confidence levels, and median response times. This gave a single statistic for each performance measure for each condition of the independent variable, for each participant. Descriptive statistics were reported as notched boxplots, an interpretation is provided (see Appendix F). A signal detection table was also reported (see Appendix G). Histograms and quartile-quartile plots in conjunction with Shapiro-Wilk tests were used to assess normality and found that much of the data was significantly skewed. As such, non-parametric tests were used. Friedman ANOVA tests were used to compare more than two conditions. If the results were significant, Wilcoxon Signed Rank tests were used to determine statistically significant differences between groups. Justification for using non-parametric tests instead of data transformation comes from past research that shows skewness is typical on variables such as accuracy and response time (Burton et al., 2010). The standard Bonferroni adjustment was used to protect against Type 1 error (Shaffer, 1995). This was reported in a significance level set at $p < .025$ when making pairwise comparisons between groups for each performance measure for the first hypothesis, and $p < .013$ for the second hypothesis. An alpha level of .05 was set for all analyses. Effect sizes for the Wilcoxon Signed Rank tests were calculated using the formula $r = \frac{Z}{\sqrt{N}}$ and interpreted using Cohen (1988) criteria as small ($r \geq 0.1$), medium ($r \geq 0.3$) and large ($r \geq 0.5$) effects. Secondly, signal detection measures were used to analyse discriminability for the child and adult conditions for hits and false alarms. The signal detection analysis was conducted using a simple Principle Components Analysis (PCA) approach outlined by Vokey (2016) and implemented in R (see Appendix H for the output from functions defined in Vokey, 2016). This approach is advantageous because it does not require specification of a likelihood function for fitting a Receiver Operating Characteristic (ROC) using maximum likelihood estimation. It provides stable *d'* (d-prime) based on the pivot point of the fitted line according to the first principle component, which eliminates the problems associated with using non-parametric versions of *d'*. Given the lack of hits in the four lowest confidence categories – these were collapsed into a single rating.

3.2 Performance with Images of Children and Adults

Figure 3.1 presents the descriptive statistics for accuracy for the Child and Adult groups as a proportion of correct decisions.

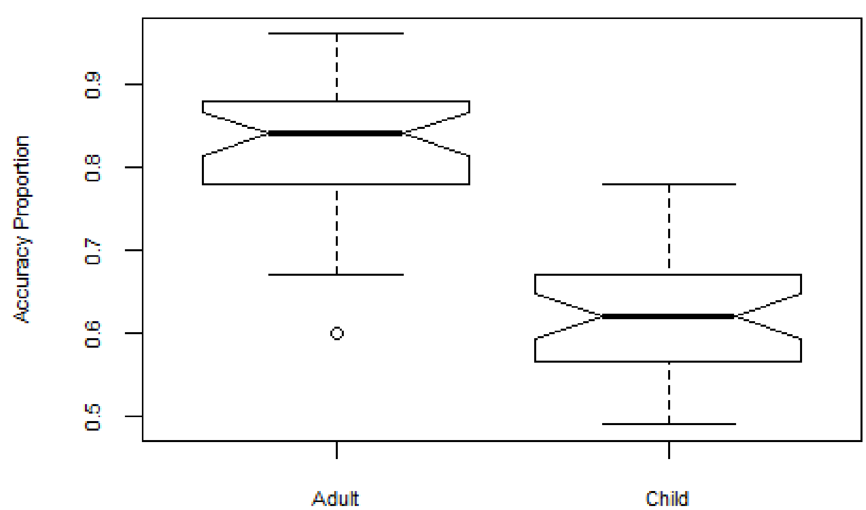


Figure 3.1. Overall accuracy for the Child and Adult image groups.

The Wilcoxon Signed Rank test revealed a significant decrease in overall accuracy for the Child group (*M* = 62%, *Mdn* = 62%), compared to the Adult group (*M* = 83%, *Mdn* = 84%), $Z = -5.16$, $p < .001$, $r = -.62$. This indicates a large effect size.

Figure 3.2 presents the descriptive statistics for confidence in decisions for the Child and Adult groups.

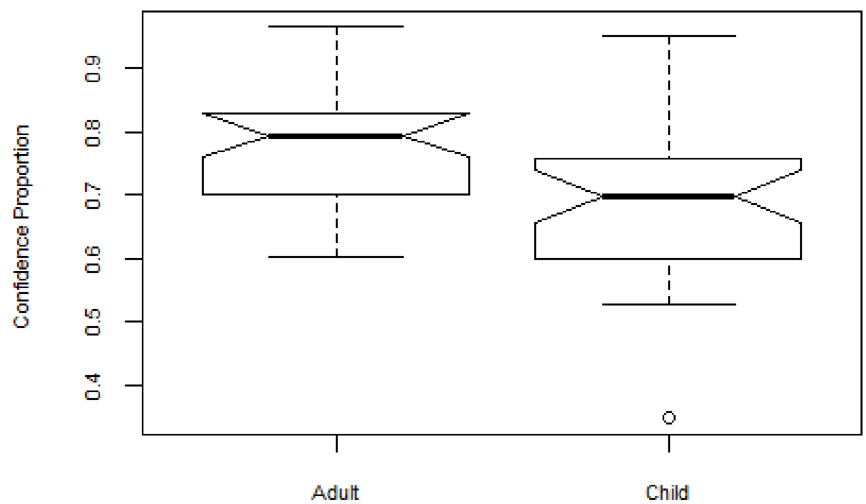


Figure 3.2. Confidence levels for the Child and Adult image groups.

The Wilcoxon Signed Rank test revealed a significant decrease in overall confidence for the Child group (*M* = 69%, *Mdn* = 69.8%), compared to the Adult group (*M* = 78%, *Mdn* = 79.4%), $Z = -5.14$, $p < .001$, $r = -0.61$. This indicates a large effect size.

Figure 3.3 presents the descriptive statistics for median response times for the Child and Adult groups.

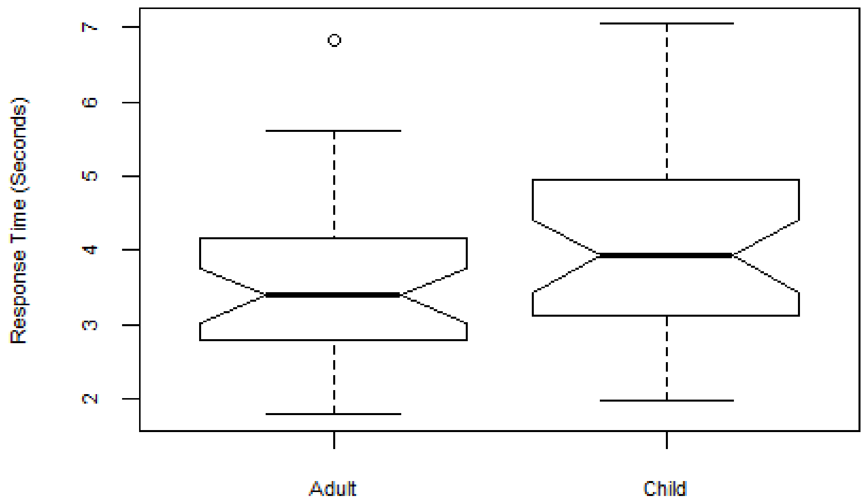


Figure 3.3. Median response time for the Child and Adult image groups.

The Wilcoxon Signed Rank test revealed a significant increase in overall response time for the Child group (M = 4.12 seconds, Mdn = 3.93 seconds), compared to the Adult group (M = 3.54 seconds, Mdn = 3.39 seconds), Z = -5.05, p < .001, r = -0.60. This indicates a large effect size.

3.2.1 Signal Detection Analysis

Figure 3.4 presents the empirical ROC for performance using Child and Adult images.

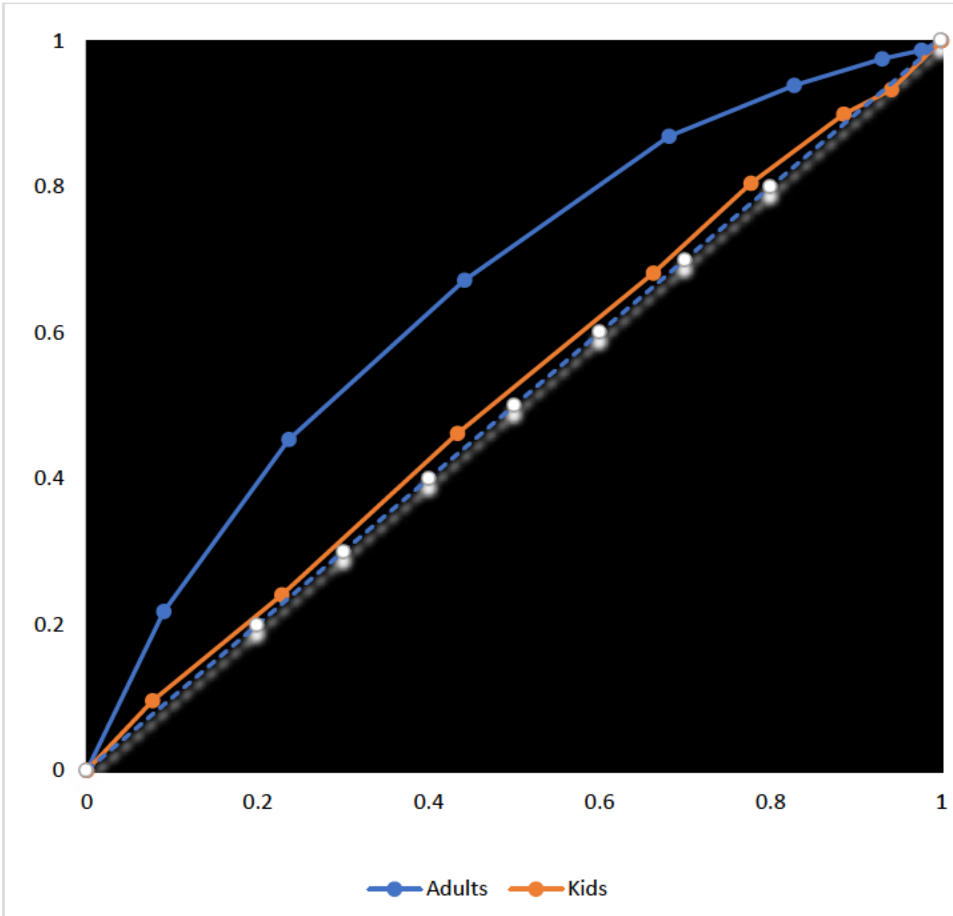


Figure 3.4. Empirical ROC for performance using Child and Adult images (dashed line shows chance performance).

Using the PCA approach it was apparent that participants were significantly better at discriminating a match (as measured by the area under the curve metric *As*) when the images were of adults compared with children, *As* = 0.66 versus *As* 0.49, respectively.

In summary, the results support the hypothesis that novices perform better conducting face matching on images of adults than children. When the novices made decisions on images of children, they were less accurate, less confident, and slower.

3.3 Performance with Images of Children and Adults on Mated and Non-Mated Pairs

Figure 3.5 presents the descriptive statistics for accuracy for the mated and non-mated Child and Adult groups as a proportion of correct decisions.

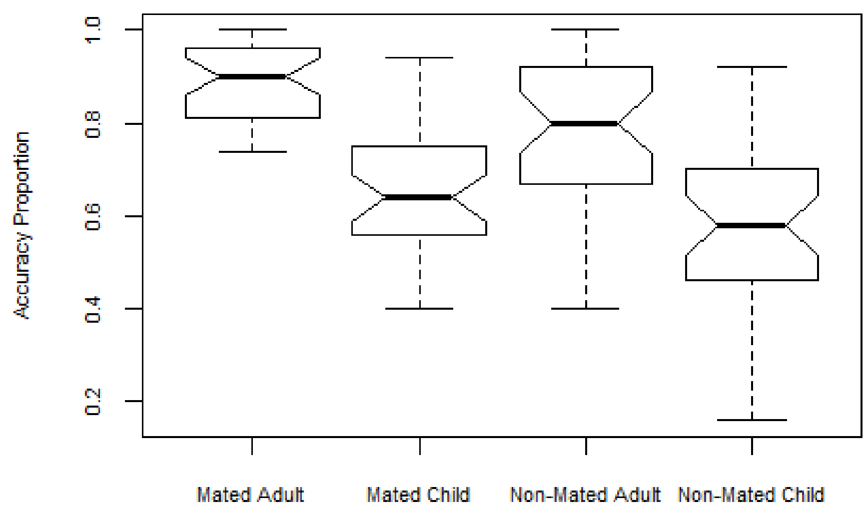


Figure 3.5. Overall accuracy for mated and non-mated Child and Adult image groups.

A Friedman ANOVA showed that there were statistically significant differences in accuracy based on the type of image presented, $\mathcal{X}^2$ (3), = 61.76, $p < .001$.

Post hoc analysis with the Wilcoxon Signed Rank test revealed a significant decrease in accuracy for the mated Child group (M = 66%, Mdn = 64%), compared to the mated Adult group (M = 88%, Mdn = 90%), Z = -5.16, $p < .001$, r = -0.62, and a significant decrease in accuracy for the non-mated Child group (*M* = 57%, *Mdn* = 58%), compared to the non-mated Adult group (*M* = 77%, *Mdn* = 80%), $Z = -5.12$, $p < .001$, $r = -0.61$. Lastly, there was a significant decrease in accuracy for the non-mated Adult group compared to the mated Adult group, $Z = -2.60$, $p = .009$, $r = -0.31$, and a non-significant decrease in accuracy for the non-mated Child group compared to the mated Child group, $Z = -1.73$, $p = .084$, $r = -0.2$. The results indicate a large effect size for the first two relationships, and a medium effect size for the third.

Figure 3.6 presents the descriptive statistics for confidence for the mated and non-mated Child and Adult groups.

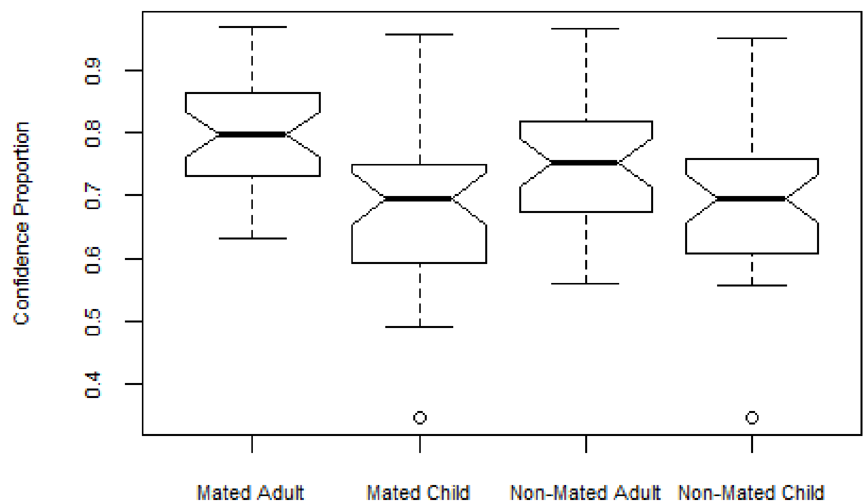


Figure 3.6. Confidence levels for mated and non-mated Child and Adult image groups.

The Friedman test showed that there were statistically significant differences in confidence based on the type of image presented, $\mathcal{X}^2$ (3), = 66.29, $p < .001$.

The Wilcoxon Signed Rank tests revealed a significant decrease in overall confidence for the mated Child group (*M* = 69%, *Mdn* = 69.4%), compared to the mated Adult group (*M* = 80%, *Mdn* = 79.6%), *Z* = -5.16, *p* < .001, *r* = -0.62, and a significant decrease in confidence for the non-mated Child group (*M* = 69.2%, *Mdn* = 69.4%), compared to the non-mated Adult group (*M* = 75.5%, *Mdn* = 75.2%), *Z* = -4.70, *p* < .001, *r* = -0.56. Lastly, there was a significant decrease in confidence for the non-mated Adult group compared to the mated Adult group, *Z* = -3.16, *p* = .002, *r* = -0.38, and a non-significant decrease in confidence for the non-mated Child group compared to the mated Child group, *Z* = -0.97, *p* = .330, *r* = -0.1. The results indicate a large effect size for the first two relationships, and a medium effect size for the third.

Figure 3.7 presents the descriptive statistics for median response times for the mated and non-mated Child and Adult groups.

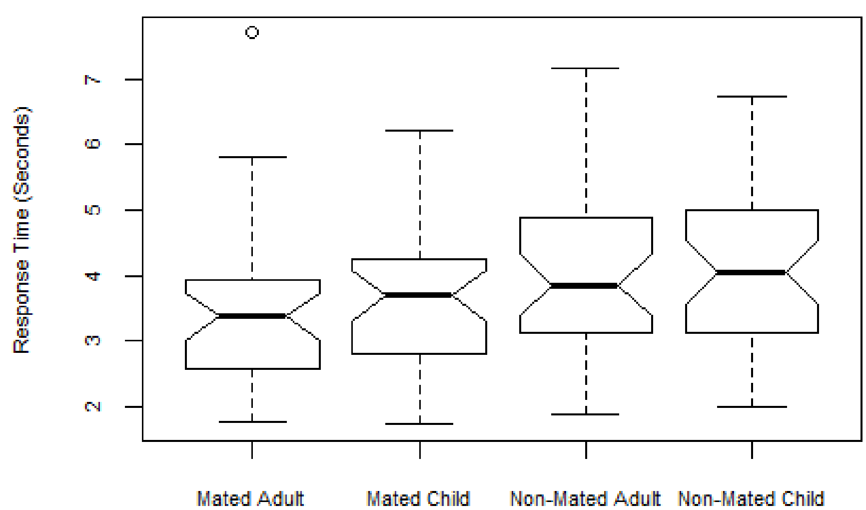


Figure 3.7. Median response time for mated and non-mated Child and Adult image groups.

The Friedman test showed that there were statistically significant differences in response time based on the type of image presented, $\mathcal{X}^2$ (3), = 39.59, p < .001.

The Wilcoxon Signed Rank test revealed a non-significant increase in median response time for the mated Child group (M = 3.74 seconds, Mdn = 3.69 seconds), compared to the mated Adult group (M = 3.45 seconds, Mdn = 3.36 seconds), Z = -2.20, p = .028, r = -0.26, and a non-significant difference in response time for the non-mated Child group (M = 4.17 seconds, Mdn = 4.05 seconds), compared to the non-mated Adult group (M = 4.03 seconds, Mdn = 3.86 seconds), Z = -1.94, p = .051, r = -0.23. Lastly, there was a significant decrease in response time for the non-mated Adult group compared to the mated Adult group, Z =-4.39, p < .001, r = -0.52, and a significant decrease for the non-mated Child group compared to the mated Child group, Z = -3.10, p = .002, r = -0.37.

Whilst some performance measures were not significantly different in some groups, the results partially support the hypothesis that novices would perform better in mated conditions.

Chapter 4: Discussion

This study aimed to understand differences in one-to-one unfamiliar face matching performance of novices when the target images were of children or adults, and when the image pairs were mated or non-mated. Previous research has primarily focussed on general performance in unfamiliar face matching. A small proportion of research has analysed facial practitioner performance on child faces (e.g., Michalski, 2017), and the few studies on novice performance matching child faces have limitations (e.g., Zeng et al., 2012; White, Dunn, et al., 2015, Ferguson, 2015; Kramer, Mulgrew, et al., 2018). Secondly, previous research has indicated performance on mated and non-mated pairs is highly dependent upon the study methodology. The results mostly supported the two hypotheses.

4.1 Overview of Performance on Child and Adult Conditions

Performance was firstly compared generally between images of children and adults. Overall, face matching was more accurate, performed with higher confidence and faster when viewing images of adults. This supports the first hypothesis. The results share similarities with Michalski's (2017) finding that practitioners perform better when matching adult faces (i.e., higher accuracy, higher confidence and faster), and are comparable with other previous research (Zeng et al., 2012; White, Dunn, et al., 2015). This indicates that matching faces of children may be naturally harder because of the significant craniofacial development during childhood (Kozak et al., 2015; White, Dunn, et al., 2015), and children having fewer discriminating features which makes them harder to distinguish them from each other (Wilkinson, 2012). The reason novices were less confident and slower when matching children's faces indicates they might also perceive the task to be harder than matching adult faces. The signal detection analysis supported this conclusion, with discriminability being worse when the images being compared were of children. The values of *c* (Appendix H) indicated more conservative responding when comparing images of children.

4.2 Overview of Performance on Mated and Non-Mated Pairs

Performance was also analysed by image pair type (i.e., mated and non-mated) on each group (i.e., adult and child). Overall performance was better when matching mated pairs compared to non-mated pairs for the adult group (i.e., accuracy, confidence and response time). Response times were significantly faster for the mated child compared to the non-mated child group, although accuracy and confidence differences were both non-significant. These results partially support the second hypothesis. The findings indicate that when viewing adult faces, mated pairs are easier to verify than non-mated pairs (i.e., higher accuracy, confidence and response time), although for child pairs there are no differences in confidence or accuracy. However, due to the response times being slower for the non-mated child group compared to the mated child group, this implies novices also perceived this task to be more difficult. In terms of performance in the adult conditions, the results are comparable to some previous research (Michalski, 2017; Burton & Megreya, 2006), and oppose others (e.g., White, Kemp, Jenkins, Matheson, et al., 2014; White, Dunn, et al., 2015).

Accuracy and confidence for mated adult pairs was better than mated child pairs, and for non-mated adults compared to non-mated child pairs. This again supports the first hypothesis and shares similarities to research on practitioners (Michalski, 2017), reinstating the notion that child faces are naturally harder to verify, across pair types. However, the difference in response times for both comparisons in the present study were non-significant, indicating that time taken to decide might have been highly reliant on whether the images were mated or non-mated. Further analysis is needed to understand this difference.

4.3 Limitations

There are some general limitations to discuss about the methodology of the present study. Due to the analysis simply comparing children and adults, there was no discrimination between more specific age ranges. This is especially important with children because of the significant amount of facial change during childhood (e.g., infants compared to nine-year old), that is different to the amount of facial change during adulthood (White, Dunn, et al., 2015). Thus, the present study does not show whether some age ranges within the child condition specifically impacted performance more than others.

Another limitation was the participant sample. Students do not represent the total population of novices in the real world. Whilst students can be all ages, they are typically young adults (participants in the present study had a mean age of 21.3 years). Different levels of life experience, for example having children or not, might impact a person's ability to discriminate between two child faces. Additionally, research has shown that older adults (e.g., 65 years of age) are less accurate in unfamiliar face matching tasks than younger adults (Megreya & Bindemann, 2015). Therefore, including a wider age range of participants could impact the performance distributions. The argument could be made that at age 65 many practitioners would be retired, although research could still analyse at what point the level of performance naturally decreases. An important point to note is that practitioners in previous studies would be motivated to perform accurately as the task represents an aspect of their job. As novices in the present study did not have a reason to do well, there is a potential limitation that the novices did not try as hard as they could.

4.4 Strengths

The present study addressed limitations in the methodologies of the general research on unfamiliar face matching. Previous studies dead-lined participants, limiting how long they had to decide whether the pair was a match or no match (e.g., White, Phillips, et al., 2015; White, Dunn, et al., 2015), used greyscale images or images unrealistic to operational contexts (e.g., Estudillo & Bindemann, 2014; White, Philips, et al., 2015), including taking the images on the same day or only a few days apart (White, Kemp, Jenkins, Matheson, et al., 2014). These are major flaws when making inferences about human ability in the operational context. Face matching in national security contexts typically involve controlled, coloured images taken up to 10 years apart (or longer), with no dead-lining, so performance in these studies may not be representative of usual work tasks. Thus, the present study appears to be one of the first comprehensive, realistic analyses of the performance of novices on a one-to-one unfamiliar face matching task with child images.

Few studies have compared the performance of novices on images of children. However, these also contain methodological limitations (e.g., White, Dunn, et al., 2015, Ferguson, 2015; Kramer, Mulgrew, et al., 2018), or did not conduct statistical analyses on the data (e.g., Zeng et al., 2012). Therefore, a significant gap in the literature about verifying children has been explored in the present study. The present study has demonstrated there are mixed performance differences between pair types (e.g., adult or child faces, and mated or non-mated pairs), which might stem from natural challenges, accompanied by a lack of experience and training. These results provide a more specific insight into human ability in unfamiliar face matching with information that can be used in future implementation of training.

More broadly, some research has suggested that practitioners do not out-perform novices (e.g., White, Kemp, Jenkins, Matheson, et al., 2015; White, Dunn et al., 2015). However, other researchers are determined to show that practitioners do in fact outperform novices, especially in tasks most alike operational activities. A major strength of the present study is sharing similar methodology to Michalski (2017), because the data collected from the two studies can be statistically compared in future research as discussed in Section 4.6. Unfortunately, due to limitations in word count this could not be explored as part of this thesis.

4.5 Implications

The present study has demonstrated that people with no training or experience matching images of faces perform with significantly lower accuracy, confidence, and slower when the images are of children than adults. This shows that verifying children in one-to-one face matching tasks is naturally hard to do, most likely due to children having less discriminating facial features (Wilkinson, 2012), and the impact of craniofacial development (Kozak et al., 2015). As previously mentioned, children are subject to numerous forms of exploitation, so it is a significant finding that they are naturally harder to verify because there are implications in terms of child wellbeing. Comparative to the present study, facial practitioners also struggle to verify children in one-to-one face matching tasks (e.g., Michalski, 2017). This suggests training may not adequately address the impact of child-specific ageing processes and having less discriminating facial features. An example of one set of current guidelines for training is FISWG, this group acknowledges that age changes the face, although the focus is on age variables such as wrinkles and weight gain (FISWG, 2010). Considering the present study's findings, measures should be taken in training to address the fact that children are naturally harder to verify.

Realistically, it would be preferable that people perform the same with mated and non-mated pairs but with high accuracy as outlined in Section 1.3.1, as this would mean people are being accurately verified. Whilst adults were easier to verify in the mated compared to non-mated condition (i.e., more accurate and confident), novices performed with no significant difference in accuracy or confidence between mated and non-mated child image pairs, meaning there was no discrepancy in ability to discriminate between child faces belonging to the same or different people. This demonstrates there is a natural difference in ability to verify that two adult faces belong to the same person, compared to not belonging to the same person, but this is not the case with children. In conjunction to performance comparing child faces being low in general, this is significant in the context of uncovering child exploitation practices, as children are naturally hard to discriminate from each other (Wilkinson, 2012). Comparatively, practitioners have been found to have higher accuracy and confidence in mated pairs for adult and child faces, compared to non-mated pairs respectively (Michalski, 2017). This is possibly due to experience, because in day to day activities practitioners are more likely to be in situations where they constantly make a match decision (e.g., passport checks). As novices have no experience or training, it is logical that they may not show these same differences.

Considering the findings, future methods to improve performance with new recruits could include using a mentor. Research has shown that performing in pairs (Dowsett & Burton, 2014), and trial by trial feedback (White et al., 2014) both increase performance comparing unfamiliar faces. Using a mentor in the early stages of training may help practitioners develop better skills for discriminating between child faces.

4.6 Suggestions For Future Research

The general aim of the present study was to conduct an analysis on novice performance in an operationally realistic scenario, prior to any training or experience. Past research has suggested that practitioners do not outperform novices in face matching tasks (White, Kemp, Jenkins, Matheson, et al., 2014), yet other studies have found practitioners to perform better (e.g., White, Phillips, et al., 2015; Wirth & Carbon, 2017). These studies have methodological limitations, consequently they may not represent true practitioner capabilities. This is a major flaw in current research on face matching performance. As such, there is a significant need to produce quality empirical research between practitioners and novices. Therefore, the most likely future research to be undertaken will be to conduct a statistical analysis between the present study and Michalski (2017). This comparison would be an extensive, comprehensive and realistic analysis of performance between people with and without training or experience on a one-to-one face matching task. This would address the implications found in the present study about natural challenges in human ability, by assessing whether practitioners are better at matching images of children than novices. Additionally, this would provide a general comparison of performance between practitioners and novices. Therefore, likely implications of this analysis would include determining whether training and experience improves human ability to verify children, and in performance in general.

As previously mentioned, literature has shown that females display fewer facial changes between the ages of 12 and 18 years (Chakravarty et al, 2011). One final limitation of the present study was not addressing this. Future research should analyse whether there is a difference in face matching performance when viewing images of children dependent on their gender, which could be done using the present studies data as gender was recorded for each image pair. This analysis would show whether gender based facial changes impact the verification of children in a face matching context. Additionally, if a difference is found this could also be compared to Michalski (2017), to see whether training and experience aids practitioners to verify children of both genders with similar accuracy.

Addressing one limitation of the present study, due to child passport photos needing to be updated every five years, future research should compare performance differences making decisions on child images where the photos are taken up to five years apart, compared to 10. Michalski (2017), found that practitioners only performed significantly better (images between ages 0–4, 5–10, 11–15 years) when images were taken up to five years apart rather than 10, but only in the non-mated condition (the mated condition only had a 1% difference). However, due to the present study finding novices to perform with no significant difference between the mated and non-mated child groups, it would be beneficial to research further considering the current standards for child passports.

4.7 Conclusion

Being able to verify children is critical in a range of contexts. Not only is this important in terms of local law enforcement, and national security for normal verification checks, but to aid in the prevention of child exploitation and radicalisation. The present study demonstrates that children are naturally harder to verify in one-to-one unfamiliar face matching tasks. This is significant because of the common use of unfamiliar face matching. The present study has also found that it is easier to determine two adult faces belong to the same person, than to determine they are different people, but there is no difference comparing child faces. Future research should compare the performance of novices to trained facial practitioners, to investigate whether training and experience improves the verification of children, and in overall performance in one-to-one face matching tasks.

References

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).

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2015). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, *30*(1).

Bruce, V. (1988). Essays in cognitive psychology. Recognising faces. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc

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

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

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

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

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behaviour Research Methods*, *42*(1), 286-291.

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

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195,* 312-314. doi: 10.1126/science.831281

Cohen, J. (1988). Statistical Power Analysis for the Behavioural Sciences. *Hillsdale, New Jersey: Lawrence Erlbaum Associates, 2nd.*

Chakravarty, M. M., Aleong, R., Leonard, G., & Perron, M. (2011). Automated analysis of craniofacial morphology using magnetic resonance images. *PLoS ONE, 6*. doi: 10.1371/journal.pone.0020241

Department of Foreign Affairs and Trade (n.d). *Children's passports*. Retrieved from https://www.passports.gov.au/passports-explained/childrens-passports

Dowsett, A. J., & Burton, A. M. (2014). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology, 106*(3).

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).

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *Perception, 5*(7), 589-601.

Ferguson, E. L. (2015). *Facial identification of children: A test of automated facial recognition and manual facial comparison techniques on juvenile face images (Doctoral dissertation)*. University of Dundee, United Kingdom.

FISWG. (2010). Guidelines and Recommendations for Facial Comparison Training to Competency. Retrieved from http://www.fiswg.org/FISWG_Training_Guidelines_Recommendations_v1.1_2010_11_18.pdf.

FISWG. (2012). Guidelines and Recommendations for Facial Comparison Methods. Retrieved from http://www.fiswg.org/FISWG_GuidelinesforFacialComparisonMethods_v1.0_2012_02_02.pdf

Graves, I., Butavicius, M., MacLeod, V., Heyer, R., Parsons, K., Kuester, N., & Johnson, R. (2011). The role of the human operator in image-based airport security technologies. In E. A. L. Jain, & C. Abeynakake (Eds.), *Innovations in Defence Support Systems* (pp. 147-181). Heidelberg, Germany: Springer. doi: 10.1007/978-3-642-17764-4_5

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261-2271. doi:10.1016/S0042-6989(01)00097-9

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

Heyer. (2013). *Understanding one-to-many unfamiliar face matching in the operational context: The impact of candidate list size, expertise, and decision aids on the performance of facial recognition system users (Doctoral dissertation).* University of Adelaide, Australia.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*(1-2), 1-19.

Kemp, R. I., Caon, A., Howard, M., & Brooks, K. R. (2016). Improving unfamiliar face matching by masking the external facial features. *Applied Cognitive Psychology, 30*(4).

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

Kozak, F. K., Ospina, J. C., & Cardenas, M. F. (2015). Characteristics of normal and abnormal postnatal craniofacial growth and development. In P. W. F. M. M. Lesperance (Eds.), *Cummings Pediatric Otolaryngology* (pp. 55-80). Philadelphia, PA: Elsevier.

Kramer, R. S. S., & Ritchie, K. L. (2016). Disguising superman: How glasses affect unfamiliar face matching. *Applied Cognitive Psychology, 30*(6).

Kramer, R. S. S., Mulgrew, J., & Reynolds, M. G. (2018). Unfamiliar face matching with photographs of infants and children. *PeerJ, 6,* e5010.

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition, 172*, 46-58.

Lanitis, A. (2008). *Evaluating the performance of face-aging algorithms*. Paper presented at the Proceedings form the 8th IEEE International Conference on Automatic Face & Gesture Recognition.

Mandal, B., Lim, R. Y., Dai, P., Sayed, M. R., Li, L., & Lim, J. H. (2016). Trends in machine and human face recognition. In M. C. M. Kawulok, & B. Smolka (Eds.), *Advances in Face Detection and Facial Images Analysis*. Heidelberg, Germany: Springer. doi: 10.1007/978-3-319-25958-1

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception, 44*, 5-22. doi: 10.1068/p7825

Megreya, A. M., & Bindemann, M. (2018). Feature instructions improve face-matching accuracy. *PLoS ONE, 13*(3).

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

Menon, D., White, D., & Kemp, R. I. (2015). Identity-level representations affect unfamiliar face matching performance in sequential but non simultaneous tasks. *The Quarterly Journal of Experimental Psychology, 68*(9), 1777-1793.

Michalski, D. J. (2017). *The impact of age-related variables on facial comparisons with images of children: algorithm and practitioner performance (Doctoral dissertation).* The University of Adelaide, Australia.

Park, J., Newman, L. I., & Polk, T. A. (2009). Face processing: The interplay of nature and nurture. *Neuroscientist*, *15*(5), 445-449.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J., Castillo, C. D., Chellappa, R., White, D., & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, super recognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1721355115

Rhodes, G. (1988). Looking at faces - 1st order and 2nd order features as determinants of facial appearance. *Perception, 17*, 43-63.

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

Royer, J., Blais, C., Barnabe-Lortie, V., Carre, M., Leclerc, J., & Fiset, D. (2016). Efficient visual information for unfamiliar face matching despite viewpoint variations: It's not in the eyes! *Vision Research, 123,* 33-40.

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.

Sewel, A., & Hulusi, H. (2016). Preventing radicalisation to extreme positions in children and young people. What does the literature tell us and should educational psychology respond? *Educational Psychology in Practice, 32*(4), 343-354. doi:10.1080/02667363.2016.1189885

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46, 561–584

Spaun, N. A. (2007). *Forensic biometrics from images and video at the Federal Bureau of Investigation.* Paper presented at the IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS'07, Washington DC.

Towler, A., White, D., & Kemp, R. I (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*, 47-58.

United Nations Office on Drugs and Crime (UNODC) (2016). *Global report on trafficking in persons.* Retrieved from http://www.unodc.org/documents/data-and-analysis/glotip/2016_Global_Report_on_Trafficking_in_Persons.pdf

U.S. Department of state (UDOS (2007). *Trafficking in persons report*. Retrieved from https://www.state.gov/documents/organization/82902.pdf.

Vokey, J. R. (2016). Single-step simple ROC curve fitting via PCA. *Canadian Journal of Experimental Psychology*, 1-16.

Wechsler, H., & Li, F. (2014). Biometrics and robust face recognition. In S. H. V. Balasubramanian, & V. Vovk (Eds.), *Conformal prediction for reliable machine learning: Theory, adaptations and applications* (pp. 189-215). Waltham, MA: Elsevier.

Wei, X., & Li, C.T. (2017). Face recognition technologies for evidential evaluation of video traces. In M. T. C. Champod (Eds.), *Handbook of biometrics for forensic science* (pp. 177-193). Cham, Switzerland: Springer International Publishing.

White, D., Burton, A. M., Kemp, R. I., & Jenkins, R. (2013). Crowd effects in unfamiliar face matching. *Applied Cognitive Psychology*, *27*(6), 769-777.

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of automatic face recognition software. *PLoS ONE*, *10*(10).

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

White, D., Kemp, R., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS ONE, 9*(8), 1-6.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences, 282*, 1814-1822.

Wilkinson, C. (2012). Juvenile facial reconstruction. In C. W. C. Rynn (Eds.), *Craniofacial identification* (pp. 254-260). New York, NY: Cambridge University Press.

Wirth, B. E., & Carbon, C. C. (2017). An easy game for frauds? Effects of professional experience and time pressure on passport-matching performance. *Journal of Experimental Psychology: Applied, 23*(2), 138-157. doi: 10.1037/xap000011

Zeng, J., Ling, H., Latecki, L. J., Fitzhugh, S., & Guo, G. (2012). Analysis of facial images across age progression by humans. *ISRN Machine Vision, Article ID 505974*.

Appendices

Appendix A: University of Adelaide Information Sheet

PARTICIPANT INFORMATION SHEET

PROJECT TITLE: PERFORMANCE OF UNTRAINED PEOPLE ON FACE MATCHING
HUMAN RESEARCH ETHICS COMMITTEE APPROVAL NUMBER: [redacted]
PRINCIPAL INVESTIGATOR: Dr Carolyn Semmler and Dr Dana Michalski
STUDENT RESEARCHER: Thomas Pearce
STUDENT'S DEGREE: Honours in Psychology

Dear Participant,

You are invited to participate in the research project described below.

What is the project about?
This project aims to test students on a face matching task, comparing performance differences between matching children and adult faces. The results will be compared to practitioners performance, and the outcome of this study may influence training programs if the difference is similar to practitioners, prior to training. This is important as facial matching is used in areas such as processing, access control, and investigative applications, and thus is important to national security.

Who is undertaking the project?
This project is being conducted by Thomas Pearce, Dr Carolyn Semmler and Dr Dana Michalski. This research will form the basis for the degree of B. Psych (Hons) at the University of Adelaide under the supervision of Dr Carolyn Semmler and Dr Dana Michalski.

Why am I being invited to participate?
You are being invited as you are a student and this project is interested specifically in the performance of students on a face matching task. This study will contribute to research in facial matching which could have potential implications in areas such as passport processing.

What am I being invited to do?
You are being invited to complete a face matching task in which you will be presented with two faces and asked to identify whether they are the same or different people. This will be done on computers in the Hughes building room 219.

How much time will my involvement in the project take?
The task will take approximately 1 hour of your time and will be completed with one session.

Are there any risks associated with participating in this project?
There are no foreseeable risks, although in the case where an incident occurs you would be referred to the University Counsellor.

What are the potential benefits of the research project?
The research may result in changes to how Defence and National Security Agencies apply methods of face matching in passport identification, access control, and investigative applications.

Can I withdraw from the project?
Participation in this project is completely voluntary. If you agree to participate, you can withdraw from the study at any time up until submission of the thesis.

What will happen to my information?
Participation is anonymous, and so data will be de-identified after completion of the study.
After completing the study, you will be able to provide an email address if you wish to be given your results.
All data will be securely stored on the standalone Biometrics cluster stored in Building 75 of the Edinburgh site of Defence Science and Technology (DST) Group**.** Only the researchers will have access to the data. Upon publication, the de-identified data may be released.
The data will be used within an Honours thesis, participants will not be identified in publications, only non-identifiable data will be published. There is a possibility that the non-identifiable data may be used in future research.

Your information will only be used as described in this participant information sheet and it will only be disclosed according to the consent provided, except as required by law.

Who do I contact if I have questions about the project?

- [redacted]
- [redacted]
- [redacted]

What if I have a complaint or any concerns?
The study has been approved by the Human Research Ethics Sub-Committee in the School of Psychology at the University of Adelaide (approval number 18/57). This research project will be conducted according to the NHMRC National Statement on Ethical Conduct in Human Research (2007). If you have questions or problems associated with the practical aspects of your participation in the project, or wish to raise a concern or complaint about the project, then you should consult the Principal Investigators, Dr Carolyn Semmler or Dr Dana Michalski. If you wish to speak with an independent person regarding concerns or a complaint, the University's policy on research involving human participants, or your rights as a participant, please contact the Human Research Ethics Committee's Secretariat on:
Phone: +61 8 8313 6028
Email: hrec@adelaide.edu.au
Post: Level 4, Rundle Mall Plaza, 50 Rundle Mall, ADELAIDE SA 5000
Any complaint or concern will be treated in confidence and fully investigated. You will be informed of the outcome.

Appendix B: Defence Science and Technology Information Sheet

DST GUIDELINES FOR VOLUNTEERS

Thank you for taking part in Defence Science and Technology (DST) Group Research. Your involvement is much appreciated. This pamphlet explains your rights as a volunteer.

DST ethics review process

- DST Group has developed an approval process for low-risk research to ensure that human research complies with the requirements of the NHMRC (2007) *National Statement on Ethical Conduct in Human Research* and the Department of Defence (2007) *Health Manual Volume 23 Human Research in Defence – Instructions for Researchers.*
- If you are told that the project has DST ethics approval, this means that the Chief of Division or the DST Low Risk Ethics Panel has reviewed the research proposal and has agreed that the research is low-risk and is ethical. Ethical clearance through the Department of Defence and Veteran Affairs Human Research Ethics Committee (DDVA HREC) is not required for low-risk research.
- DST approval does not imply any obligation on commanders to order or encourage their service personnel to participate or to release troops from their usual workplace to participate. Obviously, the use of any particular personnel must have clearance from their commanders but commanders should not use DST Group approval to pressure personnel into volunteering.

Voluntary participation

- As you are a volunteer for this research project, you are under **no obligation** to participate or continue to participate. You may withdraw from the project **at any time** without detriment to your military career or to your medical care.
- At no time must you feel pressured to participate or to continue if you do not wish to do so.
- If you do not wish to continue, it would be useful to the researcher to know why, but you are under no obligation to give reasons for not wanting to continue.

Informed consent

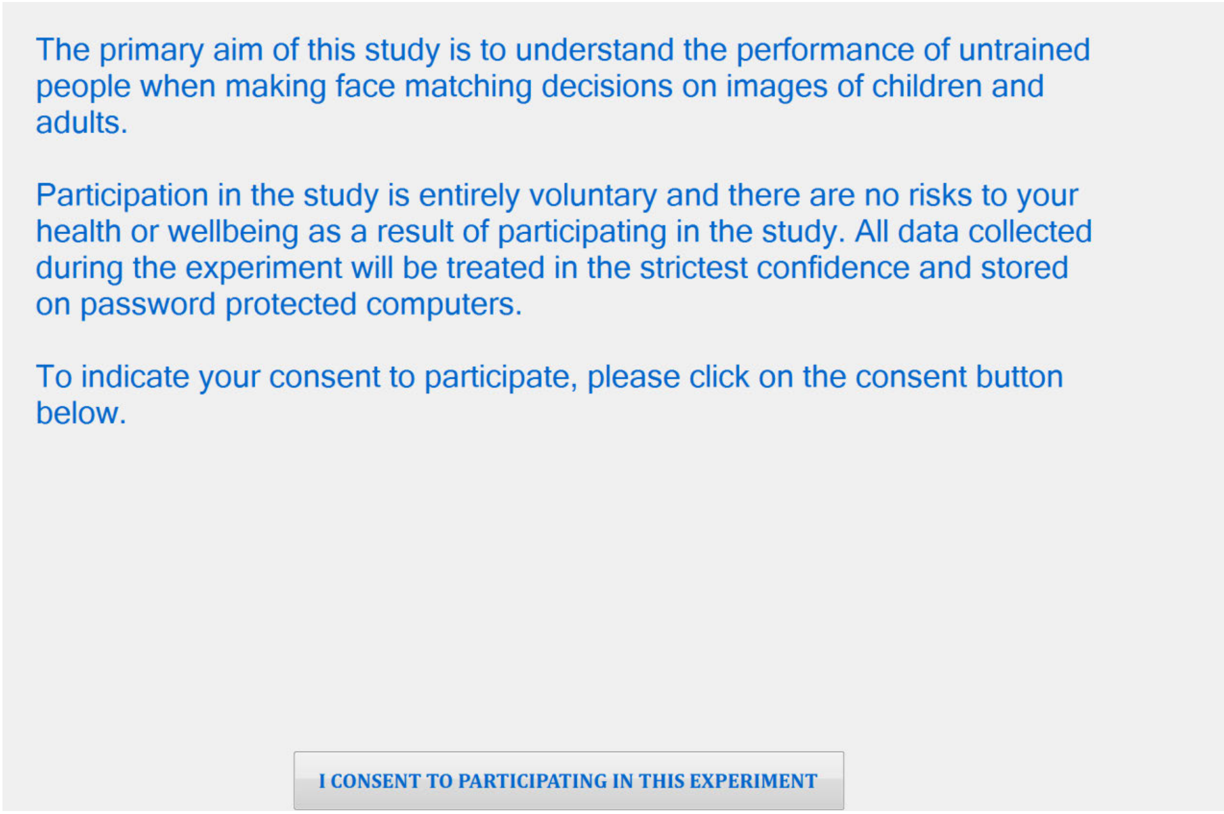
- Before commencing the project you will have been given an information sheet which explains the project, your role in it and any risks to which you may be exposed.
- You must be sure that you understand the information given to you and that you ask the researchers about anything of which you are not sure.
- You should ensure you are satisfied that you understand the information sheet and agree to participate, and keep a copy.
- Before you participate in the project you should also have been given a consent form to sign. You must be happy that the consent form is easy to understand and spells out what you are agreeing to. If you are happy you should sign the consent form and keep an un-signed copy of the consent form.

Complaints

- If at any time during your participation in the project you are worried about how the project is being run or how you are being treated, then you should speak to the researchers.
- Alternatively, you can contact the Chair of the DST Low Risk Ethics Panel. Contact details are:

[redacted]

Appendix C: Interface Consent



Appendix D: Consent Form

Human Research Ethics Committee (HREC)

CONSENT FORM

1. I have read the attached Information Sheet and agree to take part in the following research project:

Title:	**THE PERFORMANCE OF UNTRAINED PEOPLE ON FACE MATCHING**
Ethics Approval Number:	[redacted]

2. I have had the project, so far as it affects me, and the potential risks and burdens fully explained to my satisfaction by the research worker. I have had the opportunity to ask any questions I may have about the project and my participation. My consent is given freely.

3. I have been given the opportunity to have a member of my family or a friend present while the project was explained to me.

4. Although I understand the purpose of the research project, it has also been explained that my involvement may not be of any benefit to me.

5. I agree to participate in the activities outlined in the participant information sheet.

6. I understand that my participation is anonymous, and I am free to withdraw from the project at any time.

7. I have been informed that the information gained in the project may be published in a thesis.

8. I have been informed that in the published materials I will not be identified and my personal results will not be divulged.

9. I agree to my information being used for future research purposes.
   Yes ☐ No ☐

10. I agree to my information to be shared on an online digital repository. It will be in the form of non-identifiable data-sets

Yes ☐ No ☐

11. My information will only be used for the purpose of this research project and it will only be disclosed according to the consent provided, except where disclosure is required by law.
12. I am aware that I should keep a copy of this Consent Form, when completed, and the attached Information Sheet.

Participant to complete:

Name: ______________________ Signature: ________________________ Date: ____________________________

Researcher/Witness to complete:

I have described the nature of the research to ____

(print name of participant)

and in my opinion she/he understood the explanation.

Signature: ___________________ Position: _________________________ Date: ____________________________

Appendix E: Poster

Can you guess if these faces belong to the same person?

We are looking for people to come test their face matching skills on a simple, similar task to the above demonstration. This will be part of a study which is analysing the difference between students matching faces of children and adults.

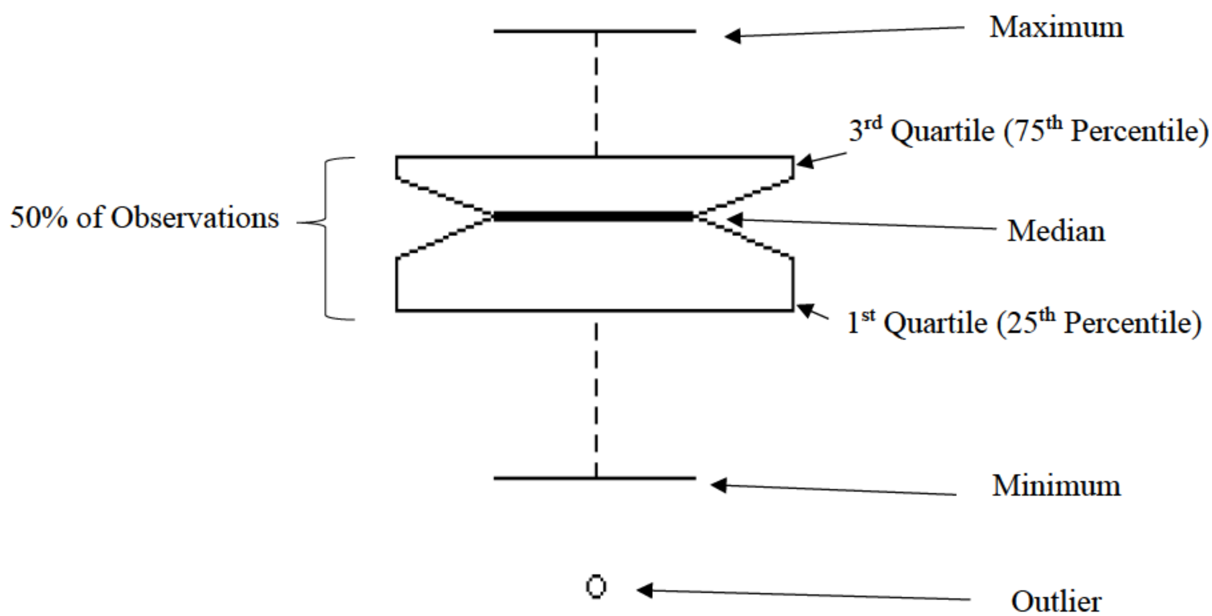
You will be able to:

- View your own results to know your own abilities
- Contribute to research that may be influential within government agencies and jobs such as passport examining
- Attain course credit if you are in first-year psychology

This study has been approved by DST Group Human Research Ethics Committee and the University of Adelaide Human Research Ethics Sub-committee [redacted]).

[redacted]

Appendix F: Notched boxplots explanation



This is a diagram of a notched boxplot. Several components make up a notched boxplot:

Median: This is the middle quartile and is represented as a line inside the box. Half of the observations are greater than the median and half are below the median. The box contains 50% of observations.

1st Quartile: The bottom line of the box is the 25th percentile. 25% of the data has values below this line.

3rd quartile: The top line of the box is the 75th percentile. 25% of the data has values above this line.

Minimum: At the end of the lower “whisker”, is the minimum data value.

Maximum: At the end of the upper “whisker”, is the maximum data value.

Outliers: The circles. Any observations differing substantially from the rest of the data.

Appendix G: Signal detection table

	Hit			Miss			CR			FA		
Conf	A	C	M	A	C	M	A	C	NM	A	C	NM
0	0	1	1	1	7	8	2	1	3	1	1	2
10	0	18	18	4	12	16	2	14	16	0	8	8
20	1	24	25	3	22	25	13	25	38	6	7	13
30	8	37	45	10	27	37	19	35	54	14	28	42
40	19	40	59	13	40	53	59	58	117	18	42	60
50	58	113	171	29	81	110	68	107	175	40	82	122
60	111	147	258	30	86	116	133	148	281	57	86	143
70	315	262	577	33	113	146	299	227	526	93	173	266
80	349	264	613	35	101	336	296	181	477	80	155	235
90	377	173	550	33	64	97	248	149	397	57	114	171
100	348	114	462	23	54	77	260	101	361	35	58	93

These are the descriptive statistics used in the signal detection analysis.

*Conf = the confidence percentage level for each condition, A = adult condition, C = child condition, M = mated condition, NM = non-mated condition. CR = correct rejection, FA = false alarm.

- Hit = choosing same on a mated pair

- Miss = choosing different on a mated pair
- Correct rejection = choosing different on a non-mated pair
- False alarm = choosing same on a non-mated pair

Appendix H: ROC outputs

```
Output from PCArocplotN

$parms
         dist         s           b       sd
1   Adults_FA 1.0000000  0.00000000 1.000000
2 Adults_Hits 0.9278329  0.55794946 1.077780
3     Kids_FA 0.8972987 -0.11289031 1.114456
4   Kids_Hits 0.8683479 -0.04756269 1.151612

$dprime
                    d_1         d_2         d_a         d_e         d_p
 d_YNp
Adults_FA    0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 0.00000000
Adults_Hits  0.60134689  0.55794946  0.57843075  0.57883590  0.53026184
 0.37495174
Kids_FA     -0.12581130 -0.11289031 -0.11882713 -0.11900109 -0.15229270
-0.10768720
Kids_Hits   -0.05477377 -0.04756269 -0.05078824 -0.05091417 -0.09807233
-0.06934761
                  A_z      A_zp         w
Adults_FA   0.5000000 0.5000000 0.5000000
Adults_Hits 0.6588398 0.6461518 0.6582119
Kids_FA     0.4664700 0.4571219 0.4807488
Kids_Hits   0.4856405 0.4723565 0.4990788

$crit
                    c        c_1        c_2        c_a        c_e
Adults_FA   0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
Adults_Hits 0.6487910 -0.6730780 -0.6245040 -0.6474283 -1.2939504
Kids_FA     0.3075137 -0.3241595 -0.2908679 -0.3061645 -0.6114338
Kids_Hits   0.3346239 -0.3582030 -0.3110448 -0.3321389 -0.6626347
```

```
$z
   Adults_FA Adults_Hits    Kids_FA   Kids_Hits
8 -1.3409134  -0.7797391 -1.4260769 -1.30728824
7 -0.7175939  -0.1163386 -0.7450634 -0.70447023
6 -0.1454966   0.4456920 -0.1669950 -0.09574675
5  0.4711521   1.1226717  0.4210206  0.47182815
4  0.9453641   1.5437318  0.7627318  0.85850820
3  1.4802090   1.9594289  1.2052243  1.27821533
2  1.9928991   2.2220067  1.5687300  1.49806765
```

```
$s
[1] 1.0000000 0.9278329 0.8972987 0.8683479

$b
[1]  0.00000000  0.55794946 -0.11289031 -0.04756269

$pc_max
[1] 0.5000000 0.6137917 0.4763114 0.48987

$z.m
  Adults_FA Adults_Hits     Kids_FA   Kids_Hits
  0.3836601   0.9139219   0.2313674   0.2855877



$cumdat
   Adults_FA Adults_Hits    Kids_FA  Kids_Hits
8 0.08997429   0.2177722 0.07692308 0.09555742
7 0.23650386   0.4536921 0.22811671 0.24056999
6 0.44215938   0.6720901 0.43368700 0.46186085
5 0.68123393   0.8692115 0.66312997 0.68147527
4 0.82776350   0.9386733 0.77718833 0.80469405
3 0.93059126   0.9749687 0.88594164 0.89941324
2 0.97686375   0.9868586 0.94164456 0.93294216

$pvar
[1] 0.0000000 0.9975633 0.9984755 0.9972910

$zC
  ranks Adults_Hits    Kids_FA   Kids_Hits
8    c8   1.3875550  1.3955854  1.38811538
7    c7   0.3992411  0.7544162  1.01506524
6    c6  -0.1612403  0.4396450  0.07207939
5    c5  -0.5347594 -0.5264007 -0.52574074
4    c4  -1.3471766 -0.9401113 -0.65788036
3    c3  -1.8374711 -1.1181692 -1.51148321
2    c2  -1.9006479 -1.9399078 -1.90136291

$pvals
   Adults_FA Adults_Hits    Kids_FA  Kids_Hits
8 0.08997429   0.2177722 0.07692308 0.09555742
7 0.23650386   0.4536921 0.22811671 0.24056999
6 0.44215938   0.6720901 0.43368700 0.46186085
5 0.68123393   0.8692115 0.66312997 0.68147527
4 0.82776350   0.9386733 0.77718833 0.80469405
3 0.93059126   0.9749687 0.88594164 0.89941324
2 0.97686375   0.9868586 0.94164456 0.93294216
```