Human age estimation performance based on facial images: Potential implications for refugee processing

Gemma Snyder

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

School of Psychology

The University of Adelaide

October 2018

Word Count: 9000

Table of Contents

List of Figures v

List of Tables vi

Abstract vii

Declaration viii

Acknowledgements ix

Chapter 1 – Introduction 1

1.1 Rationale 1

1.2 Facial Ageing and Visual Perception 4

1.2.1 Local feature cues. 5

1.2.2 Configural cues. 5

1.2.3 Surface cues. 6

1.2.4 Perception of cues. 6

1.2.5 Intrinsic and extrinsic factors of ageing. 6

1.2.6 Own-race bias. 7

1.3 Human Age Estimation with Facial Images 8

1.4 Demographic Factors Affecting Age Estimates 9

1.5 Metacognition as a Personal Insight into Estimation Ability 10

1.6 Current Study 11

Chapter 2 – Method 13

2.1 Ethics Statement 13

2.2 Participants 13

2.3 Design 13

2.4 Materials 14

2.4.1 Operational images. 14
2.4.1.1 Image source. 14
2.4.1.2 Image selection. 14
2.4.1.3 Image set. 15
2.4.2 Experimental application and equipment. 15
2.5 Procedure 16
2.6 Dependent Measures 17
2.6.1 Measuring estimation performance...17
2.6.1.1 Accuracy. 17
2.6.1.2 Directional estimation error. 17
2.6.2 Metacognition measures. 18
2.6.2.1 Bias score. 18
2.6.2.2 Precision range. 18
Chapter 3 – Results 19
3.1 Data Screening and Assumptions 19
3.2 Accuracy of Estimates 20
3.3 The Effect of Image Gender on Estimation Accuracy 21
3.4 Directional Estimation Error 22
3.5 Bias Score 24
3.6 Precision of Range 25
3.7 Features Considered in Age Estimations 26
Chapter 4 – Discussion 28
4.1 Accuracy of Age Estimates 28
4.2 Image Gender Affecting Estimation Accuracy 29
4.3 Directional Estimation Error 30

4.4 Overconfidence 31

4.5 Implications..32

4.6 Limitations 33

4.7 Strengths 34

4.8 Directions for Future Research 35

4.9 Conclusion 37

References...38

Appendices..48

Appendix A – Recruitment Poster 48

Appendix B – DST Intranet Site Recruitment Post 49

Appendix C – Participation Consent..50

Appendix D – Demographic Questions 51

Appendix E – Instructions 52

Appendix F – Practice Examples..53

Appendix G – Post-experiment Questions..54

Appendix H – Participant Information Sheet...55

Appendix I – DST Guidelines for Volunteers 57

Appendix J – Verbal Briefing Delivered to Participants Prior to the Experiment...............58

Appendix K – Testing Normality using Shapiro-Wilk Tests..60

List of Figures

Figure 1. Boxplot of the directional estimation error by the true age of the stimuli. 22

Figure 2. Scatter plot of the relationship between true age and estimated age. 23

List of Tables

Table 1. Descriptive statistics of the mean absolute error (MAE)..20

Table 2. Descriptive statistics of the mean absolute error (MAE) across gender....................21

Table 3. Descriptive statistics for directional estimation error..24

Table 4. Descriptive statistics for bias score and precision range 25

Table 5. Self-reported factors considered during age estimations...27

Abstract

Government agencies responsible for refugee processing are often challenged to develop an accurate identification profile for asylum seekers when reliable documentation is unavailable. As processing and support systems are designed differently for children and adults, it is critical to determine an accurate age estimate. However, current methodologies to estimate an individual's age lack accuracy, verification or violate ethical standards, presenting the need to explore an alternative age estimation procedure. The present study aimed to explore human performance of age estimation, using facial images, for refugee processing purposes. As a within-subjects design, participants (N = 46) undertook a perceptual task to estimate the age of both children (11–17 years) and adults (18–24 years) from a facial image. Estimates were more accurate (i.e., closer to the true age) for children than for adults, although there was a consistent tendency to overestimate the true age. If this methodology was utilised for refugee processing, trends of over-, as opposed to under-estimation, increase the likelihood of a child being incorrectly labelled as an adult, than an adult labelled as a child. Future research could aim to further develop a task performance baseline by defining group-specific biases of estimators and estimations. The results of this study provide government agencies with an initial understanding of task performance and the potential biases in human perception.

Declaration

This thesis contains no material which has been accepted for the award of any other degree of 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.

Gemma Snyder

October, 2018

Acknowledgements

I would like to express my gratitude toward my supervisors; it has been a pleasure to work with each and every one of you. To my internal supervisor from the University of Adelaide, Dr Carolyn Semmler, I am grateful for all of the knowledge you have imparted and in particular, your statistical expertise. Thank you for your enthusiasm and support throughout this year, I greatly appreciate it. To my external supervisor from DST, Dr Dana Michalski, your passion for this field has been inspiring and motivated me throughout the year to consider the importance of this research. Thank you for the assistance provided, whether it be your feedback or your positive encouragement–it all added up. I also wish to extend my appreciation to Dr Veneta MacLeod from DST for your insight within this field and invaluable advice throughout the year. Thank you.

In addition, I would like to acknowledge the staff of DST for the resources provided to me and for accommodating the data collection within your organisation.

Finally, a special thank you to my friends and family who provided feedback throughout the year, edited my thesis or supported me with your ongoing encouragement. I truly value all of your contributions.

Chapter 1

Introduction

Within various settings including child exploitation, sale of restricted goods, eyewitness testimonies and refugee processing, the age of an individual is often unknown or documentation is unreliable. Age estimation refers to the determination of an individual's age (Grd, 2013). A correct estimate is often critical for legal processes. In refugee settings, accurate age estimation is necessary to ensure appropriate processing and support systems are provided. As an under-researched area, this thesis explored human age estimation performance for refugee processing purposes.

1.1 Rationale

The global plight of refugees has demanded national authorities evaluate their age estimation procedures to ensure appropriate processing. A refugee is someone whose application for a country's protection was approved, after fleeing their origin country based on fear of persecution, violence or war (Amnesty International, 2017). An asylum seeker is someone applying for refugee status. The United Nations High Commissioner for Refugees (UNHCR, 2017) reported 22.5 million refugees and 2.8 million asylum seekers in 2016. A total of 51% were children (defined as under the age of 18 years) with 75,000 being unaccompanied minors. The highest origin countries were within the Middle East. In these situations, it is often difficult to obtain identity documentation as people are fleeing the country with limited time or fail to meet the requirements (Refugee Council of Australia, 2010). The United Nations Children's Fund (UNICEF, 2018) reported that 13% of children under five in the rural Middle East were not registered at birth. Host countries typically recognise the vulnerability of children and enact different processing and support systems. In Europe, a regime safeguards refugee children from wrongful detainment (Feltz, 2015). A child is entitled to benefits including access to housing, health care and education (Aynley–

Green, 2011). Similarly, Australian policies provide unaccompanied minors protection from detainment (Button & Evans, 2017). A child in an adult immigration detention facility risks threats to safety and wellbeing, with an average detainment of 436 days (Australian Border Force, 2018). Given the outcomes, correct identity determination is a concern for government agencies (Feltz, 2015; Thevissen, Kvaal, Dierickx, & Willems, 2012).

The misclassification of refugees based on inaccurate assessments, violates ethical responsibility, by failing to uphold basic human rights (UNICEF, 2017). Yet, misclassifications are not uncommon. A recent report disputed the acceptance of 13.6% of child refugee applications (Drury, 2018). The benefits received financially impacted the host country. Adults were assigned to child-based environments including schooling, foster care and childcare, prompting concerns of the safety risks posed to children. Of equal importance is an accurate estimation of a child claiming to be an adult. This situation presents for victims of people smuggling, child trafficking, sexual exploitation and forced marriage (Thevissen et al., 2012). An accurate age determination is critical for the child's protection and for prosecution of offenders. Failure to correctly recognise a child prevents accessibility to purpose designed support. A testimony from 16 year-old Zacharia, wrongfully deemed 18, demonstrated the effect of this decision (Feltz, 2015). Proof of age documents were considered counterfeit and Zacharia was consequently refused benefits and became homeless.

UNICEF (2017) attempts to ensure fair and appropriate processing; however, there is no systematic, universal regulation for age estimation. A report by the European Asylum Support Office (EASO, 2018) found medical methodologies as common forms of age assessment in refugee processing. Twenty-three member states of the European Union (EU) incorporated carpal X-ray assessments where bone development is compared to an image sample of upper socio-economic, Caucasian Americans (Feltz, 2015). However, skeletal maturation is culturally specific, resulting in low cross-ethnic reliability. Furthermore, delays in bone mineral density are associated with malnutrition, reportedly affecting 44.9% of refugee children (Kueper, Beyth, Liebergall, Kaplan, & Schroeder, 2015). Combining these factors forms an error approximating two years above and below the true age (Chaussain and Chapuis 2007 as cited in Feltz, 2015). Additionally, the method is at odds with medical ethics whereby the risk of potentially harmful ionising radiation has no foreseeable medical benefits (Feltz, 2015). The Australian Government considered the lacking validity substantial enough to discontinue X-ray assessments from age determination procedures (Hurley & Beaumont, 2016).

The evidence against such medical examinations highlights the need for an alternative method. Non-medical methods have been incorporated into the age assessment process (EASO, 2018). Age assessment interviews are employed by 17 EU states and within Australia, whereby interviewers construct the applicant's chronological age using life accounts (Hurley & Beaumont, 2016). Methodological variation prevents qualification of the assessment's validity, raising ethical questions of the use of an unproven practice. Additionally, time and resource constraints restrict the feasibility of employing multiple interviewers for inter-rater reliability. Subjective interpretations and biases are therefore likely to strongly influence interview outcomes.

Estimates based on facial appearance appear to be a viable option with minimal time, ethics violations or expense required. Research suggests that even with poor quality images, recognition is highly reliant on facial cues (Burton, Wilson, Cowan, & Bruce, 1999). The body, although useful for identification, does not contribute to an accurate identification beyond visual perception of the face in isolation (O'Toole et al., 2010). For static displays (i.e., images) faces were observed as the primary focus, demonstrating the natural attentiveness toward facial cues (O'Toole et al., 2011). Therefore, facial images appear to be an appropriate design for exploring the accuracy of age estimation based on physical appearance.

In summary, there is no scientific method available to accurately determine the chronological age of a person, highlighting the need for a valid research based approach to age estimation. The employment of this strategy in a practical setting would allow for multiple assessors, decreasing the influence of individual biases and minimising the applicant's required presence. The aim of this thesis was to explore human performance of age estimation, using facial images, for refugee processing purposes.

1.2 Facial Ageing and Visual Perception

Facial processing is a quantitatively unique form of visual perception that relies on a holistic representation of the face (i.e., cues are perceived in conjunction) (Hole & George, 2011). The fusiform face area in the inferior temporal cortex is specialised for facial recognition (Kanwisher & Yovel, 2006). From just 30 minutes of age, infants visually track faces over other objects, demonstrating the innate informational value of faces (Johnson, Dziurawiec, Ellis, & Morton, 1991).

Bruce and Young (1986) modelled facial processing, whereby different cortical processes govern factors of identification, expression, gender and age. In essence, the utilisation of facial cues varies dependent on the task. Age and sex represent the initial stages of the identity processing sequence (Ellis, 1986 as cited in Young & Ellis, 2013). Age is one of three features accounting for the variation between faces in categorisation tasks, highlighting the importance of age in facial processing (Laughery, Rhodes, & Batten, 1981). This value is attributed to the complex, yet available cues to an ageing face, divided into three domains of local, configural and surface cues (Hole & George, 2011), as discussed in the following sections.

1.2.1 Local feature cues. Local feature cues refer to the nose, mouth, eyes and ears. Throughout a person's lifetime, the ears and nose constantly grow while gravitational pull causes cartilage breakdown, resulting in a drooped appearance (Han, Otto, & Jain, 2013). Research with automatic age estimation technology determined these features as the most crucial aspect for correct age estimation (Han et al., 2013). With age, all facial features increase in size and coarseness (Kozak, Ospina, & Cardenas, 2015). A growth spurt occurs at approximately 11 years for females, with minimal change after this age (Ricanek, Mahalingam, Albert, & Vorder Bruegge, 2013). For males, the spurt extends from 11 to 16 years (Ferrario, Sforza, Poggio, & Schmitz, 1998).

George and Hole (1998) demonstrated the importance of facial features for human age estimation. The manipulation of older persons features onto younger faces increased perceived age by approximately 40%. Similar results were obtained by masking children's facial features (Jones & Smith, 1984). Interestingly, despite consistent ageing of the ears and nose, the study found the eyes were the critical factor for correct age estimations in these manipulations. Despite limited research in this field, the evidence supports that humans are attuned to changing local feature cues, affecting their judgment abilities.

1.2.2 Configural cues. Configural cues are represented on a global scale. Cardiodial strain is the geometric transformation of the face and proposed to be the most distinctive cue to childhood ageing (Rhodes, 2009). From birth to 20 years of age, the forehead slopes backwards and become less prominent on the cranium's surface (Geng, Fu, & Smith-Miles, 2010). Interstitial spaces are filled by the cheeks and facial features while the chin becomes more protrusive. The relative growth of features to face size is a process generally complete by 14 years of age (Ricanek et al., 2013). Perception is remarkably sensitive to these changes, with the mere presence of a face outline affecting age estimates (Pittenger & Shaw, 1975).

After 20 years of age, there are few configural cues to ageing, suggesting human perception would rely on local and surface cues.

1.2.3 Surface cues. Surface cues are a global effect of ageing on the skin. The loss of the skin's elasticity causes fine-scale wrinkling. Results suggest that age estimation can be influenced on the sole basis of wrinkles (Montpare´ & McArthur, 1986; Tiddeman, Burt, & Perrett, 2001). Removing the appearance of wrinkles reduced apparent age by 10 years (Fink & Matts, 2007). To a lesser extent, the removal of pigmentation affected age judgments, reducing estimations by five years (Fink & Matts, 2007). These skin changes vary between genders. Female faces reportedly maintain 'babyish' features (i.e., reduced wrinkles) for longer (Enlow, Pfister, Richardson, & Kuroda, 1982). Males conversely undergo more visible changes during puberty, with facial hair beginning at approximately 14 years of age (Marshall & Tanner, 1970). The onset of acne occurs around 13–14 years for males and females, another maker of puberty (Lee, 1980). Puberty represents a distinctive change in age, however, the variability is high with some individuals never displaying acne or only patchy facial hair.

1.2.4 Perception of cues. Ageing processes are critical to understanding how people discriminate age. Whilst these cues are defined separately, they are not perceived in isolation and it is this overlapping representation that develops ones age calibration model (Hole & George, 2011). The combination of cues available supports the human capacity for age estimation.

1.2.5 Intrinsic and extrinsic factors of ageing. Perceiving cues can be biased by preconceived timelines of the ageing process. Visible ageing is the result of interacting intrinsic (i.e., genes) and extrinsic (i.e., environment) factors. The uncontrollability of extrinsic factors can alter ageing processes outside of expectations. Exposure to weather elements causes acceleration and accentuation of the natural ageing process (Albert, Ricanek,

& Patterson, 2007). Sleep deprivation and stress has a similar affect on wrinkles while lifestyle factors (i.e., malnutrition) delay physical development, characterised as a younger appearance (Taister, Holliday, & Borrmman, 2000).

Extrinsic factors also encompass controlled interventions to delay visible ageing. Research suggests females are more attentive to their physical appearance by using skin care products, reducing the visible appearance of ageing (Voelkle, Ebner, Lindenberger, & Riediger, 2012). Makeup is worn for similar reasons, predominately by females (Dantcheva, Chen, & Ross, 2012). When perceiving these differences, attractiveness had a stronger negative correlation for age estimates of females compared with males (Henss, 1991), suggesting that such interventions successfully reduced visible ageing. However, also influencing these perceptions are social norms regarding makeup (e.g., when females begin to wear makeup) (Guo, Wen, & Yan, 2014). Considering the research, there is sufficient evidence to suggest that the influence of these extrinsic factors, would affect perceptions of age.

1.2.6 Own-race bias. Ageing is an inevitable process and the presentation of these changes has been partially attributed to one’s ethnic origin (Vashi, Maymone, & Kundu, 2016). These differences can develop an own-race bias, where faces from one’s own race are perceived easier than other races. A performance deficit was observed when Caucasian individuals estimated the age of African versus Caucasian faces (Dehon & Brédart, 2001). The reverse effect was not found, as there was no performance difference by the African participants. Dehon and Brédart (2001) proposed that as the African subjects resided in predominantly Caucasian areas, familiarity with relevant cues eliminated effects of the own-race bias. Although the limited research causes an inconclusive effect of ethnicity on perception, ethnic differences in ageing supports the likelihood that difficulty increases when estimating dissimilar ethnicities. The present study did not investigate observer ethnicity, however, the faces presented varied between the perceiver and stimuli's ethnicity and so the potential for performance differences was addressed.

1.3 Human Age Estimation with Facial Images

Human age estimation from facial images is currently an exploratory area of research. As such, the findings regarding human accuracy are conflicted. With few studies available, this could represent sampling or methodological differences. Categorising facial images of Caucasians individuals into an age range of 18–25 years resulted in 91.1% accuracy, which significantly decreased to 65.6% accuracy for an age range of 55–75 years (Anastasi & Rhodes, 2006). Whilst categorisation provides a broad overview, a more descriptive measure of accuracy is the mean absolute error (MAE) to describe the magnitude of estimation error. It is expected that adults' ages will similarly be less accurately estimated, with fewer cues available and increased exposure to extrinsic factors. Humans are presented with the challenge of managing more variability associated with adult ageing, based on differing severity and onset times.

Research has supported this claim through finding an age-based effect on perception. An absolute error of 2.83 years for faces aged 15–24, was profoundly lower than 5.25 years for faces aged 56–65 (Sörqvist & Eriksson, 2007). Humans even outperformed automatic estimation technology for faces under 15 years of age (Han et al., 2013). Whilst error rates are inconsistent across studies, the positive correlation between chronological age and the magnitude of estimation error remains consistent (Moyse, 2014). In essence, children were more accurately estimated than adults.

Comparing age estimation differences from images of children versus adults is often explored in the context of alcohol sales. A pivotal finding was the age overestimations of minors, which would result in an illegal sale (Vestlund, Langeborg, Sörqvist, & Eriksson, 2009; Rowe & Willner, 2001). Described as directional estimation error, the degree of overestimation was reportedly higher for the younger age group, 15–19 years, than the older age group, 20–24 years, with ‘experts’ (i.e., alcohol salespersons) in this study, outperforming a control sample, suggesting that exposure and training may increase accuracy (Vestlund et al., 2009). Exposure may allow cognitive processes to map changing facial dimensions and features, building a perceptual representation of faces. Minimal consideration towards the distinct contributions of bias to age estimation performance limited generalisations of these results. Humans are capable of controlling the decision they arrive at, with motivational reasoning biasing the accessing, construction and evaluation of beliefs (Kunda, 1990). Biased decision-making occurs when there is motivation to achieve a particular conclusion. Therefore, the rationalisation of the task can constrain decision accuracy. A task format of assessing age for alcohol purchase, versus refugee processing, entails motivational outcomes that activate different belief systems. In alcohol sales, an overestimation increases the likelihood of a successful sale, potentially explaining the observed overestimations of younger faces. An alternative hypothesis proposes overestimation in social quantification tasks reflects a regression to the population mean (Vestlund et al, 2009). People younger than the median age of 37.2 years in Australia (Australian Bureau of Statistics, 2016) would be overestimated under this hypothesis.

1.4 Demographic Factors Affecting Age Estimates

The nature of perceptual experience varies based on cognitive, biological and social factors. These factors are often reflected by demographics, affecting estimators and estimations. A key demographic explored is the age of participants; with those aged 18–45 years significantly outperforming those aged 55–75 years in facial age estimation (Anastasi & Rhodes, 2006). These observations were attributed to an own-age bias, whereby perceivers are more accurate at estimating their own age group than perceivers not within the same age group (Moyse & Brédart, 2012; Rowe & Willner, 2001; Rhodes, 2009). The expertise hypothesis proposes this result occurs due to increased exposure to one's own age group, a concept supported across various task formats (Moyse & Brédart, 2012; Sörqvist, Langeborg, & Eriksson, 2011).

Another key variable is gender, with the few studies conducted finding male faces were more accurately estimated than females (Voelkle et al., 2012). For faces aged 20–45 years, a MAE of 6.85 years was reported for males versus 7.09 years for females, a significant difference (Dehon & Brédart, 2001). In the limited research with children, male faces aged 11–15 years were more accurately estimated than females, during the period of puberty change (Ferguson & Wilkinson, 2016). The true explanation is unknown; however, gender differences in visible ageing (see Sections 1.2.1, 1.2.3 and 1.2.5) supports the likelihood of differences in perceiving and estimating age.

1.5 Metacognition as a Personal Insight into Estimation Ability

Accurate decision-making requires factual knowledge, as well as metaknowledge. This is the ability to monitor and control decision processes – knowing when to answer and the limits of one's abilities. Metacognitive insight varies between individuals (Russo & Schoemaker, 1992) and measurement of this construct is often explored by quantifying the relationship between confidence and accuracy.

Overconfidence is where the subjective confidence is higher than the objective accuracy of that judgment. This distortion is common and deeply rooted in beliefs, likely from limited judgment feedback (Russo & Schoemaker, 1992). Although not researched in age estimation, general concepts of the anchoring bias suggests individuals may anchor their estimates on familiar people with a known age identity. This develops overconfidence, particularly if the range of age-known individuals is limited. The resulting incorrect estimate is defined as misestimation, a form of overconfidence (Russo & Schoemaker, 2016). Misprecision is the tendency to estimate quantities within a range too narrow to contain the true value. To determine a range estimate requires an accuracy-informativeness trade-off (Yaniv & Foster, 1995). Imprecise judgments are less informative due to the large range provided, but more likely to be accurate (i.e., contain the true value). The logical proposition is that these judgments demonstrate a degree of decision uncertainty. However, people are often willing to accept a degree of error to fulfil cognitive and social demands to be informative and precise (Yaniv & Foster, 1995).

Although some level of precision is necessary, prioritising informativeness over accuracy is not practical within refugee processing settings where it could result in a misclassification. Yet, there was no evidence obtained of calibration accuracy for age estimation, presenting a need for further research.

1.6 Current Study

The present study aimed to explore human performance of age estimation, using facial images, for refugee processing purposes. Although incorrect estimations may seem inconsequential in everyday settings, the prominence of age in dictating social interactions and description of unfamiliar people, suggests otherwise (Voelkle et al., 2012). Nonetheless, understanding human performance is still largely under-researched and as such, the present study is exploratory. Previous research was not identified for age estimation using refugee-processing contexts. The research exploring child/adult estimate differences (Vestlund et al., 2009; Rowe & Willner, 2001) was contextually constrained and thus prevented generalisations. Furthermore, limited image sets potentially skewed the results and compromised internal validity. Vestlund et al. (2009) required participants to view only 40 images, while Rowe and Willner (2001) only assessed four ages.

To address these limitations, the present study provided participants with the research rationale of refugee processing. Target stimuli were selected as people of Middle Eastern appearance, representing operational purposes and enhancing content validity. As the own-race bias suggests the stimuli's ethnicity affects perception, this design increased the likelihood of accurately depicting task performance. The image set represented a comparatively large sample of 140 individuals. Within this, the conditions (children and adults), individual ages and gender were equally represented.

A within-subjects design enabled direct performance comparisons between participants by reducing the influence of individual differences (Pelham & Blanton, 2012). Therefore, fewer participants were required to detect an effect. One disadvantage is the potential for carryover effects, when one response is directly affected by a second stimulus. These order effects were counteracted through randomisation of the stimuli between participants. Another disadvantage is possible fatigue and boredom associated with participation in both conditions. To minimise this effect, participation in the study was voluntary to encourage intrinsically motivated participants. Intrinsic motivation implies individuals engaged in the task for personal satisfaction and were therefore, motivated by success of the task (Ryan & Deci, 2000).

Based on the limited research, general hypotheses guided this exploratory study. The research was focused on comparing performance differences when estimating children versus adults, although some variables lacked evidence to initially hypothesise superiority between the groups. The hypotheses were:

1) Children's ages will be more accurately estimated than adults.

2) Age estimation will be more accurate for male faces than female faces.

3) The ages of children and adults will be overestimated

4) Participants will be overconfident in their age estimations of children and adults.

Chapter 2

Method

2.1 Ethics Statement

The University of Adelaide Human Research Ethics Committee approved this study (approval number: 18/44). The Defence Science and Technology Group (DST) Ethics Review Panel also approved this study (NSID 02-18).

2.2 Participants

A total of 46 participants completed this study (16 female, 30 male), aged 21–80 years (*M* = 39.76, *SD* = 13.23). The majority identified as Caucasian (91%), with a minority identifying as Asian (9%).

Participants were employees of DST Edinburgh, South Australia and were recruited on a voluntary basis. Incentives were not provided aside from refreshments and the opportunity to receive individual results, suggesting the participants were intrinsically motivated to complete the study. Additionally, the privacy restrictions of the operational images required secure and classified computers, available at DST. Participants were recruited through poster advertisements at DST (see Appendix A) and on DST's intranet site (see Appendix B).

Inclusion criteria for participation was (a) corrected-to-normal or normal vision and (b) aged 18 years and over.

2.3 Design

A within-subjects design was used for this computer-based study. The two conditions for the experiment were: (a) images of children and (b) images of adults. Participants completed both conditions to reduce the influence of individual differences in analyses (Pelham & Blanton, 2012). To counteract order effects, the images were randomly presented for each participant and counterbalanced between participants. Each participant viewed the same image set.

2.4 Materials

The materials in this study were images, displayed on a computer.

2.4.1 Operational images. The image set was developed to represent operational purposes.

2.4.1.1 Image source. The images were sourced from a government agency, provided to DST for research purposes. The database was selected as it contained important demographic information (i.e., age and gender) of the people in the images. Consistency was ensured as the images were passport style: good quality, colour, plain background, hair off the face, uniform lighting, frontal face, eyes open and mouth closed. The absence of background cues limited contextual biases that potentially influence estimates as a confounding variable (e.g., if a toy was visible participants may be biased to estimate a child). Therefore, it was ascertained that estimates were based on facial cues, useful for operational purposes where contextual information can be limited.

2.4.1.2 Image selection. Researchers undertook a selection process for suitable images defined by:

(a) Age range

The age range was selected to equally represent children and adults for equitable comparisons. Following recommendations designed for reference material (Solari, 2001), an even distribution of stimuli for each age was selected. For an equal representation while allowing for sufficient observations per age (n = 10), the age range was determined for children as 11–17 years and adults as 18–24 years.

(b) Gender

The visible rate of ageing varies by gender, potentially causing a confounding effect on the relationship between true and estimated age. Gender was equally represented (five female, five male) within each age to control for this and allow equitable comparisons.

(c) Ethnicity

People of Middle Eastern appearance were selected, to represent the highest origin countries of refugees (UNHCR, 2017). This enhanced ecological validity by reflecting operational settings. Unfortunately, ethnic information was not supplied by the database, requiring a subjective judgment. To minimise personal bias, three researchers independently selected images. Images used were agreed as an appropriate ethnic representation.

(d) Headscarf wear

Images of people wearing headscarves were excluded as this isolates cues to internal features of the face, affecting facial recognition (Toseeb, Keeble, & Bryant, 2012). Headscarves may have added additional variance to the cues used for estimates, causing a lack of consistency in outcomes.

(e) Neutral expression

Neutral facial expressions were selected as it is a variable that increases estimation accuracy when compared to non-neutral expressions (Voelkle et al., 2012).

2.4.1.3 Image set. From the images that met requirements, a total of 140 images were selected. Of this, 70 images were of children (11–17 years) and 70 images were of adults (18–24 years). For each year of age within these ranges, 10 images (five female, five male) were selected.

2.4.2 Experimental application and equipment. The experimental interface was custom designed in Visual Studio 2015 C+. The software stored consent, demographic information, experimental data, post-experiment questions and e-mail registration for personal results (see Appendix C to G). Images were randomised and counterbalanced by the application. The study was conducted in DST's computing laboratory. Ten computers were used, each with a 23 inch monitor and a 1920 x 1200 screen resolution.

2.5 Procedure

One hour was allocated for each session, however participants could use more time as required. Prior to commencing the experiment, participants received the Participant Information Sheet that included a brief study rationale (see Appendix H) and DST Guidelines for Volunteers (see Appendix I). After participants understood these items, a verbal briefing that entailed a summary of the experiment commenced (see Appendix J). Participants indicated their consent to participating by clicking the designated button on the experimental interface (see Appendix C) and then entered their allocated unique identification number. This allowed for re-identification if participants no longer agreed to the inclusion of their data post-experiment. The following screens requested demographic information of age, gender and ethnicity (see Appendix D). Participants described their vision to ensure compliance with the criterion of normal or corrected-to-normal vision.

The next screen presented instructions of the study (see Appendix E). Before the experiment, participants completed one practice trial with two parts (see Appendix F). Part One presented one facial image and required participants to determine if the person was a child, defined as under 18 years, or adult, defined as 18 years and over. Once they clicked 'CHILD' or 'ADULT', they were unable to change their answer. Part Two asked participants to estimate the person's age. The quantity provided had to fall within the boundary of the previous decision (e.g., if 'CHILD' was selected, they were unable to enter an age older than 17). Participants were asked to provide a range that they were 80% confident that the true age fell within (see Section 2.6.2.1). Although Part One could be determined by Part Two responses, it was included to focus participants toward the child/adult distinction, reflecting operational purposes.

Participants then completed age estimation tasks for 140 images in the same manner of the practice example. Following completion of the study, an open-ended question was asked regarding judgment factors considered during the task (see Appendix G). Participants supplied their e-mail address if they desired their results.

2.6 Dependent Measures

Measurements in this study were selected to analyse estimation performance and metacognition.

2.6.1 Measuring estimation performance. Based on Brown and Siegler's (1993) recommendations, two complementary calculations, accuracy and directional estimation error, described age estimation performance. Distinguishing these calculations was critical, as it was possible to display minimal bias (i.e., systematic over/under-estimation) while being highly inaccurate and vice-versa.

2.6.1.1 Accuracy. Accuracy was defined as the magnitude of estimation error. The mean absolute error (MAE) was calculated by the participant's mean absolute estimation deviation from the true age. Group means were then calculated for analyses. A MAE value closer to zero indicated increased accuracy (i.e., estimated ages were closer to the true ages).

2.6.1.2 Directional estimation error. Directional estimation error described the direction of estimation bias as under- or over-estimated from the true age. The mean signed difference between the estimated value and true value was calculated for each participant. Group means were then calculated for analyses. Negative values indicated average underestimations (i.e., estimated ages were younger than true ages) and positive values indicated overestimations (i.e., estimated ages were older than true ages).

2.6.2 Metacognition measures. Measures of bias score and precision range were employed to assess metacognitive abilities.

2.6.2.1 Bias score. Overconfidence was represented by a bias score, calculated by the given confidence level (i.e., 80%) minus the proportion of correct responses. A correct response was defined as the estimated age range containing the true age. A bias score of zero represented perfect calibration. Therefore, positive deviations from zero signified overconfidence (i.e., less than 80% of responses were correct) and negative bias scores signified under-confidence (i.e., more than 80% of responses were correct). The bias score has good internal consistency (Michailova, 2010), viable for describing self-monitoring cognitions (Stankov & Crawford, 1997).

To avoid large ranges required for 100% confidence and potentially forcing overconfidence (i.e., if participants were not correct in every estimate), the given confidence level was set at 80%.

2.6.2.2 Precision range. The precision range was calculated as the size of the estimated age range. A mean range was calculated for each participant dependent on the conditions. Small ranges indicated a precise estimate but did not qualify if the range was accurate (i.e., if the true age fell within the range).

Chapter 3

Results

Results were calculated for three aspects of performance: estimation performance, metacognition and reported features. A total of 6,440 estimates were obtained.

3.1 Data Screening and Assumptions

Data was initially screened for errors and normality. While no missing values were identified, outliers were calculated by the z-distance. Potential causes for outliers were reviewed (Osborne & Overbay, 2004). The experimental interface ensured consistency between responses to each image, decreasing the likelihood of errors. The outliers were determined to be legitimate cases of variability and remained within the dataset. Normality was tested using the Shapiro-Wilk Test (see Appendix K) and confirmed by visual analysis of Q-Q plots and histograms. Some data was normally distributed (directional estimation error and bias scores) and parametric tests were used. Paired samples t-tests compared mean differences between the conditions and effect sizes were reported as Cohen's *d* for repeated measures (Cohen, 1988). Effect sizes were interpreted as a: small = 0.2, medium = 0.5 and large = 0.8 effect. Given the importance of under- or over-estimation and overconfidence in refugee processing settings, a one-sample t-test was employed to determine if estimates significantly deviated from zero (noting zero indicates an accurate estimate/calibration). Additionally, a simple linear regression was applied to investigate the relationship of true age and estimated age. A standardised residuals scatterplot confirmed assumptions of homogeneity of variance and normality were met. Multicollinearity assumptions were also met (VIF = 1.00).

For other data (accuracy and precision range), calculation of *z*-values for skewness and kurtosis found values larger than the 1.96 criterion for normality, based on a sample size of 46 (Kim, 2013). This violated assumptions of normality and non-parametric tests were selected over data transformation as variables including accuracy have previously displayed skewness (Michalski, 2017), suggesting these distributions are valid representations. A Wilcoxon Signed-Ranks Test was used to compare differences between groups. Effect sizes were calculated as $r = \frac{z}{\sqrt{N}}$ and interpreted as a: small = 0.1, medium = 0.3 and large = 0.5 effect (Cohen, 1988). A Friedman Analysis of Variance (ANOVA) was used to compare more than two groups. When results were significant, post-hoc pairwise comparisons were conducted using the Wilcoxon Signed-Ranks Test with a Bonferroni correction. The correction reduced the chance of a Type I error, resulting in a significance criterion of $p < .013$. For all other analyses, two-tailed significance was set at $p < .05$.

3.2 Accuracy of Estimates

Table 1 provides descriptive statistics of the accuracy of age estimates. Outcomes are reported in years.

Table 1

Descriptive statistics of the mean absolute error (MAE)

Statistic	Overall	Child	Adult
N	46	46	46
M	3.22	2.58	3.86
Median	3.06	2.55	3.51
SD	0.94	0.71	1.34
Minimum	2.08	1.36	2.07
Maximum	6.62	5.19	8.06

The hypothesis that images of children would be more accurately estimated than images of adults was tested. In defining the stimuli as a child or adult, 81.8% of classifications were correct. Correct categorisation occurred more often for adults (*M* = 82.9%) than for children (*M* = 80.6%). However, descriptive statistics in Table 1 suggested estimates deviated less from the true age for children than for adults. A Wilcoxon

Signed-Ranks Test confirmed this observation and showed a statistically significant difference and large effect, $Z = -5.59$, $p < .001$, $r = -0.82$. This supports the hypothesis, as children's ages were more accurately estimated than adults' ages.

3.3 The Effect of Image Gender on Estimation Accuracy

Table 2 reports descriptive statistics of age estimation accuracy when viewing images of female and male faces for both age groups. Outcomes are reported in years.

Table 2

Descriptive statistics of the mean absolute error (MAE) across gender

Statistic	Child		Adult	
	Female	Male	Female	Male
N	46	46	46	46
M	2.80	2.36	3.84	3.89
Median	2.61	2.34	3.64	3.29
SD	0.94	0.58	1.12	1.75
Minimum	1.51	1.20	2.17	1.83
Maximum	5.77	4.60	7.37	9.49

It was hypothesised that estimations for male faces would be more accurate than females. Descriptive statistics (see Table 2) suggested children's ages were more accurately estimated than adults' ages, regardless of the person's gender. A Friedman ANOVA was conducted to compare if accuracy significantly differed for females and males of both age groups. Accuracy was statistically different across the groups, $\chi^2(3) = 63.66$, $p < .001$. Post-hoc pairwise comparisons were performed with a Bonferroni correction. Analyses revealed small, but statistically significant differences in accuracy between images of male children and male adults ($Z = -1.65$, $p < .001$, $r = -0.24$) and female children and female adults ($Z = -1.22$, $p < .001$, $r = -0.18$). Descriptive statistics in Table 2 suggested that for estimates of both females and males, accuracy was superior for children versus adults. There was a negligible effect with no statistical difference in accuracy between male children and female children ($Z = 0.63$, $p = .076$, $r = 0.09$) or male adults and female adults ($Z = -.20$, $p = 1.00$, $r = -0.03$). The hypothesis was not supported, as differences were only reflective of the age groups and gender did not affect accuracy.

3.4 Directional Estimation Error

The direction of estimation error was used to explore the hypothesis that the ages of children and adults would be overestimated. A visual representation of how the 6,440 estimates deviated from accuracy by age is presented in Figure 1. The horizontal line at the origin represents accuracy.

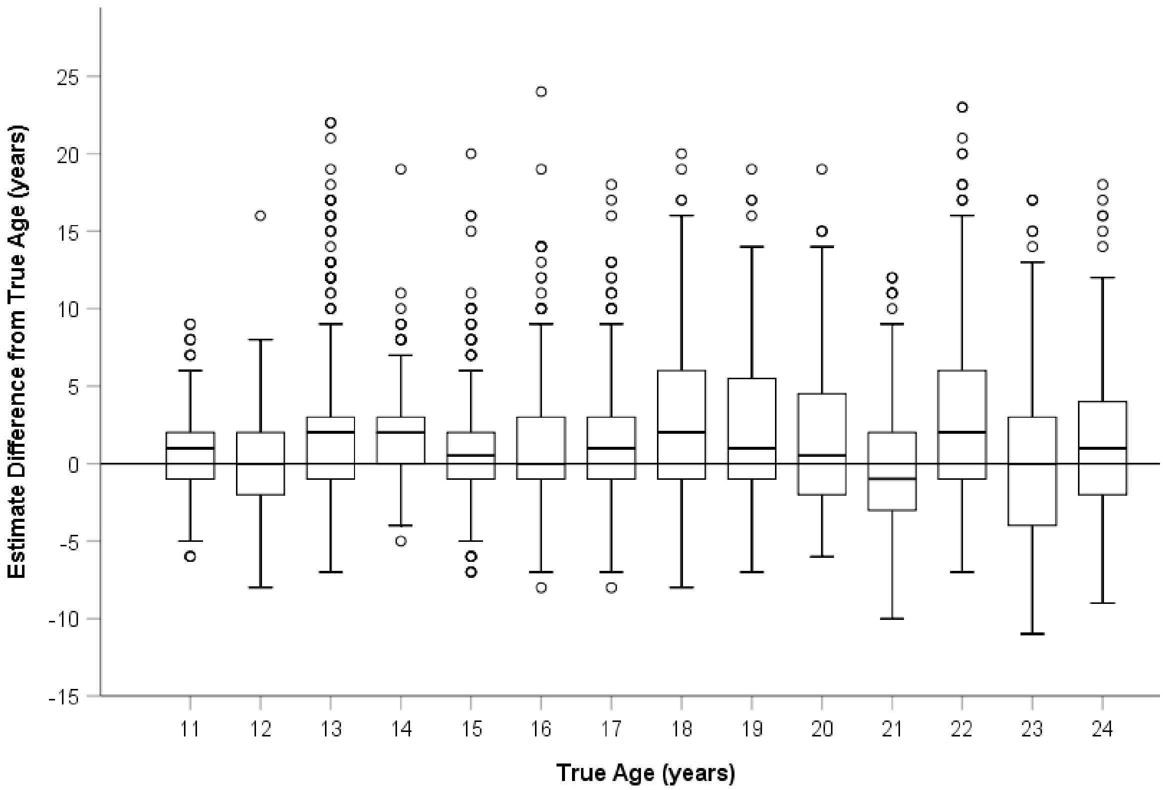


Figure 1. Boxplot of the directional estimation error by the true age of the stimuli.

The boxplot shows that almost every age was overestimated. Outliers were also present for every age indicating a large variance, with age being overestimated by up to 24 years and underestimated by 11 years.

To visualise the overall effect of increased true age on estimated age, a linear regression is displayed in Figure 2.

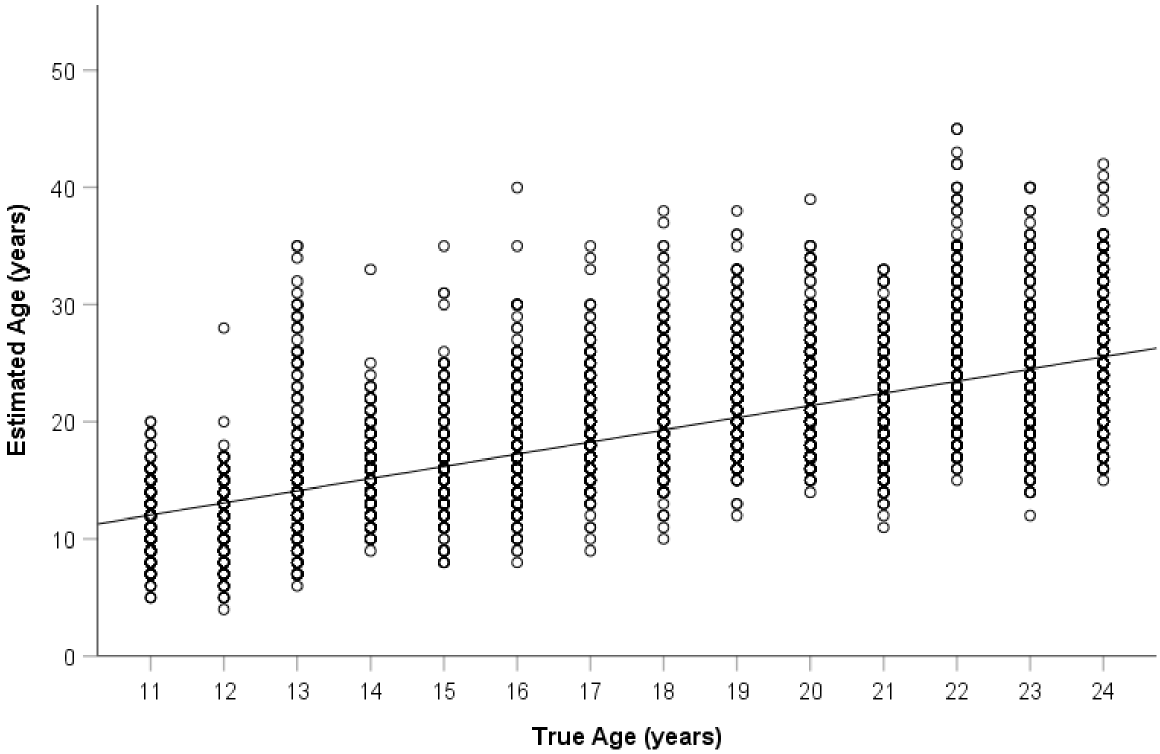


Figure 2. Scatter plot of the relationship between true age and estimated age.

A significant linear regression equation was identified, $F(1,6438) = 6242.66$, $p < .001$, $R^2 = .49$. Only 49% of the variance in estimates was explained by the true age. Participants' age estimates were equal to .62 + 1.04(True Age) when age is measured in years. Estimated age increased 1.04 years for each year increase of the persons' true age, indicating a tendency of overestimation. This was confirmed by a one sample t-test of directional estimation error scores, finding significant positive deviations from zero, $t(45) = 4.87$, $p < .01$, $d = 0.72$.

To explore if the directional estimation error differed for children and adults, descriptive statistics are reported in Table 3. The outcome is reported in years.

Table 3

Descriptive statistics for directional estimation error

Statistic	Overall	Child	Adult
N	46	46	46
M	1.29	1.07	1.52
Median	1.32	1.04	1.84
SD	1.80	1.31	2.65
Minimum	-1.71	-1.17	-3.00
Maximum	6.05	4.50	7.60

The positive values suggested a tendency to overestimate the true age of both children and adults. A one sample t-test confirmed estimates were positive and statistically significant from zero with a large effect for children ($t(45) = 5.53$, $p < .001$, $d = 0.82$) and medium effect for adults ($t(45) = 3.89$, $p < .001$, $d = 0.57$), supporting that children and adults were both systematically overestimated. However, a paired samples t-test suggested that the degree of overestimation did not significantly differ for children or adults, $t(45) = -1.44$, $p = .156$, $d = 0.21$.

In summary, the hypothesis that the ages of children and adults would be overestimated was supported, although there was no difference in the degree of overestimation.

3.5 Bias Score

Descriptive statistics of the bias score and precision range are provided in Table 4. A bias score of zero represented perfect calibration, positive deviations from zero signified overconfidence and negative scores signified under-confidence. Precision range outcomes are reported in years and values closer to zero signified a more precise range.

Table 4

Descriptive statistics for bias score and precision range

	Bias Score			Precision Range		
Statistic	Overall	Child	Adult	Overall	Child	Adult
N	46	46	46	46	46	46
M	0.22	0.18	0.25	5.50	4.67	6.33
Median	0.21	0.17	0.23	4.85	4.23	5.63
SD	0.14	0.15	0.17	1.96	1.57	2.43
Minimum	-0.04	-0.11	-0.07	2.69	2.59	2.79
Maximum	0.57	0.49	0.66	12.17	8.90	15.44

To assess if participants were under- or over-confident, the bias score was analysed. Descriptive statistics suggested participants were overconfident in both conditions (i.e., the proportion of correct responses were less than the given confidence level, 80%). A one-sample t-test confirmed significant positive deviations from an accurate calibration with a large effect overall ($t(45) = 10.53$, $p < .01$, $d = 1.55$), for children ($t(45) = 8.32$, $p < .01$, $d = 1.23$) and for adults ($t(45) = 10.23$, $p < .01$, $d = 1.51$). A paired samples t-test was used to compare participants' bias scores for children and adults. A significant, medium effect was found, $t(45) = -3.33$, $p < .01$, $d = -0.75$. This suggests participants displayed a higher degree of overconfidence when estimating adults' ages than children's ages (see Table 4). The hypothesis was supported, as participants were overconfident in their estimates of children and adults.

3.6 Precision of Range

Considering the difference in bias score, the size of the range was explored to provide an insight into participants' precisions of estimates. A Wilcoxon Signed-Ranks Test showed a strong, significant difference between estimates of children and adults, $Z = -5.91$, $p < .001$, $r = -0.87$. Descriptive statistics in Table 4 showed the range was more precise (i.e., smaller) for estimates of children. Despite adjusting to a larger range for adult estimates, it proved insufficient in aligning confidence with accuracy, or with estimates of children.

3.7 Features Considered in Age Estimations

Following the experiment, participants responded in free text format describing the factors they considered when making age judgements in the experiment. As the aim was to quantify emergent themes, content analysis was employed using the technique of repetition (Ryan & Bernard, 2003). The emergent themes and number of participants reporting each theme are displayed in Table 5. The majority of participants reported using the hair and skin as indicators of age, although a number of participants also reported using external features (e.g., makeup and clothing) to assist their judgements.

Table 5

Self-reported factors considered during age estimations

Theme	Subthemes	Number of Participants (%)
Hair	Facial hair; Hairline; Eyebrows; Grooming	29 (63.04%)
Skin	Acne; Blemishes; Youthfulness; Wrinkles; Tone; Smooth; Condition	25 (54.35%)
External items	Jewellery; Makeup; Clothing	19 (41.30%)
Facial features	Ears; Nose; Eyes; Look in eyes; Lips; Cheeks; Size; Proportion	17 (36.96%)
Face dimensions	Shape; Size; Fullness	15 (32.61%)
General Appearance	Believed life experiences; Development; Expression	12 (26.09%)
Decision anchors	School children, People of similar appearing ethnicity; Common associates; University students	8 (17.39%)

Chapter 4

Discussion

This study aimed to explore human performance of age estimation, using facial images, for refugee processing purposes. The results were varied in supporting the four hypotheses.

4.1 Accuracy of Age Estimates

Children's ages were more accurately estimated than adults', supporting the hypothesis. These results corroborate findings that accuracy decreases when the chronological age increases (Moyse, 2014). The overall MAE of 3.16 years for the present study further suggested performance superiority for estimates of younger faces when compared to Sörqvist and Eriksson (2007) result of 5.25 years for ages 56–65. Conforming to expectations, participants were attuned to perceiving ageing cues as estimates were significantly related to the true age (see Figure 2). Deviations from accuracy could be attributed to (a) a non-uniform display of ageing or (b) biases in human perception. For adult faces, the increased exposure to extrinsic factors likely introduced a higher level of variability, explaining a decrease in accuracy. In comparison, the distinctiveness of childhood ageing (Geng et al., 2010) proved advantageous for perceiving age related differences. Interestingly, processes indicative of cardiodial strain, a key childhood transformation, were reportedly considered by less than one-third of participants (see Table 5). Instead, signs of puberty (e.g., facial hair and acne) were more often reported, despite inherent individual variability (Marshall & Tanner, 1970). Grooming and external items were also commonly reported and as they reflect social norms and are easily manipulated, they are largely unreliable cues with low-cross ethnic reliability. This high reliance on less predictable features could explain the inaccuracy of estimates.

Biases may further account for errors in estimates. As the participant sample was selected to reflect operational contexts where adults are estimators, it was not possible to compare the age of the estimator to test for an own-age bias. Moyse and Brédart (2012) suggested this bias could be explained by the expertise hypothesis, where increased exposure to faces within ones own-age group, improves accuracy. Considering the participants' age (M = 39.76, SD = 13.23) and recruitment was within an adult-based work environment, it is possible that participants simply lacked familiarity with younger age-related changes.

Similarly, participants' ethnicity was different to that in the images, suggesting an own-race bias may have impacted judgments. The task was potentially prone to an increased level of difficulty associated with the cultural variability in ageing (Vashi et al., 2016). Again, a lack of familiarity with these differences may contribute to errors in perceiving and judging age.

4.2 Image Gender Affecting Estimation Accuracy

Age estimation accuracy was not superior for male faces compared with female faces. The hypothesis was not supported and this finding conflicted with previous research (Voelkle et al., 2012; Dehon & Brédart, 2001; Ferguson & Wilkinson, 2016). Methodology differences could explain the discrepancy as previous stimuli were aged 20 to 80 years (Voelkle et al., 2012) and 20 to 45 years (Dehon & Brédart, 2001). As the present study represented younger ages, it could be speculated that visible gender differences are prominent with increased age. Voelkle et al. (2012) proposed that females' attention to appear younger underlies less accurate estimations of females. As females maintain 'babyish' features (i.e., reduced wrinkles) for longer (Enlow et al., 1982), physical interventions (e.g., makeup or skincare) may not have completely commenced for visible change within the younger aged stimuli.

A discrepancy remains however as Ferguson and Wilkinson (2016) reported increased accuracy for male faces aged 11–15 years. As only a visual observation was reported, this effect may not have been statistically significant. The researchers also acknowledged the limited stimuli (four male, five female) and therefore, the results were prone to skew (i.e., if one image resulted in judgments outside the norm) with low internal validity. As task demands were similar, previous stimuli and analyses limitations may explain the discrepancy.

However, the literature analysis of ageing cues did suggest an advantage for perceiving male versus female children. Key distinctions were that males exhibit facial hair as a marker of puberty (Marshall & Tanner, 1970) and rapid growth of facial features until 16 years (Ferrario et al., 1998). Reiterating discussions in Section 4.1, facial hair was a commonly reported subtheme while facial features were not (see Table 5). This suggests many participants failed to utilise the appropriate cues. Alternatively, their cognitive calibration models may have failed to appropriately weight the value of these cues, subconsciously influencing judgments. These factors potentially contributed to the lack of evidence obtained for a gender difference.

4.3 Directional Estimation Error

The ages of children and adults were systematically overestimated, supporting the hypothesis and previous research (Vestlund et al., 2009; Rowe & Willner, 2001; Voelkle et al., 2012). As different motivational reasoning was supplied to contextualise the task from previous studies, overestimation may be a cognitive bias. Although the true explanation is unknown, these findings do not dispute the hypothesis of regressing to the population mean. All selected stimuli were younger than the Australian population median (ABS, 2016) and overestimation could reflect a tendency to assimilate estimates toward this average age. This would align with the present and previous studies results (Vestlund et al., 2009; Voelkle et al., 2012). Other research only presented partial support as young adults were underestimated (Rowe & Willner, 2001). However, the assessment of only two ages within this age category (20 and 22 years) resulted in low internal validity through the creation of systematic error.

Considering the visibility of age-related changes, these results may not accurately generalise to the age categories defined by the present study.

The stronger effect to overestimate children than adults presented as an interesting finding when the degree of overestimation was not significantly different. Although not specified in the hypothesis of regressing to the mean, logically, the degree of overestimation would be stronger for children as they are further from the population mean. A limitation of the directional estimation measure is the cancellation of errors when averaged. As observed in Figure 1, the upper quartiles for adults were large relative to the lower quartiles and outliers for adults were only overestimations. Potentially, the minority values skewed mean results to depict a higher degree of systematic overestimation.

4.4 Overconfidence

Participants were overconfident in their age estimations of children and adults, supporting the hypothesis. Although not previously explored within the domain of age estimation, this finding is consistent with suggestions of overconfidence as a consistent cognitive distortion affecting beliefs and judgments (Russo & Schoemaker, 1992). Overconfidence in the present study may result from limited exposure with familiar faces of a known age identity and minimal disconfirmation of age judgments with unfamiliar faces, creating a false sense of accuracy. Anchoring decisions on people with a known age can be particularly prone to error if metacognitive judgments fail to recognise the limitations of the samples' representativeness. As discussed in Section 4.1, the participants likely experienced increased interactions with adults, potentially explaining the higher false sense of confidence when estimating adults. Despite this, participants did demonstrate insight into the challenges for perceiving adult ageing as ranges were adjusted to reflect a larger degree of uncertainty. This could be interpreted as relinquishing some degree of precision in favour of accuracy (Yaniv & Foster, 1995). In light of overall positive bias scores though, it can be speculated that precision and the need to be informative generally outweighed cognitive demands of providing an accurate estimate.

4.5 Implications

Potential implications are speculative given that this method is currently under-researched and as such, generalisations are cautioned. Applying the present findings to assess the method's feasibility within refugee processing requires recognition of the biases in perception. Although children's ages are more likely to be estimated accurately, this study suggests a high level of error involved in the isolated viewing of a face. The likelihood of over- as opposed to under-estimation would see more children labelled as adults, than adults labelled as children. Initially, these children could be wrongfully detained in immigration facilities (Feltz, 2015; Button & Evans, 2017). Post processing, restrictions in accessing child-designated support including housing and education may risk the physical and mental wellbeing of the child (Aynley–Green, 2011). Conversely, an adult is more likely to be correctly categorised and assigned to the designated processing systems, preventing the expenditure of resources designated for children and potential risks of an adult in a child-based environment. Current results indicate consideration toward an accurate estimate is not dependent on gender although the inconsistency of this finding requires further investigation. Incorporating the likelihood of overconfidence, the assessors' may be prevented from evaluating their decision as inaccurate or believe they are correct more often than reality. Unwarranted confidence that an estimate is correct would be to a lesser degree when estimating children, with more precision to the possible age range provided.

Given the potential implications, this age estimation methodology lacks veracity for application as the sole process in refugee settings. This is not to understate the potential value of this methodology; rather to highlight particular considerations required to achieve the best possible outcome. In light of current methodologies used such as carpal x-ray, with an error rate of approximately two years (Chaussain & Chapuis, 2007 as cited in Feltz 2015), humans' estimates from facial images indicates a higher level of error. However, participants were novices to the task and prone to bias, warranting further analysis.

Recognition of biases and acceptance of hypotheses seeking to explain the observed effects have implications for future research and application. If the expertise hypothesis were true, a logical implication to improve performance is increasing familiarity. Estimates of the present study were reportedly formed on cues with high variability. Therefore, developing ones perceptual space should create a new cognitive representation of the population and ethnic differences of ageing. Limited research has indicated task training may be a successful method to improve performance, as alcohol salespersons significantly outperformed novices (Vestlund et al., 2009). The naturalistic setting of training entailed two key aspects of (a) increased exposure and (b) feedback training. Sörqvist and Eriksson (2009) found feedback training did not increase performance for stimuli aged 15–24 years. Considering alcohol salespersons did exhibit improved performance (Vestlund et al., 2009), training may require prolonged exposure to increase the perceptual representation of younger faces. As overconfidence is proposed to result from a limited exposure with known faces and minimal disconfirmation with unknown faces, this training may realign metacognitive judgments with accuracy. However, as training in Vestlund et al.'s (2009) study occurred in a naturalistic setting, experience is not causal to improved performance, highlighting the need for future experimental research.

4.6 Limitations

Limitations of the study potentially minimise the generalisability. The present study focused on the initial stages of refugee processing where age decisions for a child or adult is critical. Although a conscious design, the restricted age range prevents generalisations regarding perception to other aged people. Considering differences in ageing cues and the variance in estimates, the ages not explored may also be vulnerable to incorrect categorisation.

The motivational reasoning provided to participants sought to contextualise refugee processing. An incorrect estimate likely resulted in personal dissatisfaction. In refugee settings there is an inevitably higher degree of responsibility and potential risk associated with inaccuracy (e.g., mislabelling a child as an adult). As motivation biases decision-making (Kunda, 1990), participants' reasoning and cognitive investment in their decision may have constrained judgments, potentially limiting population validity.

Finally, the image criteria may have limitations. Although the Middle East represents the highest origin countries, refugee backgrounds are extremely diverse. Limited research suggests that performance differs for other ethnicities (Dehon & Brédart, 2001) and without definite evidence; the present results are limited to generalisations of Middle Eastern appearing individuals. As images were not of refugee applicants, ageing cues did not likely represent the variation caused by extrinsic factors. Stress, sleep deprivation and malnutrition are prevalent among refugees (Kueper et al., 2015) and alter visible ageing (Taister et al., 2000). This would increase individual variation and potentially task difficulty. Performance in the present study may therefore be skewed to suggest higher levels of accuracy.

4.7 Strengths

Previous research exploring child/adult age estimation differences was limited to the contextual paradigm of alcohol sales (Vestlund et al, 2009; Rowe & Willner, 2001). It appears this was the first study to explore this task for refugee processing contexts. An appropriate selection of stimuli ethnicity and motivational circumstance provided to participants increased the study's ecological validity, allowing for specific generalisations. Further extending beyond the age estimation literature focused on accuracy, the assessment incorporated metaknowledge and self-reports of cues used in decisions as additional performance indicators. Assessing multiple dimensions of ability and experience contributed to understanding human perception and practical implications of this procedure, depicting a more realistic account.

Another strength was the appropriate methodology. This study incorporated a large image set that equally represented age and gender, addressing limitations of previous research (Vestlund et al., 2009; Rowe & Willner, 2001; Ferguson & Wilkinson, 2016). Therefore, the effect of skewed outcomes associated with the images was controlled (i.e., if estimates of one variable differed from the norm). Combined with the within-subjects design, many observations were obtained from fewer participants. This increased the study's internal validity, allowing for unambiguous performance comparisons by controlling for individual differences. In summary, these strengths suggest the findings will meaningfully contribute to the literature of human perception in the domain of age estimation.

4.8 Directions for Future Research

As an exploratory study in an emerging field, further research is required to conceptualise the nature of age estimation. Although methodological differences sought to explain the lack of consensus with previous studies, this is only speculative and highlights the need for ongoing research. Key directions for future research are suggested as: development of the performance baseline, task training and method comparisons.

Future studies could adopt a similar methodology to the present study, with the inclusion of a representative image set that addresses current limitations. For example, researching the own-race bias by manipulating the stimuli's ethnicity could explore how performance varies. Similarly, a larger stimuli age range would increase understanding of performance differences and how this relates to visible ageing cues. Broadly, this would contribute to the literature through defining group-specific biases in perception. From an applied perspective, appropriate adjustments based on biases could be recognised for refugee assessments. Of equal importance is the investigation of the natural differences in accuracy. Despite all participants being novices, performance scores varied and the true age did not explain all of the variance in estimates (see Section 3.4). Future research could continue the literature exploring demographic (e.g., age) and situational differences (e.g., exposure to children) affecting performance. For specific generalisations, immigration officials could partake as participants in future research. As their employment entails increased exposure to refugees and motivational importance of a correct decision, performance and biases may differ and these results could increase understanding of this method's viability in practical settings.

The potential for task training presented as a theme in Section 4.5. Once group-specific biases are identified, future research could focus on developing cue-related training. For example, defining how chronological age and ethnicity affects judgments could allow appropriate exposure training with age information feedback. The perceptual face space should become balanced and participants would be expected to then decrease their reliance on comparatively few decision anchors, increasing overall accuracy. As a process, increased experience could aid in developing metacognition abilities.

A comprehensive understanding of human potential using facial images for age estimation would allow for comparative studies. Considering current procedures to estimate ages are prone to error, further research is required to quantify the relative accuracy of methods. Through comparing accuracy of age estimations from facial images with currently used methods (e.g., carpal X-ray and chronological age assessments), government agencies could be informed of the best procedure or combination thereof, to achieve the appropriate outcome.

4.9 Conclusion

The present study aimed to explore human performance of age estimation, using facial images, for refugee processing purposes. The study demonstrated that estimates of children were more accurate than of adults, although both groups were significantly overestimated. Furthermore, the gender of the stimuli did not significantly affect accuracy. Finally, there was a display of overconfidence when estimating both age groups, although this bias was more pronounced with less precision when estimating adults than children. Relevant government agencies should be aware of the inaccuracy and biases of judgments that increase the likelihood of misidentification, and the potential implications if used within refugee processing. The present findings highlight the need for future research to further understand the feasibility of facial images as a method for age estimation.

References

Albert, A. M., Ricanek Jr, K., & Patterson, E. (2007). A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, *172*(1), 1–9. doi:10.1016/j.forsciint.2007.03.015

Amnesty International. (2017). What's the difference between a refugee and an asylum seeker? Retrieved July 19, 2018. Retrieved from https://www.amnesty.org.au/refugee-and-an-asylum-seeker-difference/

Anastasi, J. S., & Rhodes, M. G. (2006). Evidence for an own-age bias in face recognition. *North American Journal of Psychology*, *8*(2). Retrieved from https://search-proquest-com.proxy.library.adelaide.edu.au/docview/198016872?accountid=8203&rfr_id=info%3Axri%2Fsid%3Aprimo

Australian Border Force. (2018). Immigration Detention and Community Statistics Summary. *Department of Immigration and Border Protection*. Retrieved from https://www.homeaffairs.gov.au/ReportsandPublications/Documents/statistics/immigration-detention-statistics-june-18.pdf

Australian Bureau of Statistics (ABS). (2016). Main Features - Main Features. Retrieved July 21, 2018, from http://www.abs.gov.au/ausstats/abs@.nsf/mf/3235.0

Aynley–Green, A. (2011). The assessment of age in undocumented migrants. *A report for the office of the Denfensor del Pueblo, Madrid, Spain.* Retrieved from http://humanrights. gov. au/ageassessment/submissions/Sir% 20Al% 20Aynsley-Green% 20Kt% 20 (Submission% 2038).pdf.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological review*, *100*(3), 511.

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

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243-248. doi:10.1111/1467-9280.00144

Button, L., & Evans, S. (2017). At What Cost? The Human, Economic and Strategic Cost of Australia's Asylum Seeker Policies and the Alternatives'. Retrieved from https://www.unicef.org.au/Upload/UNICEF/Media/Documents/At-What-Cost-Report.pdf

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

Dantcheva, A., Chen, C., & Ross, A. (2012). Can facial cosmetics affect the matching accuracy of face recognition systems? In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on* (pp. 391–398). IEEE. doi:10.1109/BTAS.2012.6374605

Dehon, H., & Brédart, S. (2001). An 'other-race' effect in age estimation from faces. *Perception*, *30*(9), 1107-1113. doi:10.1068/p3122

Drury, I. (2018). Two-thirds of 'child refugees' quizzed about their age are found to be ADULTS. Retrieved July 11, 2018. Retrieved from http://www.dailymail.co.uk/news/article-5592357/Two-thirds-child-refugees-lied-age-18.html

Enlow, O. D., Pfister, C., Richardson, E., & Kuroda, T. (1982). An analysis of Black and Caucasian craniofacial patterns. *Angle Orthodontist*, *52*, 279–287. doi:10.1043/0003-3219(1982)052<0279:AAOBAC>2.0.CO;2

European Asylum Support Office EASO (2018). EASO Practical Guide on Age Assessment: Second Edition. Retrieved May 17, 2018. Retrieved from https://www.easo.europa.eu/sites/default/files/easo-practical-guide-on-age-assesment-v3-2018.pdf

Feltz, V. (2015). Age assessment for unaccompanied minors. When European countries deny children their childhood.

Ferguson, E., & Wilkinson, C. (2017). Juvenile age estimation from facial images. *Science & Justice*, *57*(1), 58-62. doi:10.1016/j.scijus.2016.08.005

Ferrario, V. F., Sforza, C., Poggio, C. E., & Schmitz, J. H. (1998). Facial volume changes during normal human growth and development. *The Anatomical Record: An Official Publication of the American Association of Anatomists*, *250*(4), 480–487. doi:10.1002/(SICI)1097-0185(199804)250:4<480::AID-AR12>3.0.CO;2-K

Fink, B., & Matts, P. J. (2007). The effects of skin colour distribution and topography cues on the perception of female facial age and health. Journal of the European Academy of Dermatology and Venereology, 22, 493498. doi:10.1111/j.1468-3083.2007.02512.x.

Geng, X., Fu, Y., & Smith-Miles, K. (2010). Automatic facial age estimation. In *11th Pacific Rim International Conference on Artificial Intelligence* (pp. 1–130). Retrieved from http://www1.ece.neu.edu/~yunfu/papers/pricai10_t4.pdf

George, P. A., & Hole, G. J. (1998). The influence of feature-based information in the age processing of unfamiliar faces. *Perception*, *27*(3), 295–312.

Grd, P. (2013). Introduction to human age estimation using face images. *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, *21*(Special Issue), 24-30. doi:10.2478/rput-2013-0005

Guo, G., Wen, L., & Yan, S. (2014). Face authentication with makeup changes. *IEEE Transactions on Circuits and Systems for Video Technology*, *24*(5), 814–825. doi:10.1109/TCSVT.2013.2280076

Han, H., Otto, C., & Jain, A. K. (2013). Age estimation from face images: Human vs. machine performance. *Biometrics (ICB), 2013 International Conference,* 1-8. doi:10.1109/ICB.2013.6613022

Henss, R. (1991). Perceiving age and attractiveness in facial photographs. Journal of Applied Social Psychology, 21, 933–946. doi:10.1111/j.1559-1816.1991.tb00451.x

Hole, G., & George, P. (2011). Evidence for holistic processing of facial age. *Visual Cognition*, *19*(5), 585-615. doi:10.1080/13506285.2011.562076

Hurley, M., & Beaumont, E. (2016). Reforming Australia's age determination procedures: Giving asylum seekers the benefit of the doubt. Alternative Law Journal, 41(1), 30–33. doi:10.1177/1037969X1604100108

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. doi:10.1016/0010-0277(91)90045-6

Jones, G., & Smith, P. K. (1984). The eyes have it: Young children's discrimination of age in masked and unmasked facial photographs. *Journal of Experimental Child Psychology*, *38*(2), 328–337. doi:10.1016/0022-0965(84)90129-2

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

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52–54. doi:10.5395/rde.2013.38.1.52

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

Kueper, J., Beyth, S., Liebergall, M., Kaplan, L., & Schroeder, J. E. (2015). Evidence for the adverse effect of starvation on bone quality: a review of the literature. *International Journal of Endocrinology*, *2015*. doi:10.1155/2015/628740

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480. doi:10.1037/0033-2909.108.3.480

Laughery, K. R., Rhodes, B., & Batten, G. (1981). Computer-guided recognition and retrieval of facial images. *Perceiving and remembering faces*, (pp. 251-271). London: Academic Press.

Lee, P. A. (1980). Normal ages of pubertal events among American males and females. *Journal of Adolescent Health Care*, *1*(1), 26-29. doi:10.1016/S0197-0070(80)80005-2

Marshall, W. A., & Tanner, J. M. (1970). Variations in the pattern of pubertal changes in boys. *Archives of Disease in Childhood*, *45*(239), 13-23. doi:10.1136/adc.45.239.13

Michailova, J. (2010). Development of the overconfidence measurement instrument for the economic experiment. Retrieved from https://mpra.ub.uni-muenchen.de/34475/1/MPRA_paper_34475.pdf

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

Montpare´, J. M., & McArthur, L. Z. (1986). The influence of facial characteristics on children’s age perception. *Journal of Experimental Child Psychology*, *42*(3), 303314.

Moyse, E. (2014). Age Estimation from Faces and Voices: A Review. Psychologica Belgica, 54(3), 255–265, doi:10.5334/pb.aq

Moyse, E., & Brédart, S. (2012). An own-age bias in age estimation of faces. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *62*(1), 3-7. doi:10.1016/j.erap.2011.12.002

O'Toole, A., Weimer, S., Dunlop, J., Barwick, R., Ayyad, J., & Phillips, J. (2010). Recognizing people from dynamic video: Dissecting identity information with a fusion approach. *Journal of Vision*, *10*(7), 643-643. doi:10.1167/10.7.643.

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

Office of the United Nations High Commissioner for Refugees (UNHCR). (2017). *Global Trends: Forced Displacement in 2016*. United Nations High Commissioner for Refugees.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*(6), 1–12. Retrieved from https://www.researchgate.net/profile/Jason_Osborne2/publication/242073851_The_Power_of_Outliers_and_Why_Researchers_Should_Always_Check_for_Them/links/0c96053606c3cb1b2f000000/The-Power-of-Outliers-and-Why-Researchers-Should-Always-Check-for-Them.pdf?origin=publication_detail

Pelham, B., & Blanton, H. (2012). *Conducting research in psychology: Measuring the weight of smoke*. Nelson Education.

Pittenger, J. B., & Shaw, R. E. (1975). Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(4), 374. doi:10.1037/0096-1523.1.4.374

Refugee Council of Australia (2010). *Myths and Facts about Refugees and Asylum Seekers,* Retrieved July 24, 2018 from https://www.refugeecouncil.org.au/docs/news&events/rw/2010/4%20-%20Myths%20and%20facts%20about%20refugees%20and%20asylum%20seekers%202010.pdf

Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(1), 1-12. doi:10.1002/acp.1442

Ricanek, K., Mahalingam, G., Albert, A. M., & Vorder Bruegge, R. W. (2013). Human face ageing: A perspective analysis from anthropometry and biometrics. In M. Fairhurst (Ed.), *Age factors in biometric processing* (pp. 93–116). London, UK: The Institution of Engineering and Technology. doi:10.1049/PBSP010E_ch5

Rowe, G., & Willner, P. (2001). Alcohol servers' estimates of young people's ages. *Drugs: Education, Prevention and Policy*, *8*(4), 375-383. doi:10.1080/0968763001001 9299

Russo, J.E., & Schoemaker, P.J.H. (1992). Managing overconfidence. *Sloan Management Review*, *33*(2), 7-17.

Russo, J.E., Schoemaker, P.J.H. (2016) Overconfidence. In: Augier M., Teece D. (eds) *The Palgrave Encyclopedia of Strategic Management*. Palgrave Macmillan, London. doi:10.1057/978-1-349-94848-2_323-1.

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, 15(1), 85–109. doi:10.1177/1525822X02239569

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. doi:10.1006/ceps.1999.1020

Solari, A. C., & Abramovitch, K. (2001). The accuracy and precision of third molar development as an indicator of chronological age in Hispanics. *Journal of Forensic Science*, *47*(3), 531–535. doi:10.1520/JFS15292J

Sörqvist, P., & Eriksson, M. (2007). Effects of training on age estimation. *Applied Cognitive Psychology*, 21, 131–135. doi:10.1002/acp.1271

Sörqvist, P., Langeborg, L., & Eriksson, M. (2011). Women assimilate across gender, men don't: the role of gender to the own-anchor effect in age, height, and weight estimates 1. *Journal of Applied Social Psychology*, *41*(7), 1733-1748. doi:10.1111/j.1559-1816.2011.00774.x

Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence*, *25*(2), 93-109. doi:10.1016/S0160-2896(97)90047-7

Taister, M. A., Holliday, S. D., & Borrmman, H. I. M. (2000). Comments on facial aging in law enforcement investigation. *Forensic Science Communications*, *2*(2). Retrieved from http://go.galegroup.com.proxy.library.adelaide.edu.au/ps/anonymous?id=GALE%7CA137863421&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=15288005&p=AONE&sw=w

Thevissen, P., Kvaal, S. I., Dierickx, K., & Willems, G. (2012). Ethics in age estimation of unaccompanied minors. *Journal of Forensic Odonto-Stomatology*, *30*(1), 85–102. Retrieved from https://www.dentalage.co.uk/wp-content/uploads/2014/09/thevissen_p_et_al_2012_ethics_age_estimation.pdf

The United Nations Children's Fund (UNICEF). (2017). A simplified version of the United Nations Convention on the Rights of the Child. Retrieved from https://www.unicef.org.au/Upload/UNICEF/Media/Our%20work/childfriendlycrc.pdf

The United Nations Children's Fund (UNICEF). (2018). Birth registration - UNICEF DATA. Retrieved July 11, 2018, from https://data.unicef.org/topic/child-protection/birth-registration/

The United Nations High Commissioner for Refugees (UNHCR). (2017). *Global Trends: Forced Displacement in 2016*. United Nations High Commissioner for Refugees.

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications, 21*, 4250. doi:10.1109/ 38.946630

Toseeb, U., Keeble, D. R., & Bryant, E. J. (2012). The significance of hair for face recognition. *PloS one*, *7*(3), e34144. doi:10.1371/journal.pone.0034144

Vashi, N. A., Maymone, M. B. D. C., & Kundu, R. V. (2016). Aging differences in ethnic skin. *The Journal of Clinical and Aesthetic Dermatology*, *9*(1), 31. Retrieved from https://www-ncbi-nlm-nih-gov.proxy.library.adelaide.edu.au/pmc/articles/PMC4756870/

Vestlund, J., Langeborg, L., Sörqvist, P., & Eriksson, M. (2009). Experts on age estimation. *Scandinavian Journal of Psychology, 50*(4), 301-307. doi:10.1111/j.1467-9450.2009.00726.x

Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging*, *25*, 265– 277. doi:10.1037/a0025065

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, *124*(4), 424.

doi:10.1037//0096-3445.124.4.424

Young, A. W., & Ellis, H. D. (Eds.). (2013). *Handbook of research on face processing*. Netherlands: Elsevier

Appendices

Appendix A – Recruitment Poster

CAN YOU DETERMINE IF THIS PERSON IS A CHILD OR AN ADULT?

When meeting a new person, age estimations are often made as a process of profiling an individual. Have you ever wondered how accurate you are at this judgment?

A Psychology Honours student from the University of Adelaide and DST Group Biometrics researchers, Dr Dana Michalski and Dr Veneta MacLeod are seeking volunteers to answer this question. This study aims to understand the accuracy of human age estimation using facial imagery for refugee processing purposes. If you wish to participate, you will be required to make age judgments from facial images.

This study is expected to take 40 minutes to one hour.

Everybody is invited to participate. Snacks will be available as well as the opportunity to receive your personal results.

DATES: 25/06/2018 – 28/06/2018
LOCATION: Within the 75 labs at DST Group Edinburgh

[redacted]

Appendix B – DST Intranet Site Recruitment Post

CAN YOU DETERMINE IF THIS PERSON IS A CHILD OR AN ADULT?

A Psychology Honours student from the University of Adelaide and DST Group Biometrics researchers, Dr Dana Michalski and Dr Veneta MacLeod are running a study and are seeking responses from individuals that are interested in participating.

The study will investigate the accuracy of human age estimation using facial imagery for refugee processing. Refugee policies are designed so that the treatment and processing of children varies from that of adults in order to provide an appropriate and fair processing for the applicant and destination country. However, the true age of the applicant is often unknown or unreliable. The outcomes of this research could provide agencies with a better understanding of the viability of this age estimation method and could contribute to software development of automatic age estimation.

This study will require you to estimate the age of people from facial images. You will also be asked to rate your confidence for each of your decisions. In addition to the experiment, you will be asked to partake in pre-study questions that will be examined for demographic and situational factors that may affect age judgments. A post study questionnaire will ask you to describe your decision making process.

Details:

Expected study duration: 40 minutes to one hour.

Date: The study will be held on 25/06/2018 – 28/06/2018

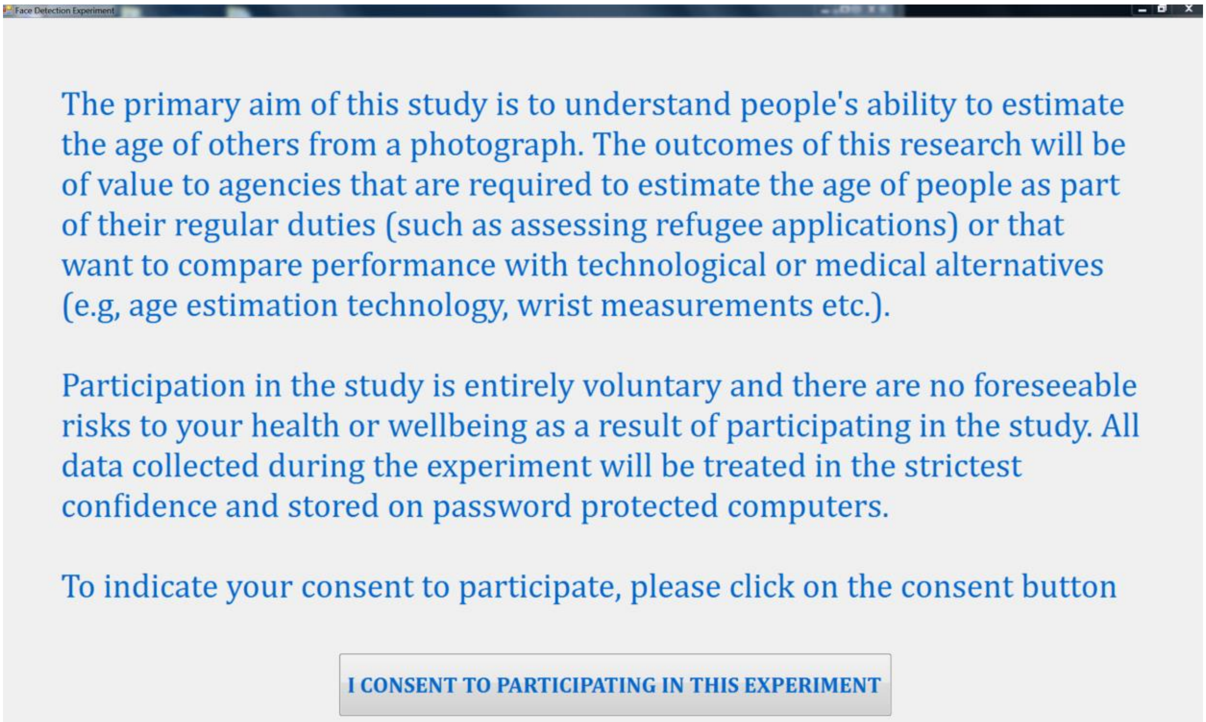
Location: Within the 75 labs at DST Group Edinburgh

Everybody is invited to participate. There will be snacks available during the study. You will also have the opportunity to receive your personal results.

If you would like to participate or need more information, please contact Gemma via email:

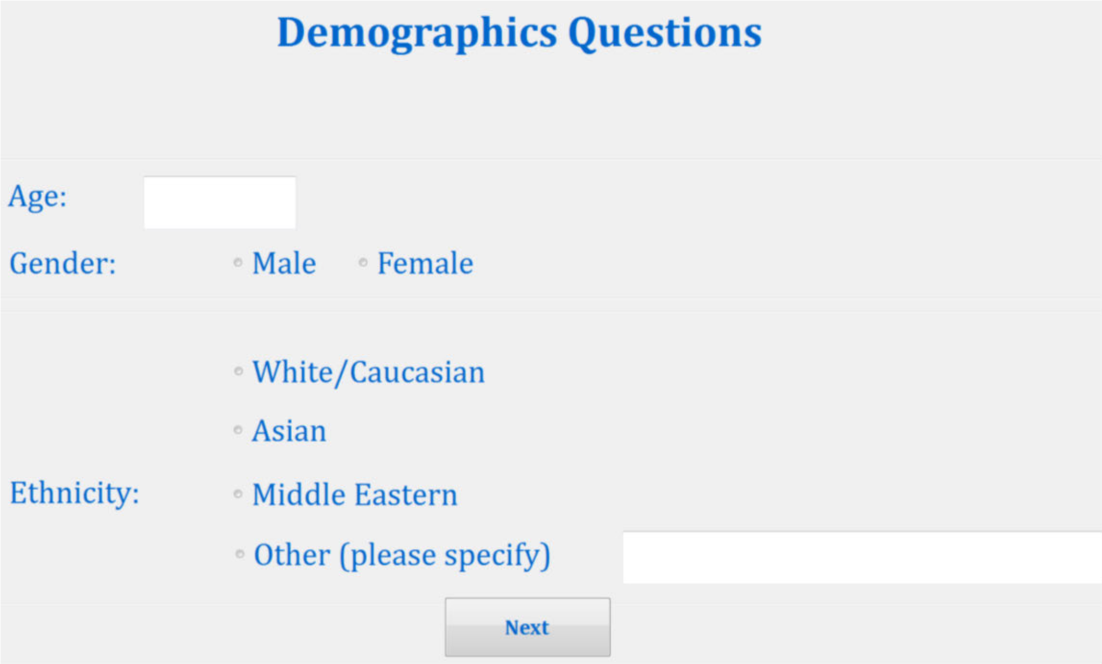
[redacted]

Appendix C – Participation Consent



Appendix D – Demographic Questions

(a)



Demographics Questions

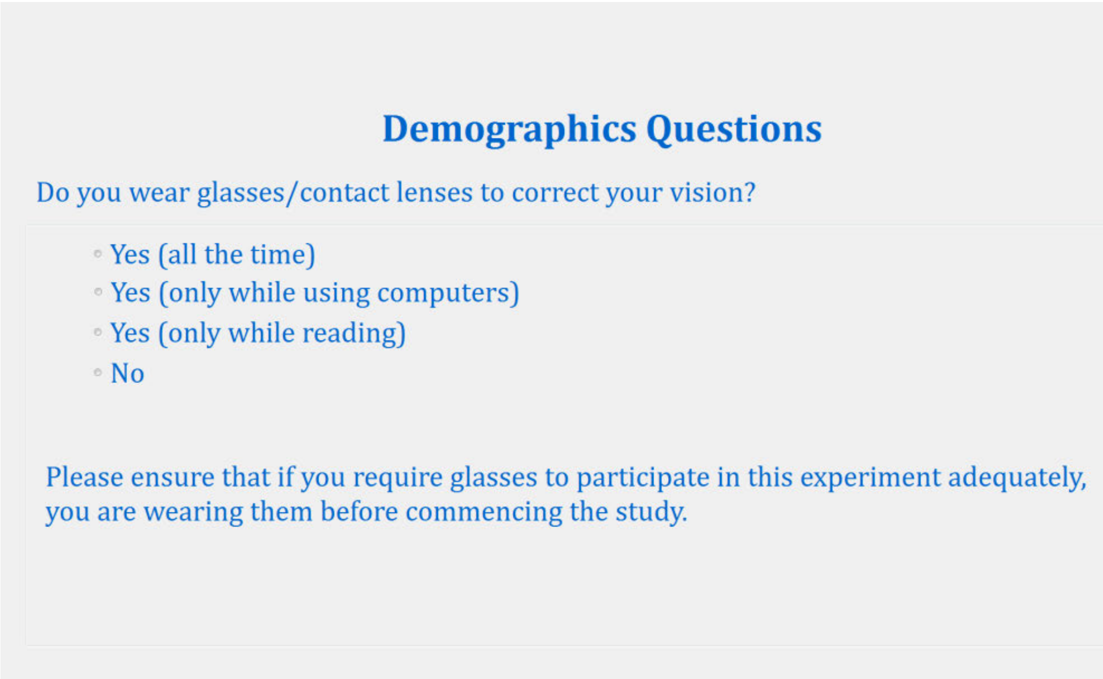
Age:

Gender: ◦ Male ◦ Female

Ethnicity:
- ◦ White/Caucasian
- ◦ Asian
- ◦ Middle Eastern
- ◦ Other (please specify)

Next

(b)



Demographics Questions

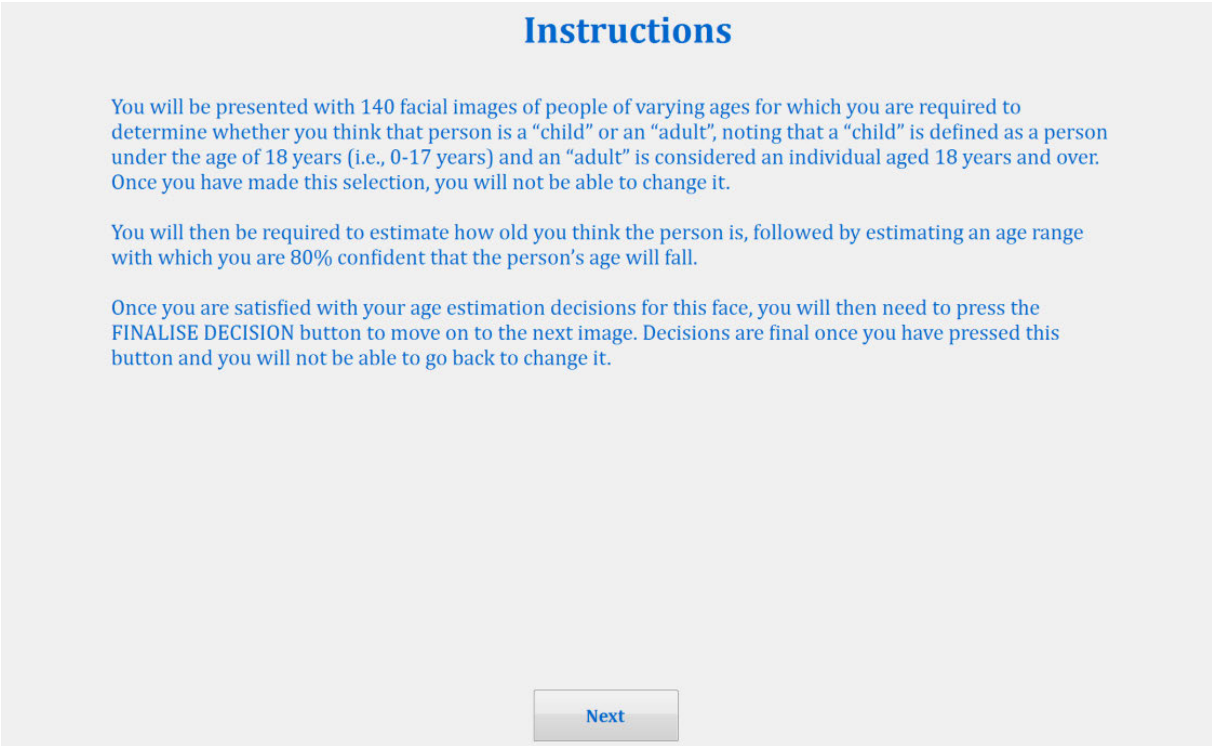
Do you wear glasses/contact lenses to correct your vision?

- ◦ Yes (all the time)
- ◦ Yes (only while using computers)
- ◦ Yes (only while reading)
- ◦ No

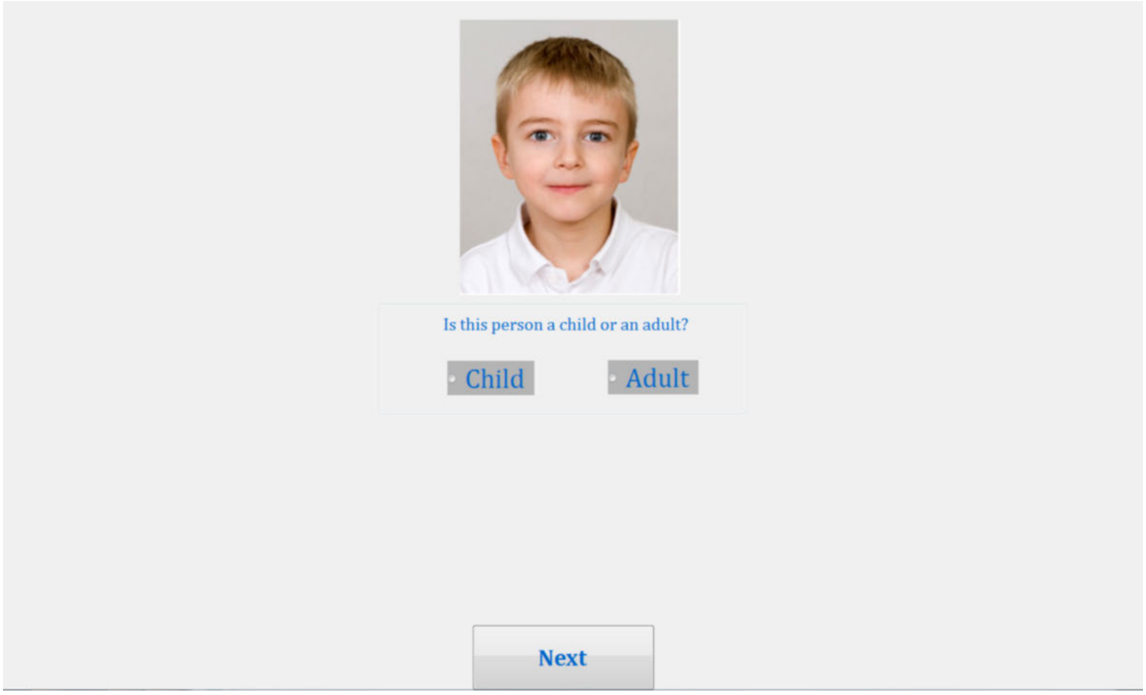
Please ensure that if you require glasses to participate in this experiment adequately, you are wearing them before commencing the study.

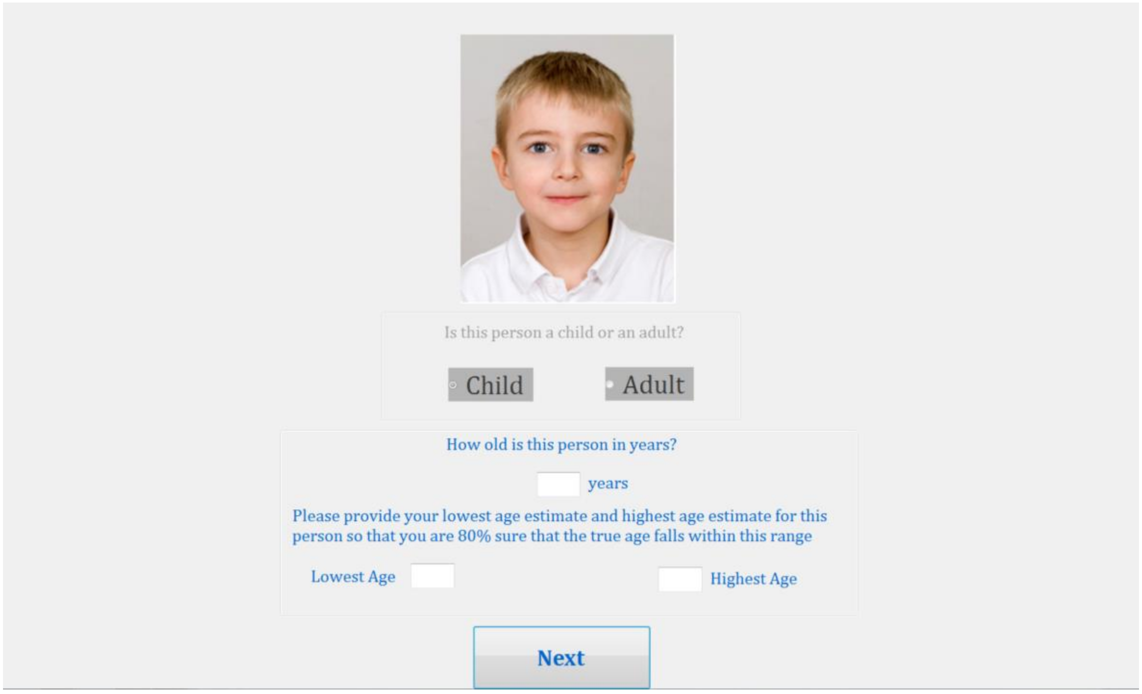
Note. (a) Requests participants' demographic information, (b) asks participants about their vision.

Appendix E – Instructions



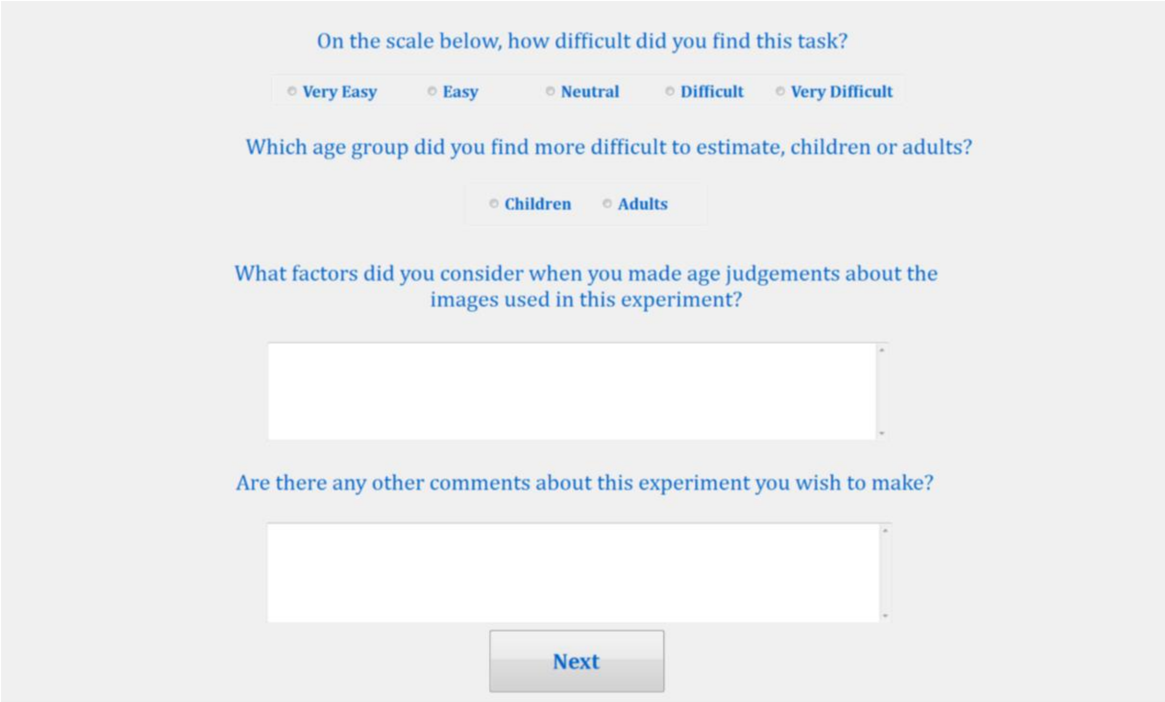
Appendix F – Practice Examples





Note. (a) is the initial practice trial requiring a child/adult categorisation and (b) is the following screen requiring an age estimate and upper and lower bounds of the estimate range. Privacy restrictions of the image database prevented the use of an image representative of the selection criteria for the practice trial.

Appendix G – Post-experiment Questions



On the scale below, how difficult did you find this task?

Very Easy | Easy | Neutral | Difficult | Very Difficult

Which age group did you find more difficult to estimate, children or adults?

Children | Adults

What factors did you consider when you made age judgements about the images used in this experiment?

Are there any other comments about this experiment you wish to make?

Next

Appendix H – Participant Information Sheet

PARTICIPANT INFORMATION SHEET

PROJECT TITLE: THE ACCURACY OF HUMAN AGE ESTIMATION FROM FACIAL IMAGES
HUMAN RESEARCH ETHICS COMMITTEE APPROVAL NUMBER: [redacted]

PRINCIPAL INVESTIGATOR: DR DANA MICHALSKI (Defence Science Technology Group)
EXTERNAL RESEARCHER: DR VENETA MACLEOD (Defence Science Technology Group)
STUDENT RESEARCHER: GEMMA SNYDER (University of Adelaide)
STUDENT'S DEGREE: BACHELOR OF PSYCHOLOGICAL SCIENCE- HONOURS

What is the project about?
Judgements of age are constantly made to create a perceptual profile of an unfamiliar individual. The accuracy of these judgements are increasingly important in settings, such as refugee processing, where the true age of an applicant is unknown or unreliable. Refugee policies are designed so that the treatment and processing of children varies from that of adults, to provide an appropriate and fair processing for the applicant and host country. Currently, age estimation is typically predicted by medical examinations including bone maturity by radiograph, physical development by medical examination and analysis of dental development. These methods can vary in accuracy, expense, time and invasiveness. The aim of this study is to understand the accuracy of age estimations and the classification of child or adult from facial images. The outcomes of this research could provide agencies with a better understanding of facial age estimations in comparison to other techniques. This project is being conducted for part fulfilment of the requirements for an Honours Degree at the University of Adelaide, under the supervision of researchers from the Defence Science and Technology Group.

Inclusion/Exclusion Criteria. You must bring and wear any required vision correction, such as glasses or contact lenses.

Your part in the study. You will be asked to conduct a series of computer-based tasks where you will make age estimation determinations about the individuals in the images. Participation in the study is entirely voluntary; there is no obligation to take part in the study, and if you choose not to participate there will be no detriment to your career or future health care. You have the right to withdraw at any time and can do this by contacting the lead investigator and quoting the Unique ID number that you will be assigned during the experiment. You should take note of this before the experiment begins. If you wish, you may receive a copy of your results by providing your email address to the researchers. Participation in this study is expected to take approximately 40 minutes to 1 hour.

Risks of participating. There are no foreseeable risks to your health or wellbeing as a result of participating in this study. Although unlikely, it is possible that you may recognise people in the images included in the study, which may induce an adverse reaction. The researcher will monitor you for any adverse reaction and you will be free to withdraw at any time with no consequences and will have the opportunity to receive counselling support from the Employee Assistance Program at DST. Any occupational health and safety issues will be identified on site and appropriate measures will be taken to control risks to participants.

Statement of Privacy. All data collected during the experiment will be treated in the strictest confidence and stored on password protected computers at DST Group and the University of Adelaide. Retained data will not include any personal or identifying information - only response time and accuracy rates, along with generic demographic data will be stored. The data will be used for this project and potentially to compare to follow-on work conducted by DST Group and will be destroyed when no longer required. You will also have the opportunity to receive a summary of the research findings. If you chose to receive your results you will be required to provide an email address. Results will be aggregated for reporting purposes to preserve anonymity.

Other relevant human research ethics considerations. In addition to receiving a copy of your own results, this research will be reported as part of an Honours thesis and potentially in the open literature or in a DST Group report in due course.

Consent. If you are willing to participate and have read and understood the above requirements and risks, your consent to participate is given by clicking on the first screen of the experimental application.

Contact Details:
Should you have any complaints or concerns about the project, please do not hesitate to contact the lead investigator in person.

Lead Investigator:

[redacted]

Alternatively, you may contact the DST Group Ethics Review Panel.

[redacted]

Yours sincerely,
Gemma Snyder (Student Researcher; The University of Adelaide), Dr Dana Michalski (Principal Investigator; Defence Science Technology Group) and Dr Veneta MacLeod (External Researcher; Defence Science Technology Group)

Appendix I – DST Guidelines for Volunteers

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 J – Verbal Briefing Delivered to Participants Prior to the Experiment

"Welcome everyone and thanks for coming along. My name is Gemma Snyder and I am conducting this experiment for my honours year thesis in the Bachelor of Psychological Science at Adelaide University. The principal investigator is Dr Dana Michalski from The National Security & Intelligence, Surveillance and Reconnaissance Division. This project is internally supervised Dr Carolyn Semmler from the University of Adelaide.

[WHS instructions]. "If there is an emergency while you are here in the building please follow the instructions of the wardens. The male toilets are down the corridor to your left and the female toilets are further down the corridor to your right. If you have any issues while completing the experiment please let me know".

"Did anyone not understand the information sheet or have any questions about it?"

"The study consists of three elements. First we will ask you a few demographic questions. Secondly we'll run through a practice trial and then you can complete the age experiment. And lastly there will be a prompt for you to enter your email address at the end of the experiment if you'd like a copy of your results – totally up to you"

"Are there any questions?"

"Please click NEXT to view a brief summary of the experiment. If you're happy to continue, please indicate your consent by clicking 'I consent to participating in this experiment'.

"Please enter your 3 digit unique identifier you received at the beginning of this experiment"

"Please click NEXT to complete the demographic questions and stop when you get to the instruction screen. If you have any questions please let me know".

"We'll start by reading through the instructions; these will make more sense when we work through an example".

“You will be presented with 140 facial images of people of varying ages, for which you are required to determine whether you think that person is a ‘child’ or ‘adult’, noting that a child is defined as a person under the age of 18 years old (ie. 0-17 years) and an adult is considered an individual ages 18 years and over. Once you have made this selection, you will not be able to change it. You will then be required to estimate how old you think the person is, followed by estimating an age range with which you are 80% confident the person’s age will fall. Once you are satisfied with your age estimation decisions for this face, you will then need to press the FINALISE DECISION button to move on to the next image. Decisions are final once you have pressed this button and you will not be able to go back to change it.”

“I would also like to add that the purpose of an 80% confidence as opposed to a 100% confidence is to avoid an extremely large range that forces overconfidence. This will allow us to assess individual confidence-accuracy calibrations, based on a correct response, 80 out of 100 times. Additionally, you will view predominately people of Middle Eastern appearance, as we are particularly interested in age estimation for refugee processing purposes.”

“Please work through the example on your own and after you have clicked NEXT, please wait for further instructions” [check that everyone has finished the examples]

“Any there any questions? Is everyone happy they know what they’re doing? You will now assess 140 images, in the same manner of the example you’ve just completed. You will be timed and your accuracy and confidence will be measured. Please work as fast and as accurately as you can”.

“Remember at the end of the experiment you can register for your results if you’d like them. Please let me know if you have any questions along the way and when you have completed the experiment. You may now begin, thanks and enjoy”.

Appendix K – Testing Normality using Shapiro-Wilk Tests

Testing Normality using Shapiro-Wilk Tests

Variable	*W*	*p*	Skewness	Kurtosis
Mean Absolute Error				
Child	.93	.006	1.22	3.00
Adult	.91	.002	1.14	1.17
Child Female	.93	.006	1.06	1.27
Child Male	.93	.008	1.13	3.76
Adult Female	.93	.007	1.10	1.56
Adult Male	.89	.001	1.15	1.07
Directional Estimation Error				
Overall	.97	.323	0.30	-0.44
Child	.97	.276	0.49	-.133
Adult	.97	.253	0.10	-0.72
Bias Score				
Overall	.97	.219	0.44	0.48
Child	.98	.528	0.28	-0.36
Adult	.98	.548	0.41	0.10
Precision Range				
Child	.88	<.001	1.20	0.84
Adult	.85	<.001	1.74	3.90

Note. df = 46 for all analyses. *SE* = 0.35 for all skewness output. *SE* = 0.69 for all Kurtosis output.