Factors affecting social perceptions of individuals: role of intrasexual competition and face perception.

Michael Jeanne Childs

A thesis submitted to Swansea University in fulfilment
for the requirement for the Degree of Doctor of Philosophy.

Department of Psychology
Swansea University

May 2024

Under the supervision of Dr Alex L Jones and Professor Jeremy Tree

Abstract

How does an individual's competitive trait and a target's facial characteristics affect social perception? This thesis looked at intrasexual competition and perception of social traits separately and contributed to different aspects of evolutionary psychology.

Previous studies showed that individuals that are more competitive would be more likely to use make-up and cosmetic procedures to alter their appearance, and that milder anti-ageing procedures and self-esteem motivations were more positively evaluated. Using surveys, I explored how women's skincare behaviours, attitudes towards aesthetic dermatology and personality traits predict females' competitiveness. I found that knowledge of aesthetic dermatology, motivations of increased self-esteem and one's anxiety towards appearing older were likely to predict female competitive trait. Additionally, using vignettes of targets with different motivations using different anti-ageing treatments, I found that females that were more competitive were more likely to evaluate targets more negatively, and that self-esteem motivations still received the most positive evaluations.

Studies on perception of facial attractiveness have identified gender, emotion, and motion biases. I examined whether these biases also apply to facial health and age using face perception tasks, and found that age estimation was similar across age groups, and facial health showed emotion and motion biases.

Lastly, previous studies have demonstrated robust other-ethnicity effects on facial recognition and I determined whether this could also be observed on facial attractiveness, health, and age by employing facial perception studies and eye-tracking techniques. I found no evidence of OEE on perceptions of facial attractiveness, health and age.

This thesis advances knowledge in two ways: first, I provided insight on which individuals would be more likely to engage in anti-ageing treatments and how targets would

be viewed by observers, and second, I demonstrated that different facial factors influence perception of social traits such as facial attractiveness, health, and age, despite the close relationship between these variables.

Declarations and Statements

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed [redacted] (candidate)

Date: 28 May 2024

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed [redacted] (candidate)

Date: 28 May 2024

STATEMENT 2

I give consent (a) for the thesis, if accepted for a research degree, to be made available for electronic sharing and for inter-library loan (subject to the law of copyright) and (b) for the title and summary to be made available to outside organisations.

Signed [redacted] (candidate)

Date: 28 May 2024

Table of Contents

I dedicate this work

to my husband, who tirelessly supported my academic endeavour; and

to my children, Bernard, Gwenllian and little Tomas who are my greatest gifts.

Acknowledgement

First, this work would not be possible if not for the perfect timing that God has provided for me and us as a family. This studentship has come at the right time, and although there were plenty of challenges along the way, we conquered them by the strength He gave us. I am forever grateful that when doors closed, You have opened others which provided more.

To Dr Jones and Prof Tree, I am incredibly lucky to have you two as my supervisors – I always thought that you have complemented each other in your expertise and have supported me in different ways, which have helped me as a researcher and an academic. Thank you for the knowledge and skills you have imparted and the opportunities you have provided for me during my studentship. I hope to collaborate with you in future projects.

To Dr Johnston, for his support in the eye-tracking experiment. To Undergraduate students who helped with data collection – Bethan Hughes, Danielle Jones, Emily Cooper, Rozalie Matejkova, Ondrej Andrew Burysek, Alex Green, Jasper Lok, Alex Woodfield, and Joseph Hutchinson.

To Timothy, for providing the emotional and mental support that I needed – you have always believed in me even when I fall apart and lose hope. Your confidence in me has given me courage to push harder and explore this career far more than I would have on my own. Thank you for reminding me of what I could achieve – and that I could be more if I wanted to. Thank you for supporting my fleeting aspiration to be a stay-at-home mom, although we both know that as much as I would love to, my restless mind would not let me. Thank you for celebrating my milestones with me – you above all else understood how important they were, even those small ones that I sometimes forget to celebrate. Thank you for being my personal cheerleader.

To my mom, thank you for your sacrifices for your family and me. If it were not for you, I would not have had the opportunity to come and live in the UK. This may not have been the path you have envisioned for me, but you paved the road that led me here.

To Claire, the most supportive mother-in-law anyone could ask for, thank you.

Lastly, I want to extend my gratitude to my community – parents and toddler's groups, mom buddies, colleagues, and friends. Thank you for checking in and providing support when I was on my own and Tim was training in the army – I do not know how I managed that difficult time, but I know that I felt supported and incredibly empowered because of you. It does take a village to raise children and I am grateful to have such a strong village behind me.

Preface

This studentship was funded by a global company and Swansea University with the focus of understanding factors that would contribute to how facial attractiveness, health and age are perceived across different age groups, gender, and ethnicities and how individuals who choose to undergo facial enhancement are perceived by other people.

The work presented in this dissertation is a culmination of three years' work where halfway through, a global pandemic has changed the direction of the main PhD project. The original project involved running a clinical trial where I test the effectiveness of a cosmetic device targeting signs of ageing (wrinkles, hyperpigmentation, and uneven skin tone, among others) across participants from different ethnic groups. Additionally, I was going to have an experiment where I show pre- and post- treatment images to participants of a wide age range and different ethnic groups, to investigate whether there are differences in the judgements of facial attractiveness, health, and age before and after the treatment as well as a function of age and ethnicity.

However, restrictions were placed nationwide in the UK due to global pandemic, which postponed the pilot phase of the clinical trial. As restrictions were prolonged, this had an impact on the time-scale I had with the studentship and I had to resort to changing my study design. This involved moving experiments online and utilising facial stimuli that were already validated and completely changing the direction of the project.

The results presented in this dissertation therefore shows the coming together of several ideas under evolutionary psychology and facial perception. I hope you, the reader, find the content enjoyable and interesting.

Michael Jeanne Childs, UK

Table List

Figure List

Definitions and Abbreviations

Aesthetic Dermatology - a range of cosmetic treatments from non-invasive treatments to more invasive procedures

Anxiety towards Aging – being anxious about appearing older

Areas of Interest (AOI) – defined facial areas for eye-tracking study

Averageness - how close a face is to facial prototypes or the average facial composition of the male or female population

Beautification strategies – behaviour used to alter one's appearance to look beautiful, e.g., make-up and cosmetic procedures

Cambridge Face Memory Test (CFMT) – standardised facial recognition test

Competitor derogation - where females aim to weaken others' value and thus increase their own

Contact hypothesis – the more exposure one has to the other-ethnicity group, the better they are at recognising them

Fluctuating asymmetry (FA) - the variation between left and right regions of traits that are usually symmetrical in a population level

Internal reliability (IR) – assesses the consistency of results across items within a test

Intrasexual competition (IC)– members of the same sex competing to attract a potential partner

Intrasexual Competition Scale (ICS) – questionnaire measuring an individual's competitiveness towards people of the same-sex.

Knowledge and Attitude toward Aesthetic Dermatology (KAAD) – subscale of the Aesthethic Dermatology and Emotional Well-Being scale measuring an individual's knowledge of aesthetic dermatology.

Operational Sex Ratio (OSR) – the ratio of competing males/females that are ready to mate to sexually competing males/females that are ready to mate at a given time.

Other ethnicity blindness – individuals having poor performance on recognising other ethnicity faces despite having good own-ethnicity performance

Other-ethnicity effect (OEE) – phenomenon where individuals are better at recognising unfamiliar faces from their own ethnic group compared to those from other ethnic group

Perceived Results from the Aesthetic Dermatology (PRAD) – subscale of the Aesthethic Dermatology and Emotional Well-Being scale measuring perceived results from procedures

Relative Fixation Counts (RFC) – proportion of overall fixations received by a facial area

Self-promotion - where one enhances their desirable characteristics to the opposite sex

Sexual dimorphism – masculine and feminine traits

Sociometer Theory – theory of self-esteem from an evolutionary psychological perspective which proposes that self-esteem is a gauge of interpersonal relationships

Chapter 1.

General Introduction

"Anyone who keeps the ability to see beauty never grows old."
Franz Kafka

The face has an important role in our daily interactions. Within a few seconds, humans perceive demographic information such as age, sex, and ethnicity, and judge their personality traits such as trustworthiness and competence. We also make judgment on whether the face is attractive - and this perception of attractiveness has social consequences.

1.1 Attractiveness and social consequences

As social beings, the face provides us with vital information which influences our day-to-day interactions. Our brains have been wired to detect faces from birth with specialized brain regions that are sensitive to faces, e.g., fusiform face area (Simion & Giorgio, 2015). From a young age, we can identify faces, which enhances our survival and aids in attachment (Barrera & Maurer, 1981). As we age, we get better at 'reading' other people's faces and we can make judgements on individuals' personality traits and attractiveness in a short period of time, as little as 35ms (Verosky et al., 2018) and also for 100-500ms (Willis & Todorov, 2006). We use these judgments to choose our friends, partners, and even who to vote for in elections (Langlois et al., 2000; Todorov et al., 2005; Willis & Todorov, 2006). The face undergoes changes as we age, however, and these changes also influence how people perceive other individuals.

Evolutionary theory posits that attractive faces evolved as a result for preference for mates that are healthy and fertile (Fink & Penton-Voak , 2002; Thornhill & Gangestad, 1999) suggesting that the face carries cues which signal one's health and fertility status. In particular, it has been suggested that women are more inclined to maintain their appearance as a function of attracting a mate, as men tend to prioritise traits such as youth and beauty while women prioritise traits such as ability to protect and provide in their potential mates

(Buss, 1988; Schmitt & Buss, 1996; Thornhill & Gangestad, 1999). This divergence in traits preferred by men and women is said to drive the consumption of treatments, which aim to maintain the appearance of youthfulness in women.

However, the social consequences of attractiveness go beyond finding a partner to reproduce with. Studies have shown that attractive people have better social outcomes, e.g., e.g. better job opportunities (Mobius & Rosenblat, 2006; Rooth, 2009), higher incomes (Judge et al., 2009), bigger circle of friends, more sexual and romantic partners, more lenient sentences, even better terms on financial loans (Ravina, 2008). More attractive people are also perceived to be more competent, more intelligent, and more trustworthy (Langlois et al., 2000). Perception of attractiveness has an influence on our daily lives in general that it is not surprising that a lot of research has been done to understand the underlying mechanisms of attractiveness.

1.1.1 What makes a face attractive?

There are a number of traits that have been shown to contribute to perception of facial attractiveness. Among these, facial symmetry, averageness, and sexual dimorphism (Fink & Penton-Voak , 2002; Jones & Jaeger, 2019; Little et al., 2011; Rhodes, 2006; Thornhill & Gangestad, 1999) have been extensively studied and have been demonstrated to be strongly linked to attractiveness.

Fluctuating asymmetry (FA), the variation between left and right regions of traits that are usually symmetrical in a population level, tend to increase during development (Perrett et al., 1999). Exposure to diseases, varying levels of nutrition, and general genetic make-up could lead to increase in FA, in turn decreases the symmetry in the body. Studies which involved manipulation of facial symmetry of male and female faces suggested that images with higher symmetry were rated to be more attractive (Fink, Neave, et al., 2006; Perrett et al., 1999; Tybur et al., 2022). Similarly, studies which used naturalistic images, i.e., no manipulations, also showed higher ratings of attractiveness for faces that were perceived to be symmetrical (Baudouin & Tiberghien, 2004; Foo et al., 2017).

Averageness, how close a face is to facial prototypes or the average facial composition of the male or female population (Trujillo et al., 2014), is another factor which contributes to perception of attractiveness. Studies which used manipulated images showed large effects of averageness on judgments of attractiveness (Jones & Jaeger, 2019), and similarly, research using naturalistic images also showed a preference for faces with higher perceived averages for male faces (Baudouin & Tiberghien, 2004; Foos et al., 2006; Jones & Jaeger, 2019). However, it is important to note that there are non-average facial characteristics that are perceived as attractive (DeBruine et al., 2007) and that attractive faces tend not to be 'average'.

Sexual dimorphism, perceptions of facial femininity and masculinity, is another factor which contribute to attractiveness. This means that male faces with features associated with typical masculinity, and female faces with more feminine features were likely to be perceived as more attractive (Rhodes, 2006). This is said to reflect male and female traits that are related to better reproduction, where more masculine faces reflect a handicap signal (Stoehr & Kokko, 2006), i.e., they have excess testosterone to have the masculine features, despite high levels of testosterone suppressing their immune system, and more feminine features signal fertility as a result of higher levels of oestrogen (Law Smith et al., 2006).

Other factors have also been demonstrated to influence facial attractiveness. One is the sex of the target, where there seems to be a gender bias in attractiveness perception. Studies have found that female faces tend to be perceived as generally more attractive than male faces (Korthase & Trenholme, 1982; McLellan & McKelvie, 1993; Morrison et al., 2013). Furthermore, an emotion bias has also been highlighted in the literature, where faces with more positive affect were perceived to be more attractive compared to those depicting negative emotions, e.g. anger and sadness (Bowdring et al., 2021; Golle et al., 2014; Morrison et al., 2013; Penton-Voak & Chang, 2008; Rubenstein, 2005; Ueda et al., 2016). Even neutral faces perceived to be depicting positive emotions were rated more attractive compared to those perceived as having a negative emotion (Jones et al., 2018).

Finally, experimental studies have also looked at the role of motion on perception of attractiveness. Studies which used between-subjects designs, that is, a different group of participants saw the static condition and another set saw the dynamic condition, found no correlation on attractiveness ratings for static and dynamic images (Rubenstein, 2005). On the other hand, those which used within-subjects design, that is, participants saw both conditions found a high correlation between the two (Roberts et al., 2009). Interestingly, studies which investigated the role of motion, emotion, and target sex found significant interactions between the three, where for female faces, those with positive emotions generally were rated as more attractive regardless of motion, whereas male faces were rated more attractive when they were moving, regardless of emotion they portrayed (Penton-Voak & Chang, 2008). From here, the role of motion on facial attractiveness is unclear due to inconsistencies in methodologies used by previous studies.

Perceptions of attractiveness are also related to other facial traits. As previously mentioned, attractiveness signals traits relevant to reproductive success. It therefore follows that these signals could also be interpreted as a cue for one's actual healthiness. Traits which signal attractiveness also signal healthiness, e.g. faces with higher symmetry and averageness show that an individual is able to cope with environmental stresses and diseases (Fink & Penton-Voak , 2002; Thornill & Gangestad, 1993), and sexual dimorphism highlight levels of testosterone and oestrogen in one's body which are both markers of health (Thornhill & Gangestad, 2006). Experimental studies show that these signals are also being picked up by observers, that is, when individuals perceive a face to be attractive, they also perceive the face to be healthy (Fink et al., 2012; Fink, Neave, et al., 2006; Matts et al., 2007).

Another trait associated with facial attractiveness and health is facial age. This association, however, shows an opposite trend, that is, the older an individual looks, the less attractive and less healthy they are perceived to be (Fink, Grammer, et al., 2006; Fink et al., 2012; Kwart et al., 2012; Matts et al., 2007; McLellan & McKelvie, 1993). This has physiological grounds – as individuals get older, they are more susceptible to having illnesses and therefore less healthy. Appearance of signs of aging such as wrinkles, sagging of facial skin, and uneven skin tone also contribute to evaluation of older age (Flament et al., 2015; Samson et al., 2010). I will discuss more details on facial aging in the next section.

A general limitation with the studies which investigated the role of motion, emotion, and target sex on perception of attractiveness is that such studies tend to use facial stimuli from a young sample and young participant pool, and therefore the effects of facial affect and motion are not yet understood for older faces. As these three traits – attractiveness, health, and age – are related to each other, I therefore aimed to untangle the contribution of these factors in Chapter 3: 'To smile or not to smile: Differential effects of gender, emotion, and motion in perception of attractiveness and health across the age range' using a wider age range of sample and participants using experimental studies to add to our knowledge of how older faces are perceived, as well as to establish how perceptions of these three socially salient traits differ as a function of facial motion, emotion and target sex.

1.2 Intrasexual competition and beautification strategies

Men and women tend to prioritise certain traits when looking for a potential partner. While men tend to favour youthfulness for successful reproduction, women tend to search for partners that have good genes for their offspring and those able to provide and protect (Thornhill & Gangestad, 1999). However, not all men or women will possess the traits that are attractive to the other sex. Intrasexual competition therefore occurs when there is a limited resource, that is, a potential partner and individuals are contending to be an attractive mate (Buunk & Fisher, 2009; Fisher & Cox, 2011). In this thesis, I focused on female intrasexual competition and how it is manifested in consumption of beautification techniques. One of the strategies commonly used by females is self-promotion, where they enhance their appearance or reputation in order to secure a mate (Buunk & Fisher, 2009).

Enhancement of facial appearance is not new. From use of make-up, accessories, piercings and tattoos of ancient civilisations, to more recent uses of (more invasive) surgical procedures, people tend to strive to adhere to a contemporary ideal of beauty. For example, Ancient Egyptians used kohl both for adornment and protection against the glare of the sun and provide anti-bacterial benefits (Draelos, 2015). In the 11th century, lessons from Trotula de Ruggiro included use of make-up for youthful appearance as well as treatment of skin diseases (Cavallo et al., 2008). Fast forward to the advent of facial reconstructiowhere between 1887 to 1898 more formal publications talked about rhinoplasty started to occur, and as a devastating consequence of World War I, soldiers suffered facial injuries and deformities which needed surgical treatments (Dolsky, 1999). Now, we have an array of treatments, ranging from less invasive use of dermal fillers to define a facial feature, to a more invasive facelift (The Aesthetic Society, 2021), all aiming to provide either a temporary or long-term changes in facial appearance.

Studies have shown that female’s frequency of use and amount of spending on make-up and use of cosmetic procedures are associated with female intrasexual competition trait (Wagstaff, 2018). Make-up is used to temporarily alter facial colour, e.g. increase facial contrast (Jones et al., 2015; Porcheron et al., 2013), mimic red lips (Stephen & McKeegan, 2010), red cheeks and brighter under-eye areas (Jones et al., 2016), and change the apparent size of facial features (Jones et al., 2018). These alterations in appearance are associated with increased perception of attractiveness and health, and decreased age (Jones et al., 2018; Jones et al., 2016; Russel et al., 2016; Stephen & McKeegan, 2010). Make-up also influences the perceptions of the target, making them more attractive to men and perceived as more dominant by other women (Mafra et al., 2020; Mileva et al., 2016; Sulikowski et al., 2022). Use of anti-ageing treatments could lead to more long-term effects, e.g., use of botox and face-lift to significantly reduce signs of aging. Target individuals are rated as more attractive, healthier, and younger (Nellis et al., 2017).

As women age, levels of oestrogen declines, reducing collagen and elastin levels and in turn results to decline in skin elasticity and overall dermal health (Lephart, 2018). This leads to women appearing older, less healthy, and less attractive. Aging is inevitable, and individuals roughly undergo the same process of how our skin ages (Michaud et al., 2015), however, the rate on how these changes occur differ significantly due to both intrinsic, i.e., genetics and hormonal changes, and extrinsic factors, i.e., lifestyle choices such as dieting and exercise, exposure to sunlight and illnesses. Although internal aging is not visible to us, aging manifests itself on the skin - the largest organ of the body. The highest consumers of anti-ageing procedures such as botox, dermal fillers, and facelifts in the US in 2020 were middle-aged women (30-55 years old, American Society of Plastic Surgeons, 2020). As advances in aesthetic dermatology continues, anti-ageing procedures have now been more accessible and affordable to many.

However, make-up and cosmetic procedures are not the only beautification strategies available for women. The use of skin care products, e.g. cleansers, toners, and moisturisers, is another industry that has a focus on beautification. Albeit more of a long-term investment compared to make-up and cosmetic surgery, women’s consumption of skin care products are at par with these two strategies (Kumar et al., 2006; Lopaciuk & Loboda, 2013). In chapter 2.1, ‘Intrasexual competition and attitudes towards cosmetic (topical and surgery) use’, I therefore explored whether the amount of spending and frequency of use of skin care products will predict intrasexual competition trait as observed from other beautification strategies, alongside with other traits such as desire to undergo anti-ageing procedures, attitudes towards aesthetic dermatology, and personality traits.

Although appearing younger makes women even those past their ‘peak fertility age’ valuable to potential mates, this could cause animosity from same-sex individuals who are also vying for potential partners. There may be other reasons why women would want to appear younger and engage in anti-ageing treatments.

1.2.1 Motivations to use beautification strategies

With the various social and financial benefits of appearing younger, and thus looking more attractive and healthier, it is not surprising that there is an increasing rate of consumption of beautification strategies such as use of make-up and anti-ageing treatments. In particular, middle-aged women are observed to be the highest consumers of anti-ageing treatments in the US (American Society of Plastic Surgeons, 2020). The literature highlights three motivations as to why individuals would engage in beautification strategies.

Finding a partner. The most salient motivation could be explained by evolutionary theory, which highlights the premium of youth for women when looking for a partner. Men are more likely to value youth when looking for a mate for the goal of reproduction, as younger women are more fertile and therefore have better chances of having a healthy offspring (Maestripieri et al., 2014). Statistics show that mothers aged 40 are having the highest stillbirth rates in the UK in 2020 (Office for National Statistics, 2023), and pregnancies occurring when mothers are 35 and over having an increased risk of various health complications (Correa-de-Araujo & Yoon, 2021). The older a female is, therefore, the less ideal it is for her to reproduce.

Appearing older therefore affects a female's potential mating value, with men's evaluations of women's attractiveness steeply declining as targets get older (Henss, 1991) with those in peak fertility age (19-25) getting the highest attractiveness ratings (Sefcek et al., 2007). In England and Wales, the standard average age of mothers in 2020 was 30.7 years (Office for National Statistics, 2023), higher than the 'peak fertility age'. Women are more likely to pursue a career before marrying and starting a family, therefore reducing their potential mate value. As more women want to start a family later in life, and men still value youthfulness, they have to compete with younger females who are also vying for a potential mate. Therefore, using anti-ageing treatments could be a way of maintaining youthful appearance, and be able to compete with the younger females also searching for a mate.

Workplace. With our aging population, having older people in the workplace will be much more commonplace, with Pensions Act 2007 allowing a revision of the State Pension age at least once every 5 years, people born after 1978 may not retire until their 68th birthday (Department for Work and Pensions, 2023). However, studies show that negative stereotypes about older people in workplace exists, e.g., less motivated, more vulnerable to balance work-family commitments, unwilling to change and participate in training (Chiu et al., 2001; Ng & Feldman, 2012) and less competent (Cuddy et al., 2005). Additionally, people who underwent anti-ageing treatments have been evaluated as more competent and more hireable compared to those who did not (Tian et al., 2020). Therefore, it may be of benefit for individuals to appear younger than they are to reduce the negative stigma in the workplace.

Wellbeing in older age. Studies which looked into the psychological state of older people suggest that compared to their younger counterparts, older individuals have lower self-esteem, increased body dissatisfaction (Brooks, 2010) and have higher rates of mood disorders. Furthermore, qualitative interviews of older women have demonstrated that appearance of facial signs of aging contributes to low self-esteem as well as affecting their quality of life and body image satisfaction (Clarke et al., 2007; Hurd Clarke, 2002; Muise & Desmarais, 2010; Slevec & Tiggemann, 2010). Individuals who have undergone cosmetic treatments to tackle the signs of aging reported higher self-esteem and life satisfaction (Brooks, 2010; Muise & Desmairas, 2010; Slevec & Tiggeman, 2010).

Harris (1994) has found that when these motivations were considered, observers rated individuals who underwent anti-ageing treatments more positively when they did it for self-esteem, followed by employment and seeking for romantic partner. In addition, Chasteen et al. (2011) found that individuals who uses mild forms of anti-ageing treatments, e.g. uses creams, were evaluated more positively compared to those who engaged in more extreme, e.g. face-lift. These two studies show that evaluations of individuals who engage in age concealment vary depending on the motivation behind and type of concealment used.

As cosmetic procedures are increasingly becoming more accessible and affordable to people, I therefore investigated whether a shift in the perception of people who engage in anti-ageing treatments have changed, and whether the motivation behind engagements are perceived differently. Chapter 2.2 ‘Perceptions of individuals who engage in age concealment’ show that targets received more positive evaluations from female participants when the motivation is self-esteem, followed by employment and least for romantic purposes. In this chapter we demonstrate that the stigma around anti-ageing treatments have not changed significantly (Chasteen et al., 2011; Harris, 1994) despite the growing popularity and affordability of such treatments.

1.3 Other ethnicity effects

One of the most common phenomena in face perception is the other-ethnicity effect (OEE), where individuals are better at recognising unfamiliar faces from their own ethnic group compared to those from other ethnic group (Cenac et al., 2019; Childs et al., 2021; McKone et al., 2012; Wan et al., 2017). This effect has been widely researched and different theories as to why this occurs have been put forward. One of them is the contact hypothesis, where it was suggested that individuals have more contact and thus more exposure to faces from their own ethnic group, and therefore they are more familiar with them (Goldstein & Chance, 1985). This has been supported by studies which measured contact length with other ethnicity groups, and they found that the longer contact an individual has to other ethnicity group, the better they are at recognising faces from that particular group (Ng & Lindsay, 1994; Zhao et al., 2014). Furthermore, a number of studies have shown that the higher the contact an individual has of ethnic groups outside their own, the lower the magnitude of the OEE, that is, their performance on recognizing faces from the other ethnic group gets better (Tanaka et al., 2004; Zhao et al., 2014).

OEE has been demonstrated using perceptual and facial recognition tasks. An established face recognition task is the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006) which comes in different versions – Boston (Duchaine & Nakayama, 2006), Australian (McKone et al., 2011) , Chinese-Han (McKone et al., 2012) and most recently, Chinese-Malaysian (Kho et al., 2023). Having different, standardised version of facial recognition test allows for comparison of facial recognition abilities for both own- and other ethnicity faces, allowing intensive studies on OEE. However, there are variations in the abilities of individuals to recognise faces from their own-ethnic group to begin with, that is, there are people that are excellent (super-recognisers; Robertson et al., 2020) and there are those that are poor (developmental prosopagnosics; Duchaine & Nakayama, 2006). In chapter 4.1 'Do individual differences in face recognition ability moderate the other ethnicity effect?', we used three versions of CFMT – Boston, Australian, and Chinese-Han to test whether an individual's ability to recognise faces from their own ethnic group influences how well they recognise faces from other ethnic groups. Here we found that the magnitude of OEE is consistent across the own-ethnicity facial recognition ability. Interestingly, our study did not find a significant effect of contact in the facial recognition ability of other ethnicity faces. We concluded that this could be due to the lack of variation of contact in the sample.

OEE is not only limited in facial recognition. This phenomenon has also been observed in perception of attractiveness, where observers where rating faces that were closer to their ethnicity as more attractive (Darrach et al., 2019; Potter & Corneille, 2008; Rhodes et al., 2005). For age perception some effects were found, but in different ways. Caucasians were more accurate at estimating age of Caucasian faces than those of African faces (Dehon & Brédart, 2001) and faces with more Asian and Caucasian traits were estimated as younger compared to those with African traits (Lick & Johnson, 2018). For health perception, Rhodes et al. (2005) found that participants from Asian and Caucasian background rated Eurasian faces, i.e., mixed composites, as the healthiest.

In general, there is no agreement as to whether an OEE occurs in perception of attractiveness, health, and age. Some studies did show an OEE (Darrach et al., Potter & Cornielle, 2008; Rhodes et al., 2005) but there are also studies which demonstrate a high agreement on which faces look attractive (Coetzee et al., 2014; Kleisner et al., 2017; Rhodes, 2006; Rhodes et al., 2001) and healthy (Jones et al., 2018). Studies also show that different cultures may use different cues when judging for social traits, e.g. Caucasians tend to use face shape cues while African participants used facial colour cues when making judgments on attractiveness for African faces (Coetzee et al., 2014), and arriving at the similar ratings of attractiveness. There is also a scarcity of research on perception of age across different cultures. This is interesting as ethnic groups manifest signs of aging differently, with Caucasians having earlier onset of development of wrinkles and sagging compared to Asian and Black ethnicities (Rawlings, 2006). One explanation for this is that skin composition differs across ethnicities, e.g. higher melanin content in Asian and Black skin provides higher protection from the harmful UVA and UVB rays from the sun, which protects skin from accelerated ageing. I therefore wanted to contribute to the literature on OEE and perception of facial attractiveness, health, and age across different ethnic groups.

Interestingly, OEE on face recognition has been explored using eye-tracking studies. This technique allows the capturing of eye movements, e.g. where they looked (fixations) and how long they looked (duration) at facial areas. This provides an insight on the mechanisms that may underlie OEE, i.e., do we look differently at faces from our own compared to faces from other races and this may affect how we recognise or evaluate them? Arizpe et al. (2016) employed eye-tracking methods in their facial recognition tasks. They found a robust OEE effect, where Caucasian participants recognised Caucasian targets better than Asian and Black faces. Additionally, they found that participants looked at eye regions more when the target face was Caucasian, and looked at lips and nose regions more when they looked at Asian and Black faces. This is interesting as it implies the way we look at a face may be driving the poorer performance on recognition of other ethnicity faces. I was therefore interested in seeing whether this pattern could also be observed in other facial perception tasks. In chapter 4.2 'Looking at a face differently: Other-ethnicity effects and eye movements on evaluation of socially salient traits', I extended Arizpe et al.'s study (2016) by adding another observer group, as well as asking participants to evaluate the faces on facial attractiveness, health, and age, to explore whether the eye movement patterns could also be observed on tasks other than face recognition.

1.4 Thesis aims

Together, this thesis aims to give an insight on trends of how individuals who are likely to engage in anti-ageing techniques as a bid to maintain their youthful appearance are perceived, and how the socially salient traits that are associated with youthfulness that women are keen to preserve are perceived in different conditions.

First, I uncovered the relationship between an individual's psychological traits, e.g. self-esteem, satisfaction with appearance, and quality of life, with their skin care behaviours and desire to undergo treatments. Previous studies have shown that a female's frequency and intensity of make-up use is related to their intra-sexual competition, and I was interested in how this translates to skin care behaviours and attitude towards anti-ageing. Additionally, I explored how individuals who undergo anti-ageing procedures are contemporarily viewed by observers. As anti-ageing procedures started to be less invasive, having less recovery time, more accessible, and most importantly, more affordable to most, I wanted to investigate whether the stigma around people engaging with anti-ageing treatments has also changed.

Second, I explored which factors affect perception of facial health, attractiveness, and age. Various factors affect how a face is perceived, e.g. emotion, motion, and sex, to name a few. Previous studies which looked at perception of attractiveness used a young sample and young stimuli and I wanted to add on to the literature by incorporating a wider age range of stimuli, as well as investigating whether the trend holds true across the three social traits of interest.

Third, I explored how consistent OEE is across own ethnic group facial recognition abilities. Facial recognition studies have demonstrated a robust OEE, that is, observers are better at recognising faces from their own-group compared to faces from other-groups. I explored whether this could be observed when people make people make judgements on facial attractiveness, health, and age. Using eye-tracking techniques, I aimed to expand on the literature by investigating which facial areas Asian and Caucasian participants look at when they view faces from Asian, African, and Caucasian target faces when making judgments on facial attractiveness, health, and age.

Chapter 2.

Perception of facial ageing and anti-ageing procedures

Chapter 2.1 Skincare behaviour and intrasexual competition

Abstract

Use of make-up and cosmetic surgery is related to female intrasexual competition (IC). However, no study has yet investigated the consumption of skincare products as a beautification strategy together with motivations, desires, and attitudes towards appearing older in relation to IC. To address this, 254 women completed a survey looking into amount and frequency of skincare spending, evaluations of signs of aging and anti-aging treatments, and personality questionnaires. When all factors were considered together, we did not find that spending and use of skincare products were related to IC. An individual's anxiety towards appearing older, having positive perceptions about the results of cosmetic treatments, and desire to have a facelift were positively related to IC. Self-esteem and knowledge about cosmetic treatments were negatively related to IC. Our findings provide a novel insight which showed that unlike more salient beautification strategies such as make-up and cosmetic treatments, the use of topical skin care products do not predict IC trait in women. However, one's knowledge of and perceived results from cosmetic treatments as well as a desire to undergo an appearance enhancing treatment predicted IC traits, which is in line with studies emphasising the use of self-promotion as a strategy for competition.

Keywords

Intrasexual competition, skincare, anti-ageing treatments, cosmetics

Sexual dimorphism in humans have led to different traits becoming more desirable for one sex than the other, for example, males tend to value attractiveness and youth for females, while females look for ability to provide resource in males (Buss, 1989; Schmitt & Buss, 1996). This results in competition between members of the same sex for access to mates, known as intrasexual competition (IC) – in order to gain access to and vie for the opposite sex (Campbell, 2004; Fisher, 2004). While males compete in more aggressive forms, females, who are less likely to take risks, would compete in other means such as relational aggression, i.e., social exclusion and harming other's reputation by gossiping (Archer & Coyne, 2005) as well as enhancing their appearance relative to other females who are also vying for other men (Campbell, 2004).

It is proposed that sexual competitiveness arise when resources are limited and individuals look for certain traits that are also desirable to others. Two strategies that females use to compete were identified as *self-promotion*, where one enhances their desirable characteristics to the opposite sex; and *competitor derogation*, where females aim to weaken others' value and thus increase their own (Buss, 1989; Buss & Schmitt, 1993; Schmitt & Buss, 1996). Studies which used vignettes of individuals who used anti-ageing treatments provide an insight to these two strategies. We could argue that the target individuals were engaging in self-promotion strategy as they were aiming to enhance their appearance for various reasons. Interestingly, it was found that target women who used anti-ageing treatments for self-esteem purposes received the most positive ratings, compared to those who used it to have a better chance at finding a romantic partner (Childs & Jones, 2022; Harris, 1994). Furthermore, those who used more invasive procedures, e.g. Botox and facelifts, also received more negative evaluations. This could therefore be seen as a type of derogatory tactic, that is, the target was evaluated negatively and thus created a stigma towards individuals who actively engaged in self-promotion strategies.

Fisher and Cox (2011) found two more strategies that females use which involves mate guarding tactics – *mate manipulation*, that is, obscuring the acts of competitors hence reducing competition; and *competitor manipulation*, that is, attempting to make rivals less desirable, thus reducing the chances of having competition. Despite the two new strategies identified by Fisher and Cox (2011) from their qualitative study, their follow-up study using the IC survey has shown that the most common tactic used by both males and females is *self-promotion*. It was argued that this could be because the other tactics could lead to the individual appearing mean-spirited, a trait that is less desirable. Additionally, they also found in this study that females were more likely to use self-promotion tactics relating to their appearance, body, and athleticism than males, in line with the idea that youth and attractiveness are the traits that men look for when searching for mates (Buss, 1989).

As enhancing one's appearance is a strategy commonly used by competitors, we could infer that individuals who desire to undergo beautification strategies, e.g. use of make-up and undergoing cosmetic procedures will have higher IC trait. Additionally, we predict that individuals who are motivated to engage in appearance enhancement for romantic purposes would be more competitive compared to those who would engage for self-esteem.

Female IC and beautification strategies

Make-up

One of the most common ways to enhance one's appearance is the use of facial cosmetics, and studies have found that female faces with applied cosmetics were perceived to be more attractive than faces without (Jones & Kramer, 2016; Mileva et al., 2016; Sulikowski et al., 2022). This may be due make-up's ability to increase facial contrast, widen the eyes, even out skin tone, and increase redness on the cheeks and lips (Jones et al., 2015). Use of facial cosmetics is also a relatively low investment strategy, e.g. it is less invasive and the results are relatively instantaneous (Jones et al., 2015; Mileva et al., 2016), compared to using more invasive and treatments with longer recovery times such as facial surgeries. It has also been demonstrated that facial cosmetics use have two functions: to *attract* the opposite sex (Mafra et al., 2020; Wagstaff, 2018) and to *compete* with the members of the same sex (Mileva et al., 2016; Sulikowski et al., 2022).

A study by Wagstaff (2018) investigated whether competitive trait and competitive tactics used by women can predict how women use make-up. Using questionnaires, participants were asked about their make-up utilisation, e.g. frequency (ranging from every day and less than once a month) and quantity used (how many make-up items they use). Participants also completed personality questionnaires pertaining to social comparison attitude, IC and competitive tactics used, narcissism, attitude to sexual roles, and socio-sexual orientation. Using binary logistic regression, they found that the more females were interested in long-term mating commitments and were more willing to use competitive tactics against other females, the more likely they were to wear make-up frequently. On the other hand, *younger* females who were more interested in short-term mating commitments and higher narcissistic trait were more likely to use a higher *quantity* of make-up. From here, it was argued that the frequency and quantity of facial cosmetics used reflect two different functions, where *frequency* of make-up relates more to long-term mating strategy, and *quantity* of make-up were for short-term mating strategies. Interestingly, it also found that women in relationships also wore make-up less frequently than single females, further highlighting the function of facial cosmetic use in coveting a partner. However, although this study has highlighted that women who were more willing to use competitive tactics would engage more in use of cosmetics, IC itself was not a significant predictor of make-up use. A more recent study however, showed that IC traits predicted hypothetical purchase and use of make-up and social media consumption.

Moreover, recent studies demonstrated that IC traits predicted hypothetical purchase and use of make-up, as well as social media consumption (Wagstaff & Sulikowski, 2022), and that younger women who have higher competitiveness spent more money, used make-up more frequently, and spent more time on applying make-up compared to their counterparts (Mafra et al., 2020). These studies therefore suggest that female competitiveness and the willingness to engage in competition with members of the same-sex could be predicted by behaviour towards use of facial cosmetics. However, although use of make-up has been widely an accepted form of enhancing one's appearance, individuals who wear more make-up have been found to be less intelligent and more immoral by other women (Kellie et al., 2020). Therefore, it is of interest for this study to investigate whether other forms of non-visible appearance enhancing strategies, e.g., use of skincare products such as moisturisers and anti-aging creams, would also be indicative of IC amongst women as opposed to make-up and more invasive cosmetic procedures. To the best of our knowledge, no study has yet looked at this relationship.

Aesthetic Dermatology

Another strategy of beautification is the use of aesthetic dermatology, which is a range of cosmetic treatments from non-invasive treatments such as microdermabrasion to more invasive procedures, e.g. Botox and facelifts (Juhasz et al., 2017). These aim to enhance or alter one's appearance – most of which are used as anti-ageing treatments. With the advancement of dermatological technologies, procedures have now been more accessible to consumers (more affordable and less recovery time) and has been growing in popularity, with the most consumers being middle-aged women (aged 35-50, The Aesthetic Society, 2021). The top six non-invasive, professional treatments in 2020 included neurotoxins (e.g. botox), dermal fillers, and abrasive skin treatments (chemical peels), non-abrasive skin treatments (lasers), hair removal and skin tightening (The Aesthetic Society, 2021).

The boom in aesthetic dermatological procedures have allowed individuals to alter their appearance in a more permanent manner, even 'reversing' the effects of time, and thus enabling those who engage in it to achieve their ideal appearance (Dubbs et al., 2015). Additionally, studies have found that post-operative images of those who have engaged in cosmetic procedures have been perceived as better potential partners (Kalick, 1979), were as attractive as their younger counterparts as well as more hireable (Tian et al., 2020) compared to their pre-operative images. Such evidence demonstrates the benefits of cosmetic procedures, and therefore highlights the underlying reasons for engaging in them. Despite the growing prevalence, women still comprise the majority of cosmetic procedures (The Aesthetic Society, 2021), gave higher importance to skin appearance, and were more knowledgeable about and more positive in evaluating the results of aesthetic dermatological procedures (Martínez-González et al., 2018) compared to men. This could be attributed to the premium placed by men on attractiveness and youth on female partners (Buss, 1988), e.g. women who would want to appear 'viable' to partners may need to know more about procedures which maintain their youth and believe that such procedures would benefit them, while this aspect is not as important for men when finding partners. It is therefore reasonable to assume that similar trends on competitiveness and engagement with make-up as discussed above would be observed with cosmetic procedures.

This has been supported by Wang et al. (2021) where they investigated perceived competition and relationship with attitudes towards cosmetic surgeries. In this part of their study, they looked at the operational sex ratio (proportion of local men of reproductive age to local women aged 15-54) of each state in the US and the amount of internet searches relating to cosmetic surgery. Here, they found that states with lower operational sex ratio (OSR) that is, high competition (more women than men in their area), had significantly higher search volumes compared to those with higher OSRs (low competition, more men than women in their area). Additionally, they also looked at the number of certified plastic surgeons in each state, and as expected, those with lower OSRs had a higher density of practitioners compared to those with higher OSRs. This suggests that, at least in the US, the perceived level of competition (as indicated by the proportion of men to women) is related to the prevalence of practices which endorse appearance enhancement.

Furthermore, a study by Arnocky and Piché (2014) found that an individual's IC trait predicted their attitude towards cosmetic surgery, that is, for both males and females, the more competitive they are, the more positively they regard cosmetic surgery, as well as the more they were willing to spend on cosmetic procedures. Similarly, Bradshaw et al. (2019) found that women who were interested in short-term mating were more likely to be accepting of cosmetic procedures, but not facial cosmetics. The same study also found that both male and female participants perceived female targets who have undergone cosmetic surgery on either face or body as more willing to engage in short-term relationships compared to those who had no cosmetic surgeries. This implies that perhaps attitudes towards cosmetic procedures are more relevant to short-term mating than long-term relationships. Nonetheless, these studies imply that attitudes towards cosmetic procedures were related to IC trait.

Skincare products

Another avenue that is commonly used, but has not received a similar level of attention, is the use of topical skin care products, e.g. cleansers, moisturisers, and serums, which help improve the appearance of facial skin (Mangal et al., 2021; Messaraa et al., 2020). The market for skin care products has been valued at 130.50 billion USD (Grand Review Research, n.d.), with women comprising the majority of skin care users. However, the relationship between using skincare products as a beautification strategy and female competitive traits has not been explored as well as the other strategies. The current study therefore explored whether use of skincare products could predict IC trait, in line with results from previous studies which used other types of appearance enhancements, as well as other traits.

Female IC and Anxiety towards Aging

As mentioned previously, women are more likely to engage in appearance enhancement strategies than men. In particular, the biggest consumer of aesthetic dermatology are women between 35-50 years old (The Aesthetic Society, 2021). One of the reasons cited for this is that female reproductive ability diminishes as women age (Buss, 1989; Maestripieri et al., 2014), and therefore this is argued to be a principal reason as to why women's attractiveness value diminishes with age (Muise & Desmairas, 2010; Slevec & Tiggeman, 2010). Additionally, the reduction of oestrogen in female bodies results to signs of aging becoming more salient, e.g. drier skin, appearance of wrinkles, thinner skin appearance and decreased elasticity (Lephart & Naftolin, 2021; Thornton, 2013). Previous studies using surveys have found that women who were more anxious about appearing older were more likely to have positive attitudes towards engaging in cosmetic procedures, that is, they are more likely to undergo treatment and believe that engaging in the treatment will help their wellbeing (Furnham & Levitas, 2012; Muise & Desmarais, 2010; Slevec & Tiggemann, 2010).

With a positive relationship between anxiety towards aging and attitude towards cosmetic procedures, as well as the relationship between IC and attitude towards cosmetic procedures, we were interested to see whether a relationship between IC and anxiety towards aging also exists, that is, would women who are more anxious about the physical aspect of aging be more competitive towards other women?

Aims and Predictions

Overall, previous studies which looked at the use of make-up, a type of beautification strategy, have shown that female IC trait predicted frequency and quantity of make-up use, perceived levels of competition, and attitudes towards cosmetic surgery. The current study therefore aimed to contribute to the literature by exploring another aspect of beautification – the use of skin care products. To the best of our knowledge, no study has yet explored the relationship between skin care products (as opposed to make-up) and IC.

As previous studies have shown that types of beautification which have visible effects, e.g., application of make-up and undergoing cosmetic procedures, were strong predictors of IC in females, we hypothesise that behaviours and relating to spending and use of skincare products would be predictors of IC.

Previous studies have shown that types of motivation and desired procedures (Childs & Jones, 2022; Harris, 1994), and acceptance of signs of aging (Harris, 1994) were factors that affect how individuals who engage in anti-ageing treatments are perceived by others. We are therefore interested in exploring whether these factors are predictive of IC, as those individuals who engage in anti-ageing procedures may be viewed as competition.

As per the evolutionary perspective, younger females were more likely to be sought after by males than older females (Buss, 1989), we therefore hypothesise that individuals who were more anxious about appear older will have higher IC. In line with this, we also hypothesise that individuals who have a more positive outlook on aesthetic dermatological treatments will have a higher IC trait as this could help them maintain their youthful appearance.

Methods

Participants

254 female participants with sample age range of 18-71 ($M = 29.80$, $SD = 11.97$) were used in this study. Data was collected between 08 April 2020 and 26 June 2020 on gorilla.sc (Anwyl-Irvine et al., 2020). Participants were recruited through social media platforms (Facebook and Twitter) and student recruitment platform. Inclusion criteria were female, aged 18 and over and had access to internet to complete the survey. There were no restrictions regarding country of residence. All completed surveys were anonymised, with age the only demographic variable used in the study. Participants who provided their email addresses separately were included in a raffle draw and two were awarded £25 Amazon vouchers for their participation. Participants from the student recruitment service were also awarded two credits for their time.

Statement of Ethics

Written consent forms were acquired before participants were presented the study, which includes a statement which asked their consent for their responses to be published as part of the study. Participants accessed the study using an anonymous link and were able to withdraw by not completing the study at any time. Participants were informed that only completed tasks were included in our analysis and in turn, publication. All participants were given an option to be included in a raffle draw as compensation for their time, and a study credit (2) was awarded when the study was accessed through student recruitment platform as part of their course requirement. This study was approved by the Swansea University Ethics Committee and followed the Declaration of Helsinki (Williams, 2015).

Materials

All materials used in this study is available in an online repository (https://github.com/JeanneChilds/Chapter-2.1-Survey).

Skincare Products and Routine

Participants were asked how much they spend on skincare products each month (< £20, £21-50, £51-80, £81-100, and > £100), and how frequent they buy the products (never, every 6 months, every 3 months, once a month, and once a week).

Motivations for Appearing Younger (Harris, 1994)

This section consisted of five reasons why people engage in age concealment and participants were asked to judge the acceptability of appearing younger in each area. Items include "self-esteem" and "vanity". Items were rated on a 7-point Likert scale, where 1 = not at all acceptable to 5 = extremely acceptable (see Appendix A). The scale showed an acceptable internal reliability in this study ($\alpha = .77$)

Signs of Aging (adapted from Harris, 1994)

This questionnaire consisted of the six signs of aging used in Harris (1994) study, and we also added two items: facial sagging and uneven skin tone. Items include "white hair" and "wrinkled neck". Each item was rated on a 7-point Likert scale, where 1 = very unattractive to 7 = very attractive (see Appendix B). Higher scores indicate higher attractiveness ratings. The scale showed a good internal reliability in this study ($\alpha = .88$)

Desired Procedures (adapted from Harris, 1994)

This questionnaire consisted of 8 items pertaining to cosmetic procedures individuals would be willing to engage with if money was no object. Items include "Use a wig to cover thinning or balding hair." and "Use face-lift or other cosmetic surgeries to look younger." We also added new procedures such as chemical peels and microdermabrasion, Botox and other injectables, and hand-held products at home (e.g. radiofrequency and light therapy). Each item was rated on a 4-point Likert scale, where 0 = would never do, to 3 = definitely would do (see Appendix C). Higher scores indicate more willingness to engage in age concealment procedures. The scale showed a good internal reliability in this study (α = .83)

Physical Appearance Subscale of Anxiety about Aging Scale (Lasher & Faulkander, 1993)

This 5-item subscale pertains to physical changes brought about by aging. Example items are: "I have never lied about my age in order to appear younger." and "I have never dreaded looking old." Each item was rated on a 5-point Likert scale where 1 = definitely agree to 5 = definitely disagree, where the higher score indicates higher anxiety towards aging (see Appendix D). The scale showed a poor internal reliability in this study (α = .61)

Perceived Results from the Aesthetic Dermatology (PRAD) and Knowledge and Attitude toward Aesthetic Dermatology (KAAD) subscales from the Aesthetic Dermatology and Emotional Well-being Scale (Martínez-González et al., 2018)

The 13-item PRAD subscale pertains to the possible results individuals perceive to gain from engaging in aesthetic dermatology procedures (see Appendix E). Items include "The results of Aesthetic Dermatology can help me feel more able to overcome mistakes and weaknesses' and " … improve my relationship with my partner." The scale showed an excellent internal reliability in this study (α = .96)

The 6-item KAAD subscale pertains to how much detail the individual knows about certain Aesthetic Dermatology procedures. Items include "I know what the injectable wrinkle fillers consist on" and " I know what laser rejuvenation for the skin consists on." Each item from both subscales is on a 4-point Likert scale, where 1 = total disagreement to 4 = total agreement. Higher scores indicate more positive perceived results from aesthetic dermatology, and more knowledge about the procedures. The scale showed an acceptable internal reliability in this study (α = .79)

IC Scale* *(Buunk & Fisher, 2009)

A 12-item questionnaire which assess how competitive an individual is towards members of the same sex. Examples include: “I just don’t like very ambitious men/women.” and “I tend to look for negative characteristics in attractive men/women.” Each item is rated on a 7-point Likert scale (where 1 = not at all applicable to 7 = completely applicable), where the higher rating indicates higher competitiveness towards same-sex individuals (see Appendix F). The scale showed an excellent internal reliability in this study (α = .91).

Procedure

All participants accessed the questionnaires through an anonymous link hosted by Gorilla. Participants were presented with the Participant Information Sheet, and consent was acquired prior to completion of the questionnaires. First, participants provided their demographic information and completed the questionnaire in the order listed in the Material section.

Analytic strategy

To address our exploratory analysis, we first Z-score standardised all of the continuous predictors as well as the dependent variable (IC trait scores from the ICS questionnaire), and entered them into a Bayesian linear regression, predicting ICS scores from the other variables. As Bayesian inference requires the setting of a prior distribution, and given that we a multitude of correlated predictors, we opted to regularise the coefficients by placing a Normal prior distribution with mean zero and standard deviation of 1 on each predictor. This allows the model to shrink the associations, such that only the clearest relationships are preserved – noisy estimates are shrunk to zero (Kruschke, 2015). Given that all variables are standardised, this can be interpreted as a prior over the effect sizes, as each coefficient represents a 1SD increase in the predictor being association with the coefficient value change in IC. We then examined the posterior distribution of each coefficient, and assessed the probability that the estimate fell within certain bounds as well as its likely direction of association (i.e., the probability the effect positive or negative, (Makowski et al., 2019). Specifically, we recovered the probability the effect was within a traditionally small (±0.2SD unit change), medium (±0.5SD unit change) or large (±1SD) region. For all coefficients, we describe their posterior mean as well as the 95% credible interval (CrI), which indicates the span within which there is a 95% probability of the true association being. Models were estimated using the Bambi package (Capretto et al., 2022) in the Python programming language.

Results

Descriptive Statistics

Unstandardised and standardised (z-scores) means and standard deviations (SD), credible values, associations and probability of having a large, medium and small effects for the predictors are shown in Table 2.1.

Bayesian Model

Our regularised regression model showed that only a few variables had a least a 95% probability of being positively or negatively associated with IC trait scores (Figure 2.1). We then evaluated the probabilities of a .2, .5, and 1 SD change in ICS given a 1 SD increase in each variable, mapping broadly onto small, medium, and large effect size changes (Table 2.1).

Anxiety towards aging had a positive association with IC trait, b = 0.36 95% CrI [0.14, 0.56]. This variable had a zero probability of a large effect (1SD), just a 9% probability of a medium effect (.5) and an 83% probability of a small effect (.25). Perceived results also had a positive association, b = 0.16, 95% CrI[0.01, 0.32], with a zero probability of a large and medium effect, and a 14% probability of a small effect. Finally, the desire to have a facelift also had a positive association with IC trait, b = 0.17, CrI[-0.00, 0.34], with a zero probability of a large and medium effect, and a 20% probability of a small effect.

Motivations of self-esteem has been found to have a negative association with IC trait, b = -0.13, CrI[-0.25, .00], with a zero probability of having a large and medium effect, and a .03% of a small effect. Finally, knowledge about treatments (KAAD) also had a negative association, b = -0.17, CrI[-0.35, 0.01], with zero probability of having a large and medium effect, and a 19% probability of having a small effect.

Table 2.1

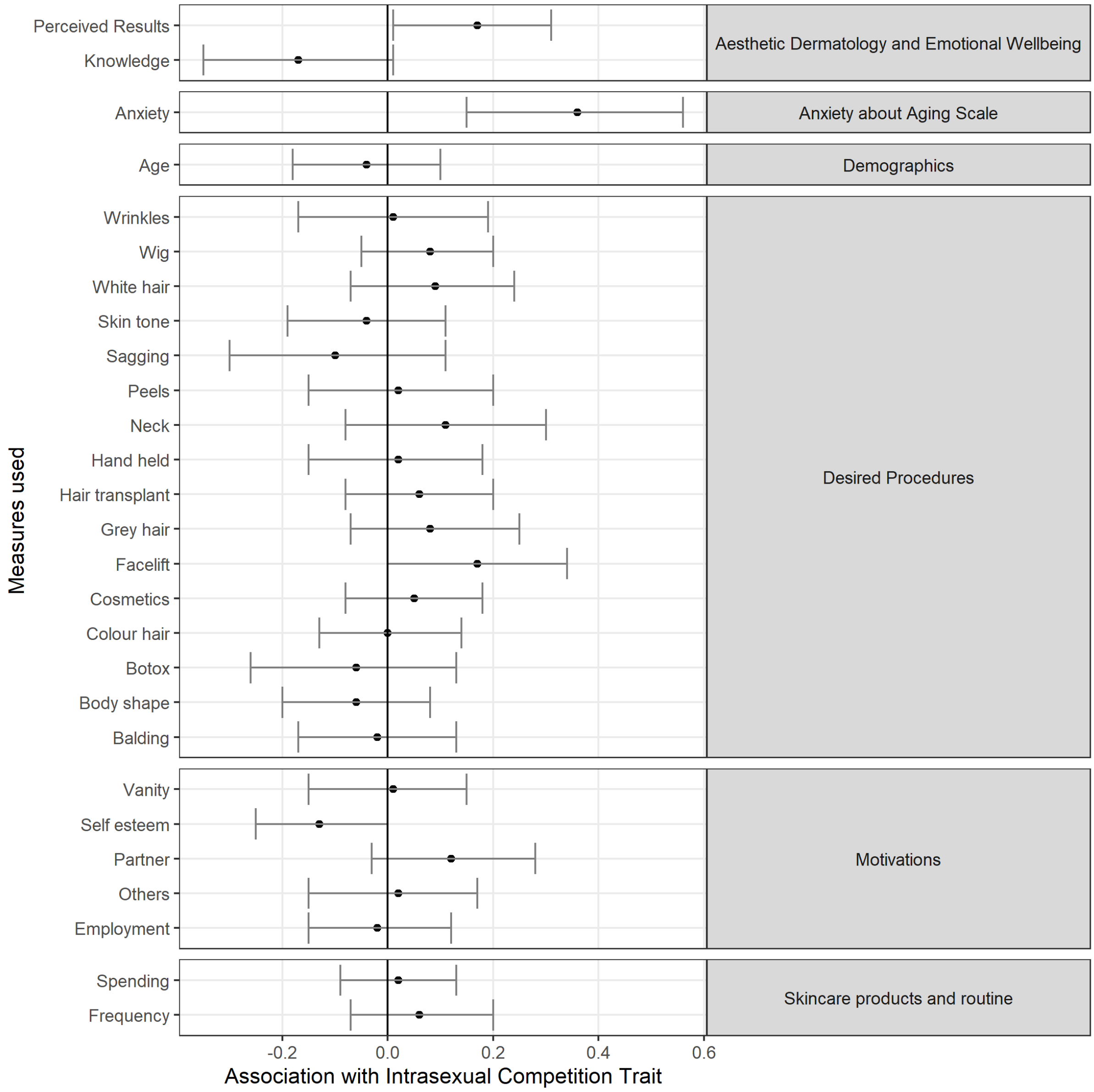
Descriptive Statistics and Probability Values of Predictor Variables.

Measure	Variables	M	SD	M	SD	2.50%	97.50%	Probability of + association	Probability of - association	Large effect (1)	Medium effect (.5)	Small effect (.25)
Demographics	Age	29.8	12	-0.04	0.07	-0.18	0.10	0.31	0.69	0.00	0.00	0.00
Skincare products and routine	Spending	1.42	0.71	0.02	0.06	-0.09	0.13	0.60	0.40	0.00	0.00	0.00
	Frequency	2.76	0.86	0.06	0.07	-0.07	0.20	0.82	0.19	0.00	0.00	0.00
Motivations	Self esteem	5.63	1.35	-0.13	0.07	-0.25	0.00	0.03	**0.98**	**0.00**	**0.00**	**0.03**
	Employment	3.81	1.77	-0.02	0.07	-0.15	0.12	0.40	0.60	0.00	0.00	0.00
	Partner	3.89	1.72	0.12	0.08	-0.03	0.28	0.94	0.06	0.00	0.00	0.06
	Others	2.79	1.57	0.02	0.08	-0.15	0.17	0.58	0.42	0.00	0.00	0.00
	Vanity	3.43	1.68	0.01	0.08	-0.15	0.15	0.54	0.46	0.00	0.00	0.00
Desired Procedures	Grey hair	4.18	1.42	0.08	0.09	-0.07	0.25	0.82	0.18	0.00	0.00	0.02
	White hair	3.98	1.42	0.09	0.08	-0.07	0.24	0.86	0.14	0.00	0.00	0.03
	Balding	2.74	1.42	-0.02	0.08	-0.17	0.13	0.38	0.62	0.00	0.00	0.00
	Wrinkles	3.46	1.29	0.01	0.09	-0.17	0.19	0.55	0.45	0.00	0.00	0.01
	Sagging	2.51	1.2	-0.10	0.10	-0.30	0.11	0.18	0.82	0.00	0.00	0.07
	Skin tone	3.27	1.24	-0.04	0.08	-0.19	0.11	0.33	0.67	0.00	0.00	0.00
	Neck	2.59	1.23	0.11	0.10	-0.08	0.30	0.87	0.13	0.00	0.00	0.08
	Body shape	3.39	1.27	-0.06	0.07	-0.20	0.08	0.18	0.82	0.00	0.00	0.01
	Colour hair	2.33	0.9	0.00	0.07	-0.13	0.14	0.50	0.50	0.00	0.00	0.00
	Wig	1.32	1.06	0.08	0.06	-0.05	0.20	0.89	0.11	0.00	0.00	0.00
	Hair transplant	1.21	1.11	0.06	0.07	-0.08	0.20	0.78	0.22	0.00	0.00	0.01
	Cosmetics	2.02	1	0.05	0.07	-0.08	0.18	0.77	0.23	0.00	0.00	0.00
	Hand held	1.46	1.12	0.02	0.08	-0.15	0.18	0.61	0.39	0.00	0.00	0.00

	Peels	1.43	1.14	0.02	0.09	-0.15	0.20	0.58	0.42	0.00	0.00	0.01
	Botox	0.73	1	-0.06	0.10	-0.26	0.13	0.28	0.72	0.00	0.00	0.03
	Facelift	0.57	0.93	0.17	0.09	0.00	0.34	**0.98**	0.02	**0.00**	**0.00**	**0.20**
Anxiety about Aging Scale	Anxiety	24.6	6.95	0.36	0.11	0.15	0.56	**1.00**	0.00	**0.00**	**0.09**	**0.84**
Aesthetic Dermatology and Emotional Wellbeing	Perceived Results	27.8	10.4	0.17	0.08	0.01	0.31	**0.98**	0.02	**0.00**	**0.00**	**0.14**
	Knowledge	12.1	4.36	-0.17	0.10	-0.35	0.01	0.03	**0.97**	**0.00**	**0.00**	**0.19**

Figure 2.1

Regularised Posterior Model.



Note: Variables with credible lines that crossed the 0.0 threshold do not have 95% probability of excluding zero.

Discussion

We conducted a survey where we asked participants about their skincare behaviours, attitudes about cosmetic products and procedures, evaluations of signs of aging, anxiety about aging, and how competitive they are with members of the same sex. We explored whether individual factors could predict competitiveness of individuals using Bayesian regression models.

Our exploratory findings suggest that when all factors were considered, only five variables clearly predicted IC scores. An individual's anxiety about appearing old and the perceived benefits from cosmetic procedures, and facelift variables were positively associated, where the higher the anxiety towards appearing old, the more positive one views the results from cosmetic procedures, and the more one is likely to undergo a facelift procedure, the higher the competitive trait of a person. Previous quantitative studies have highlighted that females who are anxious about appearing 'old' were more likely to be accepting of cosmetic surgery (Muise & Desmairas, 2010; Slevec & Tiggeman, 2010). The current study is the first to demonstrate that anxiety towards appearing older predicts female competitiveness. At the same time, we are also the first study to find that women who believe that aesthetic dermatology would bring positive outcomes in different aspects of their lives also predicted high IC trait. Taken together, this shows that women who are worried about their appearance and have positive evaluations about the effects of cosmetic treatments were more likely to be those with higher competitive trait. This is in line with the idea of using self-promotion as a strategy to compete with other females (Fisher & Cox, 2011), that is, females may perceive appearing 'old' as being less able to compete with their younger looking counterparts, however, with the use of dermatological advances, this could increase their 'odds' at competing with others, as well as having benefits in other aspects of their lives.

Conversely, motivations of self-esteem and knowledge about anti-ageing treatments were negatively associated with competitiveness, that is, the use of anti-ageing products for self-esteem reasons and the more knowledge about anti-ageing treatments, the less likely they are to be competitive towards individuals within their sex.

Our results, which showed that spending and use of skincare products do not predict IC, was unexpected. Previous studies have demonstrated that behaviours such as amount of money spent and frequency of buying beautification products would be indicative of competitiveness (Wagstaff, 2018), however, it could be that the previous study specifically talked about make-up products (Mafra et al., 2020; Mileva et al., 2016), e.g. lipstick, foundation, and concealer, rather than skincare products, e.g. facial cleansers, toners, and moisturisers. The use of make-up will result to salient and short-term change in appearance, while skincare products do not, although they are viewed to contribute to long-term maintenance and 'preservation' of the natural aging process (Clarke & Griffin, 2007). Further research comparing different beautification strategies could be beneficial in understanding the how one beautification strategy could be more associated with female competitiveness.

Next, we also found that overall, the motivation behind age concealment, evaluations of signs of aging, and desired procedures (apart from facelift) did not clearly predict ICS. This does not support findings from previous studies, which suggested that females would be more competitive when looking for partners (Wang et al., 2021). This could be due to our study not having a competitive context and therefore personal motivations behind age concealment may not be salient.

Facelift was the only cosmetic procedure that showed a positive association with IC, where the more an individual would want to have a facelift, the higher their competitive scores were. Again, this could be due to the lack of context of competition in our survey.

Overall, the current study is the first to demonstrate that evaluations of signs of aging as well as desire to engage in dermatological procedures to appear younger, when considered with other factors, do not predict female competitiveness.

One limitation of the current study are the self-report measures and correlational design employed. Participants could have answered the survey in a way that would make themselves look good, e.g. not wanting to appear vain, and therefore our data may not capture the real views of participants. Although our findings add to our understanding of how certain behaviours and personality traits are related to female IC, the correlational design of the survey means that we are unable to infer that having a high IC trait causes women to want to engage in cosmetic procedures or be anxious about appearing old.

Another limitation of our study is that we did not specify whether the evaluation of the signs of aging applies to the participants or the targets. This could therefore influence how participants would appraise the signs of aging listed, that is, some participants may think that wrinkles on other people is a negative, but think that it is a positive thing when they have it. However, the internal reliability score for these items was acceptable and therefore we could infer that the participants rated the items in a similar manner. For future research, explicit descriptions as to the direction of evaluations would be useful, that is, are the participants evaluating the signs of aging on themselves or is it on other people?

However, it is important to highlight that compared to the standardised method of using the intrasexual competition scale to correlate intrasexual competition trait and aesthetic dermatology attitudes, a different way of measuring, and therefore inferring, intrasexual competition trait was used by Wang et al. (2021). In their study, they have incorporated a competitive context by manipulating the number of men and women in their experiments and observed an interesting relationship between the number of plastic surgeons and the ratio of men and women in a given area. This has offered a different perspective as well as getting around the bias that participants could present themselves more positively, rather than their actual attitudes. Unfortunately, this study has come out after the data for Chapter 1 were collected and therefore a competitive context for these studies were not included. Future studies would benefit in adding a competitive context in their design and discussions on implications on intrasexual competition trait measurement would be welcome.

In sum, the current study has highlighted that use of skincare products, as opposed to make-up and cosmetic procedures, do not predict female IC trait, despite it being viewed as a more natural approach to aging. We demonstrate that an individual's anxiousness about looking older, self-esteem, desire for a facelift, and perceived benefits and knowledge of aesthetic dermatology were likely predictors of female competitiveness.

Next steps

From here, we saw that one's knowledge of and perceived outcomes from cosmetic procedures are likely to predict their competitiveness towards individuals of the same sex. In the next chapter, we then explore how observers' competitiveness would influence how they evaluate individuals who engage in three levels of age concealment and varying motivations.

Chapter 2.2 Perceptions of Individuals who Engage in Age Concealment[1]

Abstract

Previous literature suggests that individuals who engage in age concealment are viewed differently depending on the type of concealment used, motivations behind engagement, and to some extent, the age of the target individual. This study aimed to expand on the literature by integrating perceiver factors such as gender, age, and individual differences in intrasexual competition, alongside the individual target factors such as concealment type and motivation for use. Using a sample of 306 participants recruited online, a linear mixed model found main effects of the target's motivation and concealment type, and perceiver's gender and intrasexual competition, but not perceiver age on target evaluations. We also found that females evaluated the targets most positively when age concealment was motivated by self-esteem, followed by employment and least positively for romantic purposes whereas males did not differ on their evaluations based on motivation. Finally, we found that the higher the female participant ICS trait, the less positively they rated the targets. These findings suggest that the general perception towards the type and motivations behind the engagement have not changed despite the increasing access to age concealment, and that perceiver trait differences also play a role in how concealers are evaluated.

[1] For the published article, please see:
Childs, M. J., & Jones, A. (2022). Perceptions of individuals who engage in age concealment. *Evolutionary Behavioral Sciences.* Advance online publication.
https://doi.org/10.1037/ebs0000305

In society, more attractive people have better outcomes in different aspects of life, be it employment, friendship circles, and most importantly, finding romantic partners (Langlois et al., 2000). However, research also suggest that as people get (and look) older, the less attractive they are perceived (Samson et al., 2010). This change has been linked to older people reporting lower self-esteem and a discrepancy between how old they feel and how old they look (Clarke et al., 2007; Muise & Desmarais, 2010; Slevec & Tiggemann, 2010), which has been cited as one of the main reasons for people engaging in anti-aging procedures (Muise & Desmarais, 2010; Tian et al., 2020). Conversely, studies also show that in general, older people who engage in age concealment techniques were viewed as vain (Chasteen et al., 2011) and were generally evaluated negatively by perceivers (Harris, 1994; North & Fiske, 2013; Schoemann & Branscombe, 2011). However, these studies were conducted when anti-aging procedures were considered invasive, have long recovery times, and were expensive. The current technological advances in cosmetic dermatology have paved the way for less invasive procedures, including home-use devices (Juhász et al., 2017) which now result in faster recovery time and fewer complications compared to previous invasive procedures such as Botox, dermal fillers, and face-lifts, and are more affordable to the general population. This could therefore have an impact on how individuals who choose to engage in these procedures may be viewed.

However, one thing that has not changed is that the highest consumer group has been found to be middle-aged women (The American Society for Aesthetic Plastic Surgery, 2018). Furthermore, traditional gender roles are changing, with women opting to either not have children or postpone having children in favour of their careers, where women aged 40 years and over being the sole age group with increased conception rates (Office for National Statistics, 2020). With the shift in our societal roles and the accessibility of procedures to reduce signs of aging and maintain youthful appearance of middle-aged women, several social perceptual questions therefore arise - namely, how are middle-aged women who engage in these procedures perceived, and how might this differ as a function of the type of concealment used, their motivations, and other demographic variables?

Target Concealment Type

Anti-aging techniques have their roots in facial reconstructive surgery, a field which emerged to help disfigured soldiers integrate back to society and has since flourished as a commodity to alter individuals' appearance to reduce signs of facial aging (Chaterjee, 2007). The demand in this field has paved the way for the technology to be refined and be more accessible to consumers of all types. Previously, anti-aging techniques have been associated with high cost in terms of affordability, recovery times, and complications, and results that appear unnatural (Clarke & Griffin, 2007). Recently, technological advancements in the field have allowed the development of non-ablative techniques, e.g. using less invasive procedures such as light therapy and chemical peels, which are more affordable, shorter recovery periods, have fewer contraindications (Beilin, 2011). It is relatively unknown, therefore, whether there is a change in how observers perceive middle-aged women who conceal their age using these new techniques.

Previous studies, relying on descriptions of different target individuals who engage in varying types of age concealment, have shown more negative evaluations for more extreme procedures such as facelifts (Harris, 1994), whereas targets using mild or natural (Botox and fillers or avoiding exposure to sunlight, respectively) were rated the most positively (Chasteen et al., 2011). However, these studies were conducted when age concealment procedures were less accessible and were associated with high recovery times and complications, and with the advancement of the field, we are interested to see whether these perceptions have changed.

As Chasteen et al. (2011) found that extreme procedures received the most negative evaluations, the current study therefore aimed to compare the evaluations between *moderate* concealment types such as home-use devices which use intense pulse light technology and radiofrequency (Juhasz et al., 2017), and *major* treatments such as Botox and fillers, as these are still high on demand among consumers.

Our first hypothesis therefore is that (1) there will be a main effect of concealment type in the perception of middle-aged women who engage in anti-aging procedures, where moderate concealment would be evaluated more positively than major concealment.

Target Motivation Type

In females, signs of aging such as appearance of wrinkles, sagging skin, and uneven skin tone, has been linked to lowered female mate value (Buss, 1989; Maestripieri et al., 2014) and thus being at a disadvantage to attract potential mates. It is therefore understandable that for women who want to seek romantic partners, maintaining a youthful appearance is important (Harris, 1994; Swami et al., 2013). In line with this, research has found that older women who engaged in anti-aging procedures were rated as more attractive and healthier (Nellis et al., 2017; Tian et al., 2020) which therefore implies that age concealment could prove to be beneficial for older women who are seeking partners. However, studies have shown that such motivations were not viewed positively by others.

Using descriptions of individuals who engaged in different types of age concealment and for varying motivations, Harris (1994) found different ratings for varying motivations behind concealing one's age, with vanity and self-esteem reasons receiving the highest positive rating, followed by employment and finding a partner, and pleasing others as the lowest. Self-esteem is usually seen as a person's self-worth, often associated with feelings of adequacy. However, it has also been defined as how others value the person – in other words, one's self-esteem is a result of feedback given by other people. Leary (1999) suggests that we monitor our social relationships (sociometer theory), and thus, depending on whether our relational value increases or decreases, so does our self-esteem as a response and motivates the individual to act on it. In turn, we could argue that as we age, we lose our relational values, i.e. in general, younger perceivers judge older people more negatively (less warm and less capable) compared to younger (d = .24, Kite et al., 2005), therefore engaging in age concealment behaviour could be seen as a response in order to increase one's relational value, and therefore increase one's self-esteem.

In line with this, appearance of signs of aging has been negatively correlated with wellbeing for women (Harris, 1994; McFarland, 1999; Muise & Desmarais, 2010; Slevec & Tiggemann, 2010), therefore increase of self-esteem and positive body image in older age has been cited to be the underlying reason for engagement in age concealment techniques (Muise & Desmarais, 2010; Slevec & Tiggemann, 2010; Slevin, 2010). Additionally, Bennett et al., (2017) found in their interviews that older women (aged 69-94) engage in different appearance management behaviours such as make-up and anti-aging creams in order to promote well-being, which may suggest that self-esteem motivations play a large part when engaging in age concealment.

Recently, Tian et al (2020) used images of middle-aged and older-aged individuals pre- and post-anti-aging procedures, e.g. face-lift, eyelid surgery, and browlift, and asked undergraduates to rate them on different personality traits, employability, and attractiveness. They found that post-operative images were rated to be more hireable in comparison. As the human face is used as a cue of social status, income, and employment (Bjornsdottir & Rule, 2017; Nash et al., 2006), it is possible that current economic and labour market conditions could threaten older workers, where signs of aging are associated with negative traits such as fragility, resistance to change, and being less productive than younger workers (Hummert et al., 1997;

Perry & Finkelstein, 1999), which highlights another motivation for engaging in age concealment procedures.

Our second hypothesis therefore is that (2) there will be a main effect of motivation type on evaluations of middle-aged women who engage in age concealment, however, we are not able to give a prediction of how the different motivation types would be evaluated due to the changing societal attitudes towards romance, employment, and self-prioritisation.

Perceiver Age

Another factor that influences how individuals who engagement in age concealment are viewed is the age of perceiver. Overall, older participants were more likely to be accepting of age concealment behaviour than younger participants (Chasteen et al., 2011; Harris, 1994; Schoeman & Branscombe, 2011). It was argued that older people wanting to appear younger may threaten the social identity of younger observers, thus receiving negative evaluations (Schoeman & Branscombe, 2011). Another explanation for such evaluations could be that engaging in these behaviours may be considered atypical, and therefore older people who engage in them may be viewed negatively, e.g. desperate and vain (Harris, 1994; Schoemann & Branscombe, 2011).

We therefore hypothesise that (3a) perceiver's age will have a main effect on evaluations of middle-aged women who engage in age concealment, where older perceivers would give more positive evaluations than younger perceivers; and (3b) that this will interact with motivation type, where younger males and females would give negative evaluations for both romantic and employment reasons, but not for self-esteem.

Perceiver Gender

Recent statistics has shown that an increasing number of males are also now engaging in cosmetic procedures (The American Society for Aesthetic Plastic Surgery, 2018).

Traditionally, use of make-up and appearance enhancing methods have been attributed to females, therefore we expect females to be more accepting of such behaviours (Clarke & Griffin, 2008). Some would argue that this is due to the double standard of aging, where there is more pressure for women to look younger, as appearing older is associated with more negative evaluations, e.g. fragility, incompetence, less healthy, and less attractive (Sontag, 1979).

Evolutionary perspectives explain such phenomenon as a by-product of female reproductive function, where younger women are more favoured, particularly by men, as they are more able to produce offspring (Buss, 1989; Buss & Schmitt, 1993; Harris, 1994). In support of this theory, studies have found that men judge older women to be less attractive when they were looking for potential romantic partners (Maestripieri et al., 2014; Teuscher & Teuscher, 2007). However, research has shown that women above 30 years old have diminished likelihood of childbearing and increased maternal complications compared to women between 20-29 years old (Salihu, 2003). Therefore, as women between 35-50 years old make up the majority of those who engage in anti-aging procedures, this has implications for motivations of finding a potential mate for men, e.g. how would men evaluate women who want to appear younger to find potential mates?

From here, we hypothesise that (4a) there will be a main effect of gender on evaluations of middle-aged women, where males in general would give more negative ratings than females; and (4b) that this would interact with motivation types, where males would give the lowest ratings for romantic motivations, compared to employment and self-esteem reasons. To our knowledge, this would be the first study to explore the relationship between perceiver gender and target motivation on evaluations of age concealment.

Perceiver Intrasexual Competition

Finally, another factor which could affect how perceivers view those who engage in age concealment could be the competitiveness of the perceivers themselves. Theories of intrasexual competition posit that as there is a finite number of ideal mates, men and women would have to compete with same-sex individuals to get access to potential partners (Buss & Schmitt, 1993; Cox & Fisher, 2008; Wang et al., 2021; Wyckoff et al., 2019). As men tend to look for young, fertile partners (Buss, 1989), women who attempt to conceal their age through cosmetic means could be viewed by other women more negatively, as this would increase their possible competitors (Fink et al., 2014).

Additionally, Arnocky et al. (2019) found that women with higher ICS were more aggressive towards the target when they appeared in a sexualised manner (wearing more revealing clothes and make-up applied) compared to conventional manner (wearing long-sleeved top and no make-up applied), and that this is due to the sexualised target being perceived as lower in humanness than the conventional counterpart. These show that a female perceiver's competition trait influences how they would perceive targets who dressed more sexually. In relation to our study, we could infer that targets who engage in anti-aging procedures with the aim to find a partner (and to some extent, employment) could be viewed more negatively by female perceivers with high competitiveness as they would potentially be competing for resources.

However, there is some evidence which show that women enhance their appearance to impress other women, rather than simply attracting a mate (Mafra et al., 2020; Mileva et al., 2016; Wagstaff, 2018). Mileva et al. (2016) found that female raters judged women with make-up as more dominant than those without, implying that certain behaviour could be targeted to change how other women perceive them, rather than simply attracting a partner. Similarly, Wagstaff (2018) have found that how often women use make-up is predicted by their sexual strategies and are highly related to their intrasexual competitiveness. Another study by Mafra et al. (2020) has demonstrated that women's intrasexual competition trait and desire to attract a mate predicted frequency of make-up use. On a similar note, Wang et al., (2021) found that women focus more on their appearance when there is a higher density of women in their environment, compared to when there are more men, suggesting that when there are more competitors, the more women focus on enhancing their appearance. This could be a strategy to attract a potential mate (Buss & Schmidt, 1993), which is referred as self-promotion. Another way of competing towards others is by derogating the other person's appearance (Cox & Fisher, 2008) in order to reduce their value to potential mates. It has been shown that women engage in derogatory tactics such as gossiping and labelling the competition with negative traits, e.g. vain and desperate (Kellie et al., 2021).

From here, we hypothesise that (5a) the perceiver's intrasexual competition scores (ICS) will have a main effect on evaluations of middle-aged targets, where the higher the ICS, the more negative the evaluations will be; and (5b) that this will interact with participant age and gender, where younger female participants would be likely to have higher ICS; and (5c) this will also interact with motivation types, where those with higher ICS would give more negative evaluations to those engaging in concealment due to romantic and job reasons, compared to self-esteem.

Current study aims and motivations

Previous studies have shown that various factors influence how individuals who attempt to enhance their appearance using cosmetics and anti-aging techniques have been evaluated. Given the increasing popularity and easier accessibility of less invasive anti-aging techniques to both genders, as well as societal shifts in terms of finding a partner, competitiveness in the labour market, and the surge of 'self-care' movements, it is therefore important to explore whether perceptions of engagement in anti-aging techniques have also shifted.

Following Harris (1994) and Chasteen et al. (2011)'s methods of using vignettes, the current study aimed to investigate how different perceiver factors (age, gender, and intrasexual competition), target factors (concealment type and motivation type) and their interactions would predict overall ratings of middle-aged women who engage in age concealment.

In summary, this study explores how evaluations of middle-aged women who conceal their age would be predicted by:

(1) Concealment type – particularly moderate (use of hand-held devices) and major procedures (Botox and fillers), implying that severity of procedure could influence whether the action is acceptable;
(2) Motivation type – whether the age concealment is motivated by self-esteem, looking for employment, or seeking romantic partners, implying that reasons behind age concealment could make the action more acceptable;
(3) Perceiver's age – whether younger or older raters would have differing perceptions of target individuals, implying that one's age influences how middle-aged women are perceived for their behaviour;
(4) Perceiver's gender – whether male or female participants would be more accepting of the behaviour, implying that gender differences would exist in evaluations of women who engage in age concealment;
(5) Intrasexual Competition Scores (ICS) – whether those with high or low ICS would influence evaluations, implying that age concealment behaviours could be viewed as a way of increasing competition; and
(6) The interactions between the above variables.

Methods

Participants

493 participants accessed an anonymous link to the study on Gorilla platform (Anwyl-Irvine et al., 2020). Data was collected between 12 January 2021 and 10 February 2021. Participants were recruited through social media platforms (Facebook and Twitter) and recruitment platforms (surveycircle.com and SONA). Eligibility criteria were aged 18 and over and had access to internet to complete the survey. There were no restrictions regarding country of residence and sex of participants. Three hundred and six participants completed the tasks and were included in the analysis. Following data cleaning (see below), two participants were removed, leaving 304 participants (230 F, 74 M) with an age range of 18-67 (M = 27.50, SD = 9.51).

Statement of Ethics

Written consent forms were acquired before participants were presented the study. Participants accessed the study using an anonymous link and were able to withdraw by not completing the study at any time. Only completed tasks were included in our analysis. All participants were given an option to be included in a raffle draw as compensation for their time, and a study credit (1) was awarded when the study was accessed through SONA. This study was approved by the Swansea University Ethics Committee and followed the Declaration of Helsinki (Williams, 2015).

Materials

Vignettes (Chasteen et al., 2011)

The vignettes followed the structure from Chasteen et al.'s study (2011; see Appendix G). These consist of a description of a middle-aged woman engaging in either a moderate (non-invasive, hand-held device) or major (Botox and fillers) procedures to conceal their age, for three different reasons: looking for a job, romantic partner, or for self-esteem. For example, "Angela is a middle-aged woman who wants to maintain a more youthful appearance to look for a *romantic partner.* She regularly uses *non-invasive techniques such as light therapy* that she could use at home as part of her anti-aging routine." Each vignette follow the same format, with the motivation type and concealment type changed accordingly. The vignettes were presented on their own first in the middle of the screen with no time limit. After each vignette, participants were asked to rate each target individual on eight traits, following Harris' study (1994): admirable, attractive, conceited[r], foolish[r], interesting, pathetic[r], vain[r], and wise.

For our vignettes, we decided to describe only middle-aged targets as they are the highest consumers of anti-aging procedures (The American Society for Aesthetic Plastic Surgery, 2018). Furthermore, for the interest of time and contemporary changes in the market, we opted only to use moderate (non-invasive, hand-held devices) and major (Botox and fillers) in our vignettes as these are currently the most popular procedures. In addition, data from Chasteen et al. (2011) found that those who used mild procedure received the most positive evaluations, and those who used extreme procedures received the most negative evaluations, and we believe that this would still be the case.

Intrasexual competition scale (ICS; Buunk & Fisher, 2009, see Appendix F)

This is a 12-item questionnaire which aimed to measure how competitive an individual is towards people of the same-sex. Participants were presented a statement relating to their attitude towards same-sex individuals and were asked to rate on a 7-point Likert scale: 1 (not at all applicable) to 7 (completely applicable). Items include 'I wouldn't hire a very attractive man/woman as a colleague,' and 'I can't stand it when I meet another man/woman who is more attractive than I am.' Sums for the 12 items were calculated, with a maximum score of 84. The higher the total score, the more competitive they are with the same-sex individuals.

Procedure

Participants accessed the anonymous study link either through student recruitment sites for course credit or social media advertisement. After providing consent and demographic information (e.g. age, sex, and ethnicity) participants completed the ICS.

The participants were then presented with a total of six individuals who engage in different age concealment techniques for varying reasons. Each trial consisted of the description first – there was no time limit to the presentation of the vignette – before the participant continued to the evaluation component. The vignette was kept on the left side of the screen, with the traits to be measured presented on the right side. Each trait was followed by a sliding scale with values of 0 (Not at all) to 100 (Extremely). The traits to be evaluated were presented in two screens. The trials were presented in random to the participant to avoid order effects. The study took approximately 8-10 minutes to complete.

Data Cleaning

187 participants did not complete the tasks and therefore were removed from the dataset. In addition, we calculated the standard deviation (SD) for within each participant's responses and removed those who had a standard deviation of zero, as this meant the participant gave consistently the same answers in the study. From this procedure, one participant was removed. Furthermore, as we are looking at gender differences, and we only had one participant who identified as other, we decided to only include participants who identified as male or female. This yielded a final sample of 304.

Design and Analytic Strategy

We fitted a linear mixed effects model in R (R Core Team, 2013) using lme4 (Bates et al., 2014) with a mean rating (averaging all variables together, after reverse scoring conceited, foolish, pathetic, and vain) as the outcome variable, with fixed effects of participant age (scaled), participant gender, participant ICS (scaled), concealment type, motivation type, and their interactions. Participants were used as random effects, reflecting that the ratings come from different individuals. This statistical model allows us to investigate the differences in evaluations of people who engage in different age concealment types (moderate or major) for different motivations (romantic, job or self-esteem) between males and females, across the age and ICS distribution.

The model is as follows:

$$\textbf{Mean Rating} = \mathbf{P_{Age(scaled)}} * \mathbf{P_{Gender}} * \textbf{Motivation} * \textbf{Concealment} * \mathbf{P_{ICS(scaled)}} + \mathbf{(1|P)}$$

Note. P stands for 'participant', where the age, gender, and ICS values were collected from the participants, rather than the target vignettes.

Results

Descriptive Statistics

Table 2.2 presents the means and standard deviations of the averaged ratings given by the participants to each target vignette. Overall, female participants gave higher ratings for the targets (*M* = 56.80, *SD* = 17.00) than males (*M* = 52.88, *SD* = 15.90), moderate concealment procedures were given higher evaluations (*M* = 60.93, *SD* = 14.38) than major concealment (*M* = 50.75, *SD* = 17.54), and self-esteem motivations received the most positive evaluations (*M* = 58.75, *SD* = 16.25), followed by looking for job (*M* = 55.42, *SD* = 16.76), and looking for romantic partner received the lowest evaluations (*M* = 53.36, *SD* = 17.03). Our raw data and code could be seen in https://osf.io/pj6h8/.

Table 2.2

Means and Standard Deviations of the Mean Rating for Each Vignette from Female and Male Participants.

Gender		Romantic			Job			Self-esteem		
		Moderate	Major	Overall Romantic	Moderate	Major	Overall Job	Moderate	Major	Overall Self-esteem
Females (n=230)	Mean	58.83	48.52	**53.68**	60.73	51.76	**56.25**	65.71	55.22	**60.46**
	SD	14.9	18.14	17.36	14.41	17.72	16.75	13.83	16.75	16.22
Males (n=74)	Mean	58.26	46.47	**52.36**	58.54	47.13	**52.83**	58.34	48.54	**53.44**
	SD	11.81	17.46	15.99	14.81	16.39	16.59	12.95	15.76	15.19

Note: The maximum rating for each vignette is 100.

Perceptions model

The complete estimated coefficients for our model are shown in Appendix H. We found several significant predictors: gender (b = -4.47, t(643.9303) = -2.128, p = .034), concealment type (b = 8.92, t(1480) = 9.642, p < .001); and ICS (b = -4.33, t(643.9303) = -4.15, p < .001. We also found significant interactions between age and romantic motivation (b = -1.999, t(1840.001) = -2.166, p = .030), age and self-esteem motivation (b = -2.394, t(1480.001) = -2.594, p = .009) and gender and ICS scores (b = 4.901, t(643.930) = 2.336, p = .02). Other interactions were not significant, p > .05.

Main effects

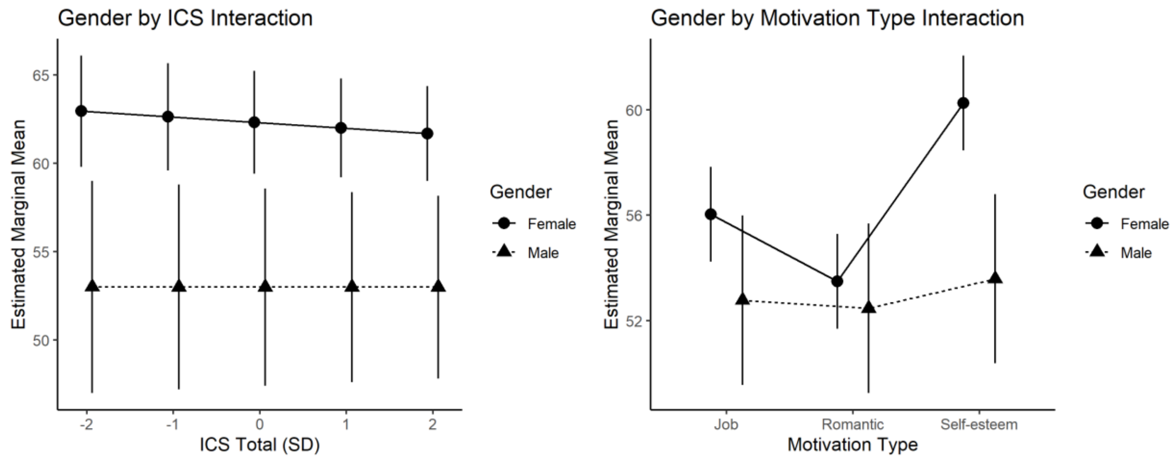
To further investigate the significance of our model, we conducted an ANOVA (using Type III sums of squares) on the fitted linear mixed model in R (R Core Team, 2012). Here we found a significant main effect of gender, F(1,296) = 4.57, p < .001, ηp^2 = .02; where females gave higher ratings (M = 56.80, SD = 17.002) than males (M = 52.88, SD = 15.90). There was also a significant main effect of motivation type, F(2, 1480) = 17.786, p < .001, ηp^2 = .02, where concealment due to romantic pursuits were rated the lowest (M = 53.36, SD = 17.03), followed by employment (M = 55.42, SD = 16.76) and self-esteem reasons (M = 58.75, SD =16.25). We also observed a significant main effect of concealment type, F(1, 1480) = 364.05, p < .001, ηp^2 = .20, where moderate treatment was rated higher (M = 60.93, SD = 14.38) than major treatments (M = 50.75, SD = 17.54); and a significant main effect of ICS, F(1, 296) = 5.115, p < .024, ηp^2 = .02, where the higher the participant's ICS, the lower the mean rating they provide..

Interactions

Figure 2.2 demonstrates the two-way interaction found between gender and motivation, F(2, 1480) = 9.02, p < .001, ηp^2 = .01. Pairwise comparisons using *emmeans* package (Russel et al., 2017) showed that male and female participants were similar in their evaluations of romantic (p = .59) and job motivations (p = .08), but were significantly different in their evaluations for self-esteem reasons ($p < .001$), where female participants gave higher ratings for (M = 60.46, SD = 16.22) than male participants (M = 53.44, SD = 15.19), $p < .001$. Furthermore, we found that within genders, male participants did not differ in their ratings across the three motivation types (all comparisons $p > .05$), whereas female participants gave significantly different ratings across the three motivations ($p < .001$) where they rated romantic reasons the lowest (M = 53.68, SD = 17.36), followed by employment (M = 56.25, SD = 16.75), and self-esteem received the highest evaluations (M = 60.46, SD = 16.22).

Figure 2.2

Illustrations of Interactions between Variables.



Note: Panel on the left depicts the interaction between participant gender and participant intrasexual competition scores. Panel on the right depicts the interaction between participant gender and target motivation type.

We also found a significant two-way interaction between gender and ICS (Figure 2), $F(1,296) = 4.95$, $p = .03$, $\eta p^2 = .02$. We conducted an estimated marginal means analysis (*emmeans* package on R, Russel et al., 2017) on the ICS scaled to their standardised scores (-2, -1, 0, 1, 2) between each gender. The pairwise comparisons yielded significant differences between male and female participants for each level of ICS, where female participants consistently gave higher ratings than male participants (all comparisons $p < .001$) regardless of ICS scores.

Discussion

The current study investigated how target factors (motivation and concealment type), and participant factors (age, gender, and intrasexual competition scores) would predict how the target individuals who engage in anti-aging behaviour would be evaluated. We presented participants with six hypothetical middle-aged women who varied on their concealment type used and motivation behind the use of anti-aging procedures. To address our hypotheses, we ran a linear mixed model which allowed us to integrate the between and repeated measures variables in a single analysis.

We found a number of main effects and two-way interactions. First, our findings support hypothesis (1a) that there will be a main effect of concealment type, where, as expected, targets who engaged in moderate concealment received more positive evaluations than those who engaged in major concealment. This supports findings from previous studies, where it was found that more invasive procedures were rated negatively than milder ones (Chasteen et al., 2011; Harris, 1994). This implies that the general attitude towards more invasive procedures has not changed. It is worth remembering that although Botox (classed as major procedure in this study) is less invasive than extreme measures such as face-lift in Chasteen et al.'s study (2011), this could be appraised as more invasive than home-use products. It is also important to point out that although the moderate procedure in this study is relatively new to the market, the premise of achieving professional results at home could be viewed as less invasive and more natural (Juhasz et al., 2017). We understand that there are other appearance enhancing procedures that are currently gaining in popularity, such as dermal fillers, which provide instant changes in appearance, however, this is outside the scope of our rationale as we wanted to compare the relatively new domain of home-use devices to those of established anti-aging procedure such as Botox (Chasteen et al., 2011). This will be a good avenue for future research, however. In general, therefore, our findings firstly demonstrate that the overall perception of anti-aging procedures remain the same – that is, the less extreme, the more acceptable it is perceived.

Our data also support our hypothesis (2a) that there will be a main effect of motivation type on evaluations of individuals who engage in age concealment. However, we predicted that the three motivations would significantly differ with each other, instead, we found that romantic and employment motivations were rated similarly, and self-esteem reasons were significantly rated higher than the other two motivations. Self-esteem motivations receiving the highest evaluations support previous studies which found that the primary goal of most women wanting to engage in anti-aging procedures was to increase their confidence and body image (Clarke & Griffin, 2008; Muise & Desmairas, 2010; Slevec & Tiggeman, 2010). Overall, this finding highlights the idea that personal wellbeing as motivation for appearance enhancement is more accepted than other motivations. This could also be viewed as women wanting to increase their relational value as they get older (Leary, 2000), as the appearance of youth is perceived to be more positive compared to appearing old (Schoemann & Branscombe, 2011).

That employment motivations were not rated significantly different from romantic reasons, however, was not expected, as previous studies have shown that middle-aged and older-aged individuals who have engaged in appearance enhancement procedures were deemed as more hireable than their counterparts (Tian et al., 2020), and therefore could be argued that appearing younger to gain or progress in one's career would be more acceptable than finding a partner. However, a study by North and Fiske (2013) have shown that older target individuals were disliked by younger raters when they did not share their wealth, compared to those who were more generous. In relation to employment, we could infer that our middle-aged targets are viewed negatively due to them taking up resources (income) that would otherwise be taken up by others, that is, the younger group. Unlike North and Fiske's study (2013), however, we did not find an interaction of age and motivation type – although this could be due to the majority of our participants being in a younger age range, and with our current labour market being saturated, the idea that middle-aged targets further competing may have influenced the negative evaluations.

Romantic motivations receiving the most negative evaluations support findings from Chasteen et al (2011) and Harris (1994). One explanation for this could be that as a target's sex and age interact in terms of how the perceivers view them (Sng et al., 2020), our participants may have viewed our target as atypical for their sex *and* age, that is, younger women are typically viewed to be more invested in finding a partner and starting a family, therefore, a middle-aged woman trying to find a romantic partner may not fit the stereotype. Further research looking at how male middle-aged targets would be viewed could be beneficial to the literature.

In contrast to our hypothesis (3a), we did not find a main effect of participant age. This contradicts previous studies which found that younger perceivers rated older targets more negatively (Chasteen et al., 2011; Harris, 1994; North & Fiske, 2013; Schoeman & Branscombe, 2011). However, although we tried to recruit a wider age range of participants, our sample is relatively young (mean age = 27.50) and therefore the effect may not have been as salient as expected. It is therefore useful to recruit more older adults for future studies.

On the other hand, our findings support our hypothesis (4a) that there will be a main effect of participant gender on target evaluations, where male participants in general gave harsher ratings than female participants. This supports previous findings by Teuscher and Teuscher (2007) where male participants rated older female targets more negatively, and findings from Harris (1994) where female participants gave higher ratings to targets overall. Additionally, as women are the main consumers of cosmetic products and procedures (The

American Society For Aesthetic Plastic Surgery, 2018), it could be that female raters viewed engaging in age concealment behaviour as more typical and therefore had more positive assessments (Harris, 1994). However, as our study only included female targets, we are not able to explore how male targets would be evaluated. This could be a useful avenue for future research.

Additionally, our results support our hypothesis (4b) that gender would have an interaction with motivation type. However, the trend we originally predicted was not observed. Instead of male participants giving the lowest ratings for romantic motivations and their evaluations increasing for employment and self-esteem motivations, our data show that male evaluations did not significantly differ across the motivation types, whereas female evaluations did. In other words, male participants in general gave lower evaluations overall, regardless of the motivation. This is of particular interest, as such findings contradict the evolutionary perspective, where we expect men to view women who want to look younger to gain a partner negatively as this potentially conceals their reproductive value, a trait suggested to be sought after by males (Buss, 1989; Buss & Schmidt, 1993). It could be that with middle-aged women wanting to appear younger (and if they want to attract a mate), the pool for potential mates would increase and could therefore be beneficial for men. This, however, still poses as an issue in terms of reproductive value, as females face more pregnancy complications and risks as they get older (Maestriperi et al., 2014).

Furthermore, although female participants were more generous in their evaluations overall, they gave harsher evaluations when the target was looking for a romantic partner, followed by employment, and gave the highest evaluations for self-esteem. This supports previous findings by Harris (1994), where it was found that vanity and self-esteem reasons received the highest ratings than looking for partner and employment. This also implies that female participants, rather than male participants, pay more attention to *how* other women consume cosmetic products, and the motivation which influence such behaviour. Previous studies have shown a similar trend, where Mileva et al. (2016) have found that female observers perceived women who wore more make-up as threat to themselves. This therefore questions the concept of double standards theory, which claims that females engage in appearance enhancement for males (Sontag, 1979) – if this is the case, we would expect males to also have significant differences in their evaluations between motivation types. However, as we did not explore our participants' attitudes and behaviours regarding anti-aging procedures, we cannot fully make assumptions as to their personal motivations.

Finally, our data also support our hypothesis (5a) that the perceiver's intrasexual competition scores (ICS) will have a main effect on evaluations, and as expected, we found that the higher the ICS, the less positive the evaluations were. This is in support of previous studies which showed that female participants would engage in derogatory tactics in order to compete with rivals (Cox & Fisher, 2008; Wyckoff et al., 2019), in this study's case, more negative evaluations towards an individual who is aiming to appear younger.

Our results did not yield a two-way interaction between motivation type and ICS, however, contradicting hypothesis (5b). This implies that those with high competitiveness view others as competitors regardless of the reason behind their appearance enhancement. This supports findings from Arnocky et al (2019), where participants with high competitiveness trait were more aggressive towards our targets. We could infer that as all our targets were engaging in appearance enhancement, this on its own could be reason enough to be viewed negatively. However, unlike Arnocky et al's study (2019), we did not ask our participants to rate our targets on their humanness and therefore we cannot assume that the same psychological mechanism is at work here.

We also did not find a three-way interaction for age, gender and ICS (5c), implying that there are similar levels of ICS across the population, and that evaluations towards the targets are similar across the age range. We did, however, found a significant two-way interaction between participant gender and participant ICS, where males had similar ratings across all levels of ICS and female's mean ratings were significantly different for each level of ICS, where those with higher ICS gave higher evaluations. It could be that as our target individuals were female, intrasexual competition towards the targets is therefore more relevant to the female participants. This supports the general idea that females who are more competitive would view others as threat, and therefore would engage in tactics to reduce their rival's potential (Wyckoff et al., 2019).

One limitation of the current study is that we only investigated how male and female perceivers would evaluate *female* targets. This therefore did not allow us to fully investigate the double standards of aging, as we cannot make conclusions as to how male age concealers would be evaluated. To overcome this, future research could include both male and female targets and compare the evaluations between the two. We could expect that male perceivers ICS would have an influence on male, but not female targets, and vice-versa. However, as mating strategy of males do not depend on them looking younger (Buss, 1989; Buss & Schmidt, 1993), we do not anticipate changes in the evaluations of male targets as a function of perceiver age and target motivation type.

Another limitation of our study is the sole use of vignettes to describe the targets. Recent study by Tian et al. (2020) have shown that participants who were shown pre- and post-treatment photos of age concealers rated the target more positively on their post-treatment appearance. They argued that seeing the results of age concealment would negate the underlying stigma about age concealment. However, as the moderate treatment in the current study aims to be less invasive while aiming to deliver similar results to professional procedures, it could be that those who engage in more extreme procedures would be rated more negatively when participants are able to compare the results side-by-side. Such studies therefore would need to consider using independent samples to reduce carry-on effects.

Another limitation of the current study is that we did not provide an explicit definition of ‘middle-age’. It was previously shown that the perception of onset of ‘middle-age-ness’ differ between younger and older people, where older participants tend to attribute middle-age onset as later than younger participants (Chopik et al., 2018; Drevenstedt, 1976). It could therefore be that the subjective views of our participants have affected how they would evaluate our target. This could also explain why we did not find a significant main effect of age, but previous studies did (Chasteen et al., 2011). For future studies, therefore, it would be important to explicitly define the target’s age as this could affect evaluations.

In sum, the current study aimed to investigate whether the perceptions of people who engaged in age concealment has changed, given the current societal shift to increasing use of concealment techniques and personal priorities. We found that in general, the less extreme procedure is still regarded more positively, and male participants did not differ in their evaluations regardless of why the target engaged in age concealment. Such findings have implications on how we interpret previous theories which suggest that females primarily engage in such behaviours to attract mates, while the evidence here suggests it may be to compete with other females.

Next steps

This chapter explored the characteristics of individuals who would be likely to engage in anti-ageing procedures and how such behaviour was perceived by observers. Next, I explore what affects facial perception of attractiveness, health, and age.

Chapter 3.

Effects of motion, emotion, and target sex on social perception

Chapter 3.1 Effects of target motion, emotion, and gender on perceptions of age, attractiveness, and health.

Abstract

Previous studies consistently demonstrate associations between facial appearance and social trait evaluations, however, most have only used *young* target faces as their stimuli. The current set of studies aimed to determine the role of facial motion, emotion, and sex on perceptions of social traits using a wider age range of target stimuli. Study 1 used static neutral and dynamic smiling moving faces to investigate how facial age, attractiveness, and health were evaluated using static and dynamic faces of Young, Middle, and Old faces. Here, we found that age evaluations were similar across face motion conditions, whereas attractiveness and health evaluations were higher when faces were moving. Study 2 built on these findings and explored the influence of positivity effects on attractiveness and health. We used static and dynamic *smiling* faces and found only significant effects of motion on attractiveness ratings. This showed that health ratings were more affected by emotion rather than motion itself. Lastly, Study 3 explored whether facial sex affects ratings of health and attractiveness. We used data from static neutral (Study 1) and static smiling (Study 2) faces and found that in general, female faces were rated as more *attractive* than males, but no gender bias was observed for health ratings. We also found that for attractiveness ratings, older female and younger male faces benefit the most from smiling. These findings suggest that despite the close relationships between these social traits, perceivers use different facial cues depending on the trait being evaluated.

How we perceive a person's face has social consequences. Research on facial attractiveness has demonstrated that individuals who have attractive faces are more likely to have favourable social outcomes ranging from being perceived as more competent and therefore have better employment prospects (Jackson et al., 1995), having better friendship and romantic partnership prospects, particularly in the Western society (Anderson et al., 2008) and even lighter legal sentencing (Langlois et al., 2000). Facial attractiveness has also been closely linked to age estimations and health perceptions, where the younger a person looks, the more attractive (Ebner, 2008; Kwart et al., 2012) and healthier (Fink, Grammer, et al., 2006; Gunn et al., 2013; Samson et al., 2011) they appear to be. Such influence has been argued to underlie the increasing consumption of procedures which aim to reduce signs of aging (Chasteen et al., 2011; Tian et al., 2020) as well as procedures related to enhancing appearance such as rhinoplasty (Parsa et al., 2021) particularly in middle-aged women, although men consumers are also increasing in numbers (The American Society for Aesthetic Plastic Surgery, 2018).

However, research on perception of these facial traits have often used static stimuli, an approach that is problematic as our daily interactions are often dynamic in nature, e.g. speech and emotion expressions. Another issue with the current literature is that the studies tend to employ young target and observer samples, raising issues about the generalisation of such findings to older adults. The current work therefore aimed to contribute to the literature by exploring whether there will be an effect of motion on the observer's age estimates and evaluations of attractiveness and health using a wider age range of targets and observers.

Motion and perception of facial characteristics

Our daily interaction with people is marked with high level of motion from speech and emotion expressions. These dynamic signals and cues allow us to direct our behaviour towards others, e.g. happy faces facilitate an approach reaction, while angry faces facilitate avoidance reactions (Zebrowitz, 2011). The use of moving stimuli has also been shown to facilitate increased recognition of individuals with developmental prosopagnosia (Longmore & Tree, 2013; Xiao et al., 2014) compared to static stimuli. This therefore implies that static and dynamic images involve dissociable processes in relation to recognition (Dobs et al., 2018; Lander & Butcher, 2015).

Use of dynamic images in facial perception studies have shown that these stimuli provide information that are not present with static images, e g., from a person's gaze pattern we are able to know the focus of their attention or their state of mind (Dobs et al., 2018). The additional information, e.g., facial symmetry (Hughes & Aung, 2018) and the idiosyncratic, spontaneous features from the movement of the face (Dawel et al., 2021) could also influence the judgement of facial attractiveness, health, and age. Previous studies using static images have shown that these three key characteristics are highly related to each other (Fink et al., 2006, 2012; Matts et al., 2007), however, this has not been explored using dynamic images and a wider age range. The current study therefore aimed to compare how static and dynamic images are perceived on these three social traits using stimuli with wider age range.

Motion and attractiveness

The influence of facial attractiveness spans across various social spheres, where more attractive people are more likely to have better outcomes in employment, interpersonal relationships, leadership positions, and legal situations (Langlois et al., 2000). These wide-reaching consequences has been argued to be a significant underlying factor in the pursuit of anti-aging techniques, particularly for middle-aged women (Chasteen et al, 2011; Tian et al., 2020). Previous research on facial attractiveness (see Table 3.1) have explored the influence of motion in perception of facial attractiveness (Penton-Voak & Chang, 2008; Rhodes et al., 2011; Roberts et al., 2009; Rubenstein, 2005) and broadly speaking, it seems that motion does not affect attractiveness ratings. However, these have used younger age group as both targets and observers, and therefore we are unable to generalise their results to older samples.

Rubenstein (2005) was the first study to explore the difference in perception of attractiveness between female static and dynamic (reading text). They found no significant difference between the ratings, however, this study also found that the ratings for static and dynamic versions of the faces had low association with each other, that is, the perception of attractiveness of a static version of a face did not match the attractiveness rating of the dynamic version of the same face. This low association could be due to the between-subjects design used, where participants only saw one condition, and therefore individual differences between participants may be driving the results rather than facial motion itself. Contrast this to a study by Roberts et al. (2009) which found that attractiveness ratings for static and dynamic (speaking) stimuli were highly correlated when participants saw both versions of the face, in other words, as ratings for static faces increase, ratings for dynamic faces also increase. They also found that same-sex ratings were higher, i.e. female participants gave higher ratings of attractiveness to female faces, and vice versa.

From here, we could argue that to minimize the effects of individual differences in attractiveness ratings, employing a repeated measures design would be more appropriate. Rhodes et al. (2011) and Kościński (2013) then further found that motion had no significant effect on evaluations of attractiveness for young targets. However, although Roberts et al. (2009) demonstrated that using repeated measures may be more reliable in investigating the evaluations for static and dynamic images, and that motion did not have an effect on attractiveness ratings, these studies only used young targets, and therefore these results may not be generalisable to older adults.

Table 3.1

Studies Which Used Static and Dynamic Face Types for Attractiveness Ratings.

Authors	Model sex	Rater sex	Design	Dynamic Context	Results
Bowdring et al. (2021)	M and F UGs	M and F UGs	Within-subjects	Smiling	Smiling faces were more attractive than neutral faces on both static and dynamic conditions
Kościński (2013)	105 M and 115 F UGs	M and F (20-28yo)	Between-subjects	Talking (Introducing themselves) and smiling	Ratings for static and dynamic images were strongly correlated
Lander (2008)	24 M and 24 F (18-26yo)	60 M and 60 F (18-27yo)	Between-subjects	Talking (telling time)	Female static faces were rated as more attractive, whereas male moving faces were more attractive.
Penton-Voak and Chang (2008)	M and F UGs	M and F UGs	Within-subjects	Neutral (reading cue) Positive emotion (plans about a holiday)	Male faces more attractive when dynamic; Female faces more attractive when smiling
Rhodes et al. (2011)	M (17-35yo)	F (17-35yo)	Within-subjects	Counting aloud and smiling	No significant difference in attractiveness ratings
Roberts et al. (2009)	M and F UGs	M and F UGs	Mixed	Self-introduction and holiday	Ratings for static and dynamic images were strongly correlated
Rubenstein (2005)	F UGs	M and F UGs	Between-subjects	Reading cue	Low association between face types

Note: M = males, F = females, UG = undergraduate students

Furthermore, although it seems that motion in general does not influence attractiveness ratings, studies have demonstrated that emotion expression may modulate this. Another finding from Rubenstein (2005) was that perceived emotions on the moving face, but not static faces, affected attractiveness ratings, where although their dynamic stimuli (reading) depicted neutral expressions, faces perceived as having a more positive emotion were rated as more attractive compared to those perceived as displaying a negative emotion. It was therefore suggested that perhaps static and dynamic faces engage distinct mechanisms, which makes dynamic images more susceptible to other salient facial traits such as emotion expression.

Penton-Voak and Chang (2008) therefore directly manipulated emotion expression in their study, where young faces either depicted positive or negative emotions, on static and dynamic versions. This study yielded several interesting findings. First, they found an overall main effect of movement, where dynamic images were more attractive than static images, in contrast with Rhodes et al (2011) and Koscinski (2013)'s findings. Second, they found an interaction of face sex and expression, where expression did not influence ratings for male faces, whereas female happy faces were rated as more attractive than female neutral faces, partially supporting Rubenstein's findings (2005). Thirdly, an interaction of face sex and motion was found, where male faces were significantly more attractive when moving, but not females. From here, we could argue that emotion expression in dynamic faces affects attractiveness ratings of female faces more than males.

Similar findings were demonstrated recently by Bowdring et al. (2021), using G-theory analysis. They employed static and dynamic images depicting neutral and smiling expressions and found that overall, motion had little effect on attractiveness evaluations, however, smiling significantly increased the attractiveness ratings of faces. Interestingly, their findings also suggest that the differences in attractiveness perceptions are influenced more by the perceiver and the targets themselves rather than emotions and motion per se. This is out of the scope of the current study, however.

From these studies, we could argue that although motion in general does not affect perceptions of facial attractiveness in young samples, this could be modulated by other characteristics such as facial expression and gender, where young female faces with positive emotions were rated higher in attractiveness. However, such effects are yet to be explored in wider age range. As these studies used young target faces and observers, we are unable to generalise these findings to older adults, the group which we could argue are the most affected with regards to social evaluations and ageing appearance. The current study therefore aims to contribute to the literature by incorporating a wider age range for both target faces and observers, as well as using both male and female faces.

Motion and age perception

Very little research has investigated age estimations of dynamic faces. As far as we know, only one study has explored this avenue (Holland et al., 2019) and most age estimation studies have used static, neutral images – although more recently, studies using static images with different emotion expressions have also been conducted (Ebner et al., 2010; Voelkle et al., 2012). In general, studies which used static images on age estimations have found an own-age bias, that is, observers were more accurate at judging the age of their own-age group compared to those of others (Ebner et al., 2010; George & Hole, 1995; Moyse & Brédart, 2012; Rhodes & Anastasi, 2012; Voelkle et al., 2012). It was argued that such a phenomenon exists due to observers having more experience with people their own-age and therefore would be more likely to be accurate at estimating ages of people within their group (Voelkle et al., 2012). Furthermore, it was also found that ages of young faces tend to be overestimated compared to other groups, and that older faces are underestimated (Henss, 1991; Vestlund et al., 2009). Such findings could be due to the rapid facial shape changes that happen in childhood and early adulthood, whereas older adult faces changes are more subtle and occurs gradually (Enlow 1982, as cited in George & Hole, 1998; Mendelson & Wong, 2012). For the current study, however, as we are interested in the evaluation of the *same face* between conditions, own-age bias effects and age over/underestimations should have little influence on facial judgements.

Age estimations studies which used static images with different facial expressions have yielded inconsistent results – on the one hand, studies which used a wider age range have found faces with positive valence were rated younger (Voelkle et al., 2012; Ebner et al., 2010), while Ganel (2015) only used young sample and found that smiling faces resulted in age overestimations, i.e. smiling made young faces looked older. These contradictions could be due to the salience of markers of aging when smiling (Samson et al., 2010), e.g. wrinkles around the eyes and mouth areas, which made young faces look older, whereas this could be expected in older faces and therefore has reduced effect.

As mentioned, only one study has investigated facial age perception using dynamic faces. Holland et al. (2019) used dynamic clips formed from morphed facial images of the static FACES database (Ebner et al., 2010). Images from this database were from Caucasian models, with each model depicting different emotions. Participants were shown 15 random dynamic clips and were asked to identify the emotions being depicted and estimate the age of the person. Their findings broadly support previous findings on age estimations, that is, age of younger faces was overestimated, and ages of older faces were underestimated (Henss, 1991; Vestlund et al, 2009). Interestingly, their results contradicted previous findings from static images (Ebner et al., 2010; Voelkle et al., 2012) where they found that age estimations did not significantly differ between facial expressions, suggesting that dynamic faces elicit similar age estimations. However, Holland et al. (2019) was not able to directly compare how people viewed a static and dynamic image of the same person, therefore, we are unable to make an assumption as to whether observers would make similar age estimations from static and dynamic images. The current study therefore addresses this issue by presenting both static and dynamic images to participants.

Motion and health perception

To the best of our knowledge, no study has yet explored the effects of motion on health perception, however, studies using static images have shown that facial attractiveness and health are highly related with each other (Jones et al., 2018). This could mean that health perception would also follow the trend for attractiveness perception (although Foo et al. (2017) found that attractiveness ratings do not honestly signal individual's health). The current study will therefore be the first to investigate the effects of motion in facial health perception across a wide age range.

Study Overview

To the best of our knowledge, this is the first set of studies to investigate the influence of motion in the perception of three key facial traits (attractiveness, age, and health) using a wider age range of both male and female targets. The current study therefore aims to contribute to the literature by asking a wider age range of participants to rate both static and dynamic faces on age, attractiveness, and health. Studies using *static* faces of young people have shown that different emotions elicit differential perception of age, attractiveness, and health, where faces with perceived positive emotions were generally appraised to be more attractive (Bowdring et al., 2021; Ebner, 2008; Penton-Voak & Chang, 2008; Rubenstein, 2005) and healthier (Henderson et al., 2016; Jones et al., 2018), but not younger (Ganel, 2015); and faces with perceived negative emotions were appraised to be less attractive (Rubenstein, 2005) and less healthy (Henderson et al., 2016). This has a significant impact on older faces, as signs of ageing result to changes in facial features, e.g. drooping of the skin in the eye regions and sagging around the lips, which are in turn associated with negative emotions (Ebner, 2008; Hess et al., 2012).

Overall, neutral expression in static images appear to elicit the most accurate age estimations across ages and generally (with exception of Ganel, 2015) of happy faces were underestimated compared to emotions with negative valence (Ebner et al., 2010; Voelkle et al., 2012), hence we decided to use images depicting a neutral expression for the static condition. In terms of dynamic images, Holland et al. (2019) have demonstrated no significant differences in the age estimation of dynamic images across emotion expressions, therefore allowing us to use any emotion expression. With regards to attractiveness and health perception, we could argue that as older adults and faces with more apparent signs of aging are already perceived as less attractive (Ebner, 2008; Kwart et al., 2012) and less healthy (Fink & Matts, 2008; Samson et al., 2011) than their younger counterparts, adding a negative valence could potentially compound this effect. Therefore, for Study 1, it was deemed important to utilise dynamic images with positive affect.

Study 1 therefore was an exploratory experiment to investigate the role of motion in the perception of these three key dimensions using a static image with neutral expression, and a dynamic image depicting a happy expression. Study 2 was conducted to disentangle the influence of positivity effects on the results from Study 1; and finally, Study 3 aimed to investigate the influence of target sex in the perception of attractiveness and health in static smiling and neutral faces.

Study 1

Aims, Hypotheses and Predictions

In general, previous studies on attractiveness, age, and health have used static images, despite findings of motion having a differential effect on other facial characteristics, e.g. identity and emotion expression (Lander & Butcher, 2015) and thus it may be more ecologically valid to use dynamic images when investigating facial characteristics. Additionally, studies have mainly used young models and raters, therefore the effects of motion on these ratings for older faces are still unclear. The primary aim of Study 1 therefore was to investigate whether motion would influence ratings of facial attractiveness, health, and age across the age range.

In line with previous research, our hypotheses and predictions were as follows:

(1.1a) There will be a negative correlation between age estimation and perceptions of attractiveness and health;

(1.1b) there will be a positive correlation between perceptions of attractiveness and health; and

(1.1c) there will be a positive correlation of ratings between static and dynamic conditions for each trait, where as ratings for static faces increase, so do the ratings for dynamic faces.

(1.2a) There will be a main effect of Face Motion for each trait, where dynamic images would be rated more positively than static images;

(1.2b) there will be a main effect of the Face Age group on ratings, where older faces would be perceived less attractive and less healthy compared to the younger group; and

(1.2c) there will be an interaction between the face type and Face Age group on ratings, however, as there were inconsistencies in the literature, we are unable to predict the direction of the interaction.

Methods

Participants

Both tasks were open for participants aged 18 and over, with no exclusions for sex, and must have access to a computer and have no or corrected visual impairment to complete the experiments. Due to the nature of stimuli, only tasks completed by Caucasian participants were included in the analysis.

53 participants (36 F, 17 M) with age range of 20-66 years (M=35.92; SD=10.14) completed the study online using an anonymous link between 15 November 2019 to 02 April 2020. No reimbursements were given to participants for this study.

Statement of Ethics

This study was approved by the Swansea University Ethics Committee and followed the Declaration of Helsinki (Williams, 2015). All participants were informed of the study aims and procedures and were explicitly informed that their participation was completely voluntary and were able to withdraw their participation by not completing the study. Only completed tasks were included in the analysis.

Materials

The FACES database contains 171 static (Ebner et al., 2010) and dynamic (Holland et al., 2019) Caucasian faces divided into three age groups – young (19-31yo), middle (39-55yo), and old faces (69-80yo). Each unique individual was depicting six expressions – anger, disgust, fear, happiness, neutrality, and sadness – and two slightly different versions of each expression were created for static images (sets A and B). The dynamic stimuli were created by morphing the images together to create a moving image – for more information, see Holland et al. (2019).

A total of 60 individuals from FACES database were randomly chosen - 20 images per age group (10 males and 10 females). For static images, all target faces were depicting a neutral expression (from set A). For dynamic images, target faces started with a neutral face, and moved once to depict a smiling face. The motion lasted for approximately 2 seconds (for more detail on creation of dynamic stimuli, see Holland et al., 2019). The experiment was created and hosted in Gorilla (Anwyl-Irvine et al., 2020).

Trial blocks were set to be presented randomly to each participant, with each block consisting of all 60 images, either static or dynamic, and were rated on one trait at a time. Participants were asked to rate the images on both stimuli conditions on three traits – age, attractiveness, and health. Participants saw all six blocks, with a total of 360 trials.

Design

For Study 1, we were interested in the effect of *Face Motion* on social ratings. We therefore employed a mixed-model design using a by-image analysis. Our between-subjects variable was Age Group (three levels: young, middle, and old) and our within-subjects variable was Motion (two levels: static and dynamic). The dependent variables were the *averaged ratings* across participants for each target face on facial age, attractiveness, and health separately.

Procedure

Participants accessed the task on Gorilla (www.gorilla.sc) using an anonymous link. Participant information sheet and consent forms were presented before data collection. Demographic information such as age and gender were collected. Participants were also asked to rate their self-perceived facial age and facial attractiveness using a 7-point Likert scale. Following the questionnaires, participants were provided with instructions as to which trait was to be judged at the beginning of each block. Each image was presented individually, with a sliding scale at the bottom of the image to indicate their answer (18-100 for age, and 0-100 for health and attractiveness). Images remained on the screen until the participant gives an answer and clicked 'Continue' for the next trial. After each block, participants were given the option to have a break before continuing the study. Upon completion of all six blocks, participants were presented with the Debrief Form. The study took approximately 25-30 minutes to complete.

Data Cleaning

To inspect the quality of our data, we first looked at the mean and standard deviations (SDs) within each participant's responses separately for each trait. Participants with an SD of 0 were removed from the final analysis, as this means that the participants gave the same rating to each face. Following this procedure, one data point was removed from the Health dynamic task.

We then averaged the responses for each face type across individual targets. This created two variables for each trait: an average for static and dynamic conditions for the target face. Homogeneity tests revealed no significant differences of variance between each variable and therefore we were able to proceed with our mixed-model ANOVA.

An intra-class correlation analysis was conducted to check for inter-rater reliability using the *pingouin* package in Python (Vallat, 2018). We found an excellent agreement for static condition, where the average measure ICC2k was .897 (95% CI: 0.85-0.93), $F(59, 3068) = 25.001$, $p < .001$, and an excellent agreement for dynamic condition, where the average measure ICC2 was .885 (95% CI: 0.83-0.92), $F(59, 3068) = 20.784$, $p < .001$. To check for reliability of responses between static and dynamic images, we ran Pearson correlations and found high associations between the face types for attractiveness, $r(59) = .948$, $p < .001$; age estimates, $r(59) = .997$, $p < .001$, and health, $r(59) = .964$, $p < .001$.

Data Analysis

For the current study, we averaged the ratings for the target faces across all participants on static and dynamic conditions separately. We therefore have two composite scores for each target face on three key dimensions, e.g. attractiveness-static and attractiveness-dynamic scores, and so forth.

First, to determine whether there were associations between ratings of overall dynamic and static images, as well as ratings between the three traits, we employed a Pearson correlations test. To investigate the effect of the target age group and the type of stimuli on perceptions of age, attractiveness, and health, we ran separate 2 x 3 mixed Analysis of Variance on the averaged ratings for each trait using JASP (JASP Team, 2024). Significant interactions were further investigated using appropriate follow-up tests.

Results

Descriptive Statistics

Table 3.2 presents the means and SDs of ratings for each face type and age group. In sum, dynamic images received more positive ratings of attractiveness, health, and age estimates. Young faces also received the most positive ratings, followed by the Middle, and Older faces across all ratings.

Table 3.2

Means and Standard Deviations of Averaged Ratings of Attractiveness, Health, and Age Estimations for Static and Dynamic Face Types Across Age Groups.

Trait	Motion	Age Groups	Evaluations	
			Mean	SD
Age	Dynamic	Young	27.31	5.96
		Middle	51.26	7.55
		Old	70.77	7.09
	Static	Young	27.46	6.00
		Middle	50.38	7.74
		Old	70.95	7.21
Attractiveness	Dynamic	Young	55.65	24.24
		Middle	41.61	24.64
		Old	38.63	25.39
	Static	Young	52.06	22.85
		Middle	39.34	23.49
		Old	30.86	22.78
Health	Dynamic[a]	Young	72.05	18.12
		Middle	58.14	20.52
		Old	50.47	22.86
	Static	Young	69.98	19.11
		Middle	55.15	22.12
		Old	44.58	25.18

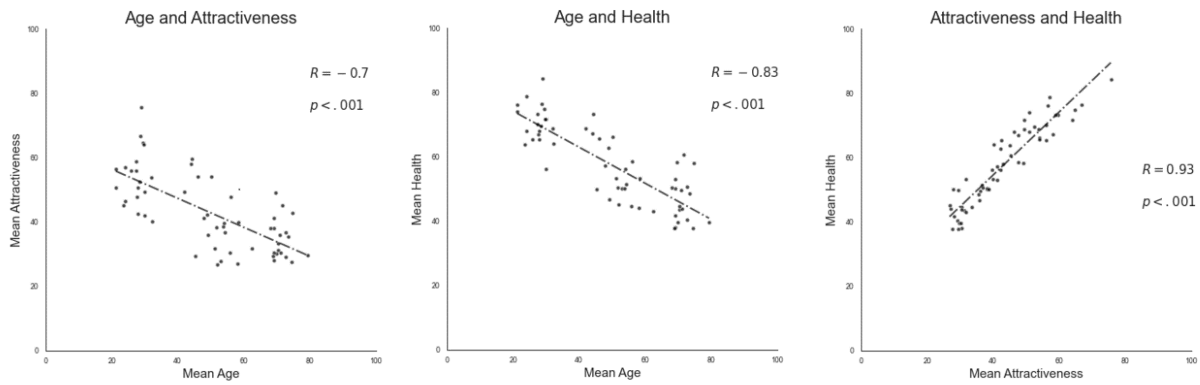
Note: Age ratings did not show a significant difference between dynamic and static images; attractiveness and health ratings were higher for dynamic images. Young target faces received the highest ratings of attractiveness and health, followed by the Middle and Old target faces for both face types. [a]Averaged ratings from n=52.

Inferential Statistics

Relationships between the three traits. We ran a Pearson correlation test to investigate how the three key traits were related to each other (Figure 3.1). Overall, we supported previous research where we found that overall age estimates were negatively correlated with attractiveness, $r(59) = -.700$, $p < .001$, and health evaluations, $r(59) = -.829$, $p < .001$, where the older the face looked, the less attractive and less healthy they were perceived to be; and attractiveness and health ratings were positively correlated, $r(59) = .932$, $p < .001$, where as attractiveness ratings increase, so do health ratings.

Figure 3.1

Relationships Between the Three Key Facial Traits.



Note: Age estimations were negatively correlated with attractiveness and health, while attractiveness and health were positively correlated with each other.

We then ran a series of Analysis of Variance to investigate the influence of motion on estimations of facial age, attractiveness and health (Figure 3.2).

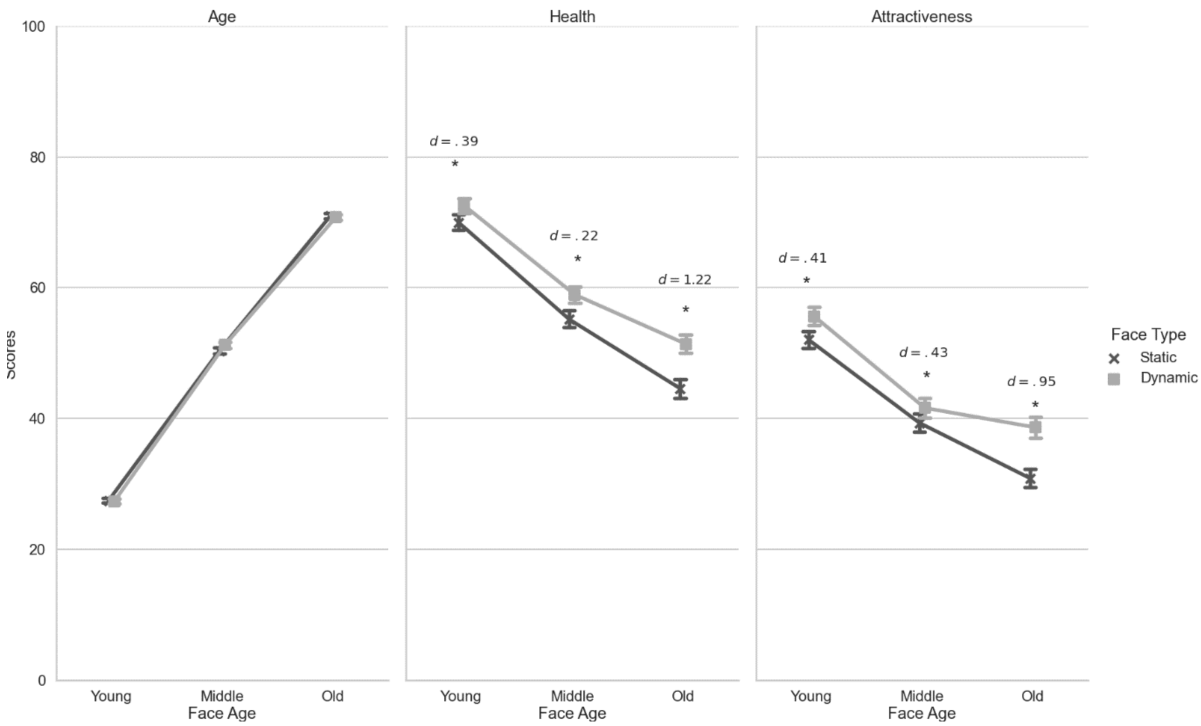
Age. We found that, in contrast with the other two traits, age estimates did not yield a significant effect of motion, $F(1,57) = .976$, $p = .327$, $\eta p^2 = .017$; but we found a significant effect of age group as expected, $F(2,57) = 644.863$, $p < .001$, $\eta p^2 = .958$, where Young faces were rated as the youngest, followed by the Middle and Old faces; and we also found a small, but significant interaction between the two variables, $F(2,57) = 3.554$, $p = .035$, $\eta p^2 = .111$, however, following post-hoc comparisons with Bonferroni correction, we found no significant difference between the face types within each age groups, $p > .05$.

Attractiveness. We found a significant effect of Face Motion, $F(1,57) = 103.289$, $p < .001$, $\eta p^2 = .644$, where dynamic images were rated as more attractive; a significant effect of age group, $F(2, 57) = 25.248$, $p < .001$, $\eta p^2 =.470$, where the Young faces had the highest ratings of attractiveness, followed by the Middle, and the Old faces; and a significant interaction between the two variables, $F(2,57) = 13.726$, $p <.001$, $\eta p^2 = .325$, where post-hoc comparisons using Bonferroni correction revealed that Young and Old faces received significantly higher ratings of attractiveness ($p < .001$) when dynamic, but not the Middle faces ($p = .072$).

Health. Similar to attractiveness, we found a significant effect of face type, $F(1,57) = 70.678$, $p <.001$, $\eta p^2 = .554$, where dynamic images were rated as healthier; a significant effect of age group, $F(2,57) = 51.588$, $p < .001$, $\eta p^2 = .644$, where Young faces were the most healthy, followed by the Middle, and Old faces; and a significant interaction between the two variables, $F(2,57) = 7.104$, $p = .002$, $\eta p^2 = .200$, where post-hoc comparisons using Bonferroni correction showed that the Middle ($p = .003$) and Old faces ($p < .001$) looked healthier when dynamic, but not Young faces ($p = .121$a).

Figure 3.2

Interactions Between Face Age Group and Face Type for Each Trait.



Note: Rating scores clustered for each trait. Overall, dynamic images received higher ratings than static images. Significant interactions found between face type and age group for all rating types, however no significant differences were found between face types within age groups for age estimation following post-hoc analysis. Error bars represent standard error. Cohen's *d* scores show the effect sizes of each significant comparison between static and dynamic stimuli within each age group for each trait.

Discussion

We sought to determine whether there would be an effect of motion in evaluations of social traits across target age groups. To do this, we presented a wide age range of participants with both static and dynamic faces and asked them to rate these faces on three key dimensions – attractiveness, health, and age estimations. To the best of our knowledge, this is the first study to incorporate dynamic images in perceptions of these three traits using a wider age range of target faces and participants.

The current findings support several hypotheses. First, our findings support hypothesis 1.1a that there will be a negative relationship between age estimation and perceptions of attractiveness and health, where we found that older faces were rated less attractive and less healthy than their younger counterparts. This supported the findings from Gunn et al. (2013) where they found that individuals who were perceived two years younger had lower risk of cardiovascular diseases, and previous findings where they found that perceived age was negatively correlated with attractiveness (Ebner 2008; Kwart et al., 2012).

Additionally, our findings also support Samson et al. (2010) which showed that when signs of facial aging are more salient, i.e., smiling could highlight signs of aging such as wrinkling around the eyes and mouth, perceptions of attractiveness decrease. The current study did not employ manipulation of the faces used in the tasks, in contrast to Samson et al. (2010), however, the nature of the database meant that there are distinct differences in the clinical signs of aging for each age group, i.e. the Old group would have had the most amount of wrinkles, followed by the Middle, and the Young faces. These findings indicate that in unmodified images, signs of aging have a negative impact on appraisals of attractiveness and health.

Our findings also support hypothesis 1.1b, where we found a positive correlation between attractiveness and health, whereas ratings for attractiveness increases, so do the ratings for facial health, supporting previous studies (Fink et al., 2012; Jones et al., 2018; Matts & et al., 2007). Our findings further support those from Roberts et al (2009) where we found a positive relationship on evaluations of attractiveness between static and dynamic conditions. Additionally, this presents a novel finding for the positive associations between static and dynamic images for age estimations and health perception. Studies from Ebner et al (2010) and Holland et al. (2019) used the same database for facial age estimations, however, data was collected separately and the differences in data collection method made it difficult to get an absolute comparison of age estimates for the images from the FACES database. The current findings further provide a good support for the use of FACES database in both static and dynamic contexts, for age estimation studies.

Furthermore, our findings from the mixed ANOVA partially supported hypothesis 1.2a that there will be a main effect of motion on ratings, where we found that dynamic images were rated higher on attractiveness and health, but *not* on age estimations. This was expected for ratings of attractiveness and health, as previous studies have found that moving, smiling faces were perceived to be more attractive (Bowdring et al., 2021; Penton-Voak & Chang, 2008) and healthier (Jones et al., 2018) than neutral faces. For age estimation, however, we expected that age estimations would also be higher in dynamic images, particularly in the Middle and Older group due to the increased salience of clinical signs of aging, especially because of the wrinkling around the eyes and mouth when smiling. However, as found by Holland et al. (2019), the smiling movement did not influence age estimation. Our findings contradicted that of Ganel (2015) where we did not find a significant age difference between static and dynamic faces in our Young and Old groups. We did, however, find that middle-aged faces were rated older when smiling. One explanation for this finding could be that when static, middle-aged faces may pass as younger than they are because the signs of aging are less visible, and would then be highlighted with a smiling motion, where they ‘gain’ age estimates. Young and Old faces may not be as affected as signs of aging for the Young faces would be less pronounced in both contexts, i.e. a floor effect, and Old faces have more texture and thus were already appraised as old, i.e. ceiling effect.

Furthermore, our findings supported hypothesis 1.2b that there will be a main effect of target Face Age group for all traits, where Young faces received the lowest age estimates, followed by Middle and Old, further validating the FACES database for age estimation studies (Ebner et al., 2010). Furthermore, Young faces received the highest ratings for attractiveness and health, followed by the Middle and Old faces. We also found partial support for hypothesis 1.2c, where we found significant interactions between target Face Age group and stimuli type for ratings of attractiveness and health, where, in line with previous studies, we found that older faces received significantly lower ratings of social perception in both face types, where although dynamic faces in general were rated as more attractive and healthier, this boost is higher for Old faces compared to Young and Middle faces.

Our findings suggest that when faces are smiling, older adults are perceived slightly more attractive and healthier compared to when they are static and depicting neutral expressions. This could have implications on stimuli selection for experiments investigating facial attractiveness and health on older adults in the future, as well as perceiving effects of anti-aging treatments. However, although these results provide novel insights to social perceptions of older adults, we are not able to conclude whether our findings for evaluations on facial attractiveness and health were due to motion itself, or the fact that faces were depicting a happy expression. We therefore decided to investigate this further by using smiling faces for both static and dynamic contexts.

Study 2

Aims, Hypotheses, and Implications

As the stimuli used in Study 1 have used confounding variables, e.g., using a different motion and emotion, we were not able to identify fully where the differences found in the mian effects were coming from. For Study 2, we therefore aimed to disentangle the role of positive emotion and facial motion on ratings of attractiveness and health. Our hypotheses follow the findings from Study 1:

(2a) there will be a main effect of Face Motion on ratings of attractiveness and health, where dynamic faces would be rated more positively;

(2b) there will be a main effect of target Face Age group on ratings of attractiveness and health, where Young faces would be rated most positively;

(2c) there will be an interaction between target Face Age group and motion, where Middle and Old faces would be receiving a boost of positive ratings when dynamic; and

(2d) overall ratings of attractiveness and health will be positively related with each other.

With regards to interpretation of our findings, if results from Study 2 support our hypotheses that motion has an effect on ratings, this means that the ratings from Study 1 were due to influence of *motion* itself, otherwise, the findings from Study 1 (and hence Study 2) were merely due to positivity effects, that is, smiling in itself makes people look more attractive and healthier.

Methods

Participants

Inclusion criteria were similar to that of Study 1. 54 participants (45 F, 8 M, 1 Other) with age range of 19-67 years (M=30.24; STD=10.02) completed the study online using an anonymous link between 13 July and 05 August 2020. Participants were not given reimbursements for this study.

Design

For Study 2, we wanted to disentangle whether the significant effects found from Study 2 were due to motion itself, or simply due to positivity effects. As we did not see a significant effect of motion on age estimates, we dropped the Age task for this study. Following the design from Study 1, we employed a mixed-model design using a by-image analysis. Our between-subjects variable was the *Age Group* (three levels: young, middle, and old) and our within-subjects variable was the *Face Motion* (two levels: static and dynamic). The dependent variables were the *averaged ratings* on attractiveness and health across all participants for each target face.

Materials

The same target individuals were used in this study as Study 1, but with a happy expression in the static condition (set A) instead of a neutral expression. For this study, participants were asked to rate the images only on facial attractiveness and health, yielding four blocks with 60 trials each, with a total of 240 trials. This experiment was also set up on Gorilla (Anwyl-Irvine et al., 2020).

Procedure

The procedure was similar to Study 1, where participants accessed the study using an anonymous link on Gorilla (www.gorilla.sc) and were presented with the Participant

Information Sheet and Consent Forms. Participants also provided the same measures as in Study 1 prior to completing the task. Study blocks were presented in random order per participant to avoid order effects. Prior to each block, participants were informed as to which trait would be rated – for this part of the study we asked participants to rate faces on attractiveness and health. Following the task, participants were provided with the Debrief Form.

Data Cleaning

Similar procedures for data cleaning was conducted for our data. Following the procedure, we removed one set of participant responses from the Health task (dynamic). The remaining responses were averaged across each face separately for the static and dynamic conditions. Homogeneity tests revealed that the Health ratings from static faces were significantly different from each other, while Attractiveness from both motion types were of equal variances. However, as we have equal amounts of stimuli for each condition, we were able to proceed with our parametric analysis.

We ran an intra-class correlation analysis to check for inter-rater reliability using *pingouin* package in Python (Vallat, 2018). Similar to Study 1, we found an excellent agreement for static condition, where the average measure ICC2 was .974 (95% CI: 0.96-.98), $F(59, 3127) = 66.973$, $p < .001$, and we also found an excellent agreement for dynamic condition, where the average measure ICC2 was .970 (95% CI: 0.96-0.98), $F(59, 3127) = 59.062$, $p < .001$. To check the reliability of evaluations between the face types, we ran Pearson's correlations and found strong, positive relationships between static and dynamic face conditions on attractiveness, $r(59) = .982$, $p <.001$, and health perceptions, $r(59) = .972$, $p < .001$, similar to that of Study 1.

Data analysis

Data analysis for this study followed that of Study 1, where scores from all participants were averaged for each target face per condition, creating a composite score for static and dynamic conditions for each trait. To investigate whether motion has an effect on ratings of attractiveness and health, we ran separate 2 x 3 mixed ANOVAs for each trait. Complementary Bayes factor analyses were also performed to determine whether the effects were under null hypothesis, given the data.

Results

Descriptive Statistics

Table 3.3 presents the means and standard deviations of attractiveness and health ratings for each face type and age group. In sum, we found a similar trend as Study 1, where overall, dynamic images were rated more positively compared to static images, and that Young faces received the most positive ratings, followed by Middle and Old faces.

Table 3.3

Means and SDs of attractiveness and health ratings for static and dynamic conditions grouped by target Face Age.

Trait	Motion	Age Groups	Evaluations	
			Mean	SDs
Attractiveness	Dynamic	Young	60.06	22.27
		Middle	44.04	22.23
		Old	39.81	20.94
	Static	Young	57.77	22.23
		Middle	42.50	23.07
		Old	37.57	21.25
Health	Dynamic	Young	74.20	18.56
		Middle	59.65	20.55
		Old	51.25	21.76
	Static	Young	74.28	18.52
		Middle	59.83	21.34
		Old	50.22	21.42

Note: a. Data for Attractiveness is from N = 54. b. Data for Health (static) is from N = 54, and for Health (static) is from N = 53. Overall, dynamic images received higher attractiveness ratings, whereas both face types received similar health ratings within the age groups.

Inferential Statistics

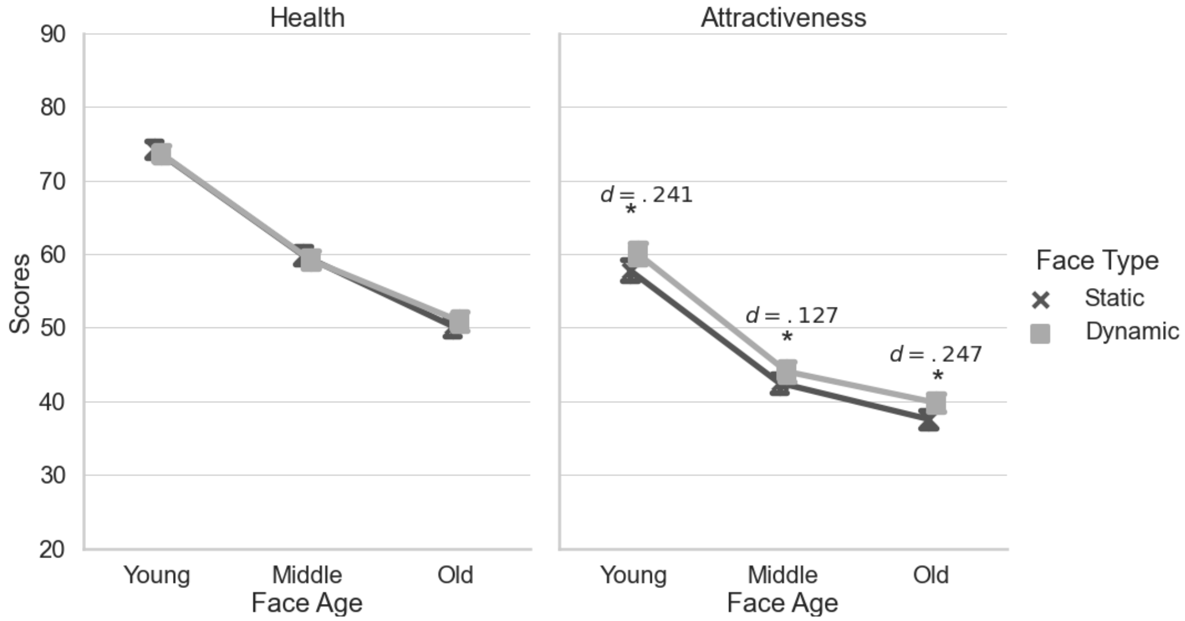
Overall correlation. In line with previous studies, we found a strong, positive correlation between attractiveness and health ratings, r(59) = .937, $p < .001$.

Attractiveness. Similar to Study 1, attractiveness ratings showed a significant effect of Face Motion, $F(1,57) = 35.355$, $p < .001$, $\eta p^2 = .383$, where dynamic images (M= 47.969, SD = 13.136) were more attractive than static images (M = 45.948, SD = 13.614); and a significant effect of age group, $F(2,57) = 21.526$, $p < .001$, $\eta p^2 = .430$, where Young faces (M=58.917, SD = 9.378) were the most attractive, followed by the Middle (M = 43.269, SD = 12.059) and Old faces (M = 38.690, SD = 8.955); however, we found no significant interaction between the two variables, $F(2,57) = .511$, $p = .603$, $\eta p^2 = .018$ (Figure 2). The complementary Bayes factor analysis showed that this is more likely under the null hypothesis ($BF_{01} = 4.743$).

Health. In contrast to findings from Study 1, we found no significant effect of Face Motion, $F(1,57) = .363$, $p = .549$, $\eta p^2 = .006$. The complementary Bayes factor analysis showed that this is more likely under the null hypothesis ($BF_{01} = 3.760\text{x}10^{+8}$). We found a significant effect of Age Group, $F(2,57) = 37.435$, $p < .001$, $\eta p^2 = .568$, where Young faces were the healthiest (M = 74.238, SD = 5.803), followed by the Middle (M = 59.742, SD = 9.458) and Old faces (M = 50.738, SD = 10.109); and no significant interaction between the two variables, $F(2,57) = .833$, $p = .440$, $\eta p^2 = .028$ (Figure 3.3). The complementary Bayes factor analysis showed that this is more likely under the null hypothesis ($BF_{01} = 17.781$). This suggests that the effect observed in Study 1 could be due to the difference in emotion, rather than motion itself.

Figure 3.3

Interaction Between Target Face Age and Motion on Evaluations of Facial Attractiveness and Health for Smiling Static and Dynamic Faces.



Note: Motion had a differential effect on attractiveness and health perceptions of smiling faces, where dynamic faces were rated as more attractive than static faces, whereas no such effect was found on health ratings. As expected, Young faces were rated as most attractive and healthiest, followed by Middle and Old faces. No interaction between Face Age and face type were found for both traits.

Discussion

Due to the difference in emotions used for static (neutral) and dynamic (smiling) faces from Study 1, we sought to further test whether the significant effect of motion found in the perception of facial attractiveness and health were due to motion itself or rather simply due to positivity effect, i.e. faces with perceived positive emotions were more likely to be more attractive and healthier. Study 2 therefore employed smiling faces for both static and dynamic conditions – if the results from the current study were similar to that of Study 1, that is, if we find significant effect of motion, we could assume that motion *does* have an effect on perception of social traits, given that the same expression was used. Otherwise, no significant effect of motion in the current study means that the significant effect of motion found in Study 1 were simply due to positivity effect.

Findings from the mixed ANOVA demonstrate interesting dissociations in evaluations of attractiveness and health. On the one hand, attractiveness ratings supported hypothesis 2.1a where we found a significant effect of motion, generally supporting findings from Study 1. Post-hoc analysis showed that Young and Old moving faces were more attractive than static faces when depicting a happy face. As expected, we found a significant effect of target Age Group, where Young faces were rated more attractive than Middle and Old faces. However, unlike Study 1, we found no significant interaction between Face Motion and Age Group. The additional Bayes factor analysis also suggested that given the data, there was likely no interaction between the two variables. These findings suggest that overall, the difference in ratings were due to motion rather than the positive affect depicted by the dynamic stimuli. However, in Study 1, Old faces had a stronger 'boost' in ratings when smiling compared to static, neutral expression. The lack of interaction here in Study 2, as well as the Middle faces not having a significant difference in ratings between the two conditions, suggests that perhaps for older adults, a positive expression is inherently more attractive than a neutral, static expression.

In contrast to Study 1, however, health ratings from Study 2 did not support hypothesis 2.1a, where we found no significant effect of motion. The complementary Bayes factor analysis suggested that there is a likely null effect of motion, given the data, implying that results from Study 1 were due to positivity effects, that is, the moving face looks healthier because of the positive affect it displays. As expected, however, we found a significant effect of Age Group, where Young faces were rated healthier than Middle and Old faces. This could reflect the general trend on individual health, e.g. as we get older, we are more likely to get ill and recovery times are slower. In contrast to Study 1, no significant interaction was found between target Age Group and Face Motion. This further suggests that the 'boost' found in Study 1 was also due to positivity effects.

Intriguingly, these contrasting trends were found despite a strong, positive correlation between attractiveness and health ratings. Our findings therefore suggest that although attractiveness and health perceptions are closely related, these evaluations are affected differently by target emotion and motion.

Study 3

So far, we have demonstrated that *motion* (Study 1) and *emotion* (Study 2) have a differential effect in age estimations and perceptions of attractiveness and health across the age range, and that static and dynamic images elicit reliable responses from observers. Furthermore, we found that overall, dynamic smiling faces elicited more positive evaluations of attractiveness and health across the age range. However, the literature on attractiveness research cited another source of variation in perception of attractiveness for *young* faces – the target face's sex, that is, young *female* faces were rated as more attractive when smiling compared to neutral expression (Penton-Voak & Chang, 2008), whereas young *male* faces were not affected by the positive expression (Rhodes et al., 2011). Note that data from Study 2 provides some support for this, where attractiveness ratings were influenced by motion more than the emotion displayed by the face. No study has investigated whether this was the case for health perceptions, however, despite the close relationship between the two.

In this part of the chapter, we aimed to investigate whether there are differences in the perception of attractiveness and health of male and female faces across a wider age range of target stimuli using the data from Studies 1 and 2. As we did not have a corresponding data for dynamic neutral condition, we included only the data from static neutral (Study 1) and static smiling (Study 2) conditions for Study 3. Using the faces as a unit of analysis, a strategy previously used in face perception studies (Jones et al., 2018; Morrison et al., 2013; Voegeli et al., 2021), we explored whether gender differences in attractiveness and health ratings would emerge as a function of the target Age Group and target Emotion.

Aims and hypotheses

For this section, therefore, our hypotheses were as follows:

(3a) there will be a significant effect of Age Group, where Young faces would receive the most positive evaluations;

(3b) there will be a significant effect of Sex in the evaluations of attractiveness and health in dynamic, smiling faces, where in general, female faces would receive more positive evaluations;

(3c) there will be a significant effect of Emotion, where smiling faces would have higher ratings of evaluations compared to neutral expression;

(3d) there will be a significant interaction between Sex and Age Group, where gender differences would emerge in the Young group and Middle group, but not in Old faces;

(3e) there will be a significant interaction between Sex and Emotion, where females would receive higher evaluations in the smiling faces, where male evaluations would not be affected;

(3f) there will be a significant interaction between Emotion and Age Group; given the data on positivity effects, we expect that older faces will receive a boost of positive ratings for attractiveness and health compared to young and middle faces; and

(3g) there will be a three-way interaction between Age Group, Sex, and Emotion.

Methods

Participants

For this study, responses for attractiveness and health tasks were collated from two previous studies. In total, we have 107 participants (79 F, 24 M, 1 Other), with age range 19-67 years old (M = 33.029, STD = 10.461). Following the data cleaning procedures from Studies 1 and 2, no data was excluded for this part of the study.

Design

This study explores the gender differences on evaluations of attractiveness and health in two different face types, across a wider age range. With the target faces as our unit of measurement, our between measures variables were target *Face Sex* (two-levels: males and females) and *Age Group* (three-levels: young, middle, old). Our within measures variable was *Emotion* (two-levels: happy and neutral). The dependent variable is the averaged rating of attractiveness and health across all participants for each target face.

Data Analysis

Responses were averaged for each target face, separately for attractiveness and health evaluations. Assumptions of sample homogeneity were met for both traits. We then ran a mixed ANOVA to investigate the effects of target emotion, sex, and age group on evaluations of facial traits. Complementary Bayes factor analyses were also performed to determine whether the effects were under null hypothesis, given the data.

Results

Descriptive Statistics

Table 3.4 presents the means and SDs for attractiveness and health ratings of for each face category. In sum, there is a general trend of females having more positive evaluations than men, and young faces having the highest ratings, followed by middle, and old groups.

Table 3.4

Means and SDs of Attractiveness and Health Evaluations by Emotion, Face Sex, and Face Age.

Trait	Emotion	Sex	Group	Mean	SD
Attractiveness	Neutral	Female	Young	55.60	10.67
			Middle	45.81	10.08
			Old	30.66	7.42
		Male	Young	48.51	6.73
			Middle	32.87	8.33
			Old	31.06	7.03
	Smiling	Female	Young	60.33	11.63
			Middle	48.70	10.12
			Old	40.39	8.24
		Male	Young	55.20	7.64
			Middle	36.30	11.88
			Old	34.76	9.97
Health	Neutral	Female	Young	69.45	8.86
			Middle	58.55	8.85
			Old	44.42	7.19
		Male	Young	70.51	4.66
			Middle	51.76	8.61
			Old	44.76	8.27
	Smiling	Female	Young	74.09	7.22
			Middle	63.52	9.46
			Old	52.82	10.12
		Male	Young	74.46	4.71
			Middle	56.15	9.59
			Old	47.63	12.50

Inferential Statistics

Attractiveness. We found a significant main effect of Face Emotion, $F(1, 54) = 65.296$, $p < .001$, $\eta p^2 = .547$, where smiling faces were more attractive (M=45.948, SD=13.614) compared to neutral faces (M=40.752, SD= 12.726). We also found a significant main effect of Face Sex, $F(1, 54) = 9.498$, $p = .003$, $\eta p^2 = .150$, where on average, females were rated more attractive (M=46.917, SD=13.080) than males (M=39.782, SD=11.891), and a significant main effect of Age Group, $F(2, 54) = 27.742$, $p < .001$, $\eta p^2 = .507$, where Young faces had the highest ratings (M=54.913, SD=9.404), followed by the Middle (M=40.920, SD=11.564), and the Old faces (M=34.216, SD=7.761).

Our analysis did not yield any significant two-way interactions: Emotion and Sex, $F(1,54) = .835$, $p = .365$, $\eta p^2 = .015$; Emotion and Age Group, $F(2, 54) = 2.708$, $p = .076$, $\eta p^2 = .091$, and sex and group, $F(2, 54)=1.622$, $p = .207$, $\eta p^2 =.057$. The complementary Bayes factor analysis showed that this is more likely under the null hypothesis ($BF_{01} = 4.97x10^{+6}$, $BF_{01} = 13.858$, and $BF_{01} =1.810x10^{+7}$) .

We did, however, found a small, but significant three-way interactions between Emotion, Sex, and Age Group, $F(2, 54) = 3.660$, $p = .032$, $\eta p^2 = .119$ (see Figure 3.4). Post-hoc comparisons with Bonferroni corrections revealed that for females, attractiveness ratings between neutral and smiling faces for Young and Middle groups were similar to each other, but for the Old group, smiling faces were significantly more attractive than neutral faces. For males, on the other hand, Middle and Old groups were similar in attractiveness ratings, but Young, smiling faces were significantly more attractive than Young, neutral faces.

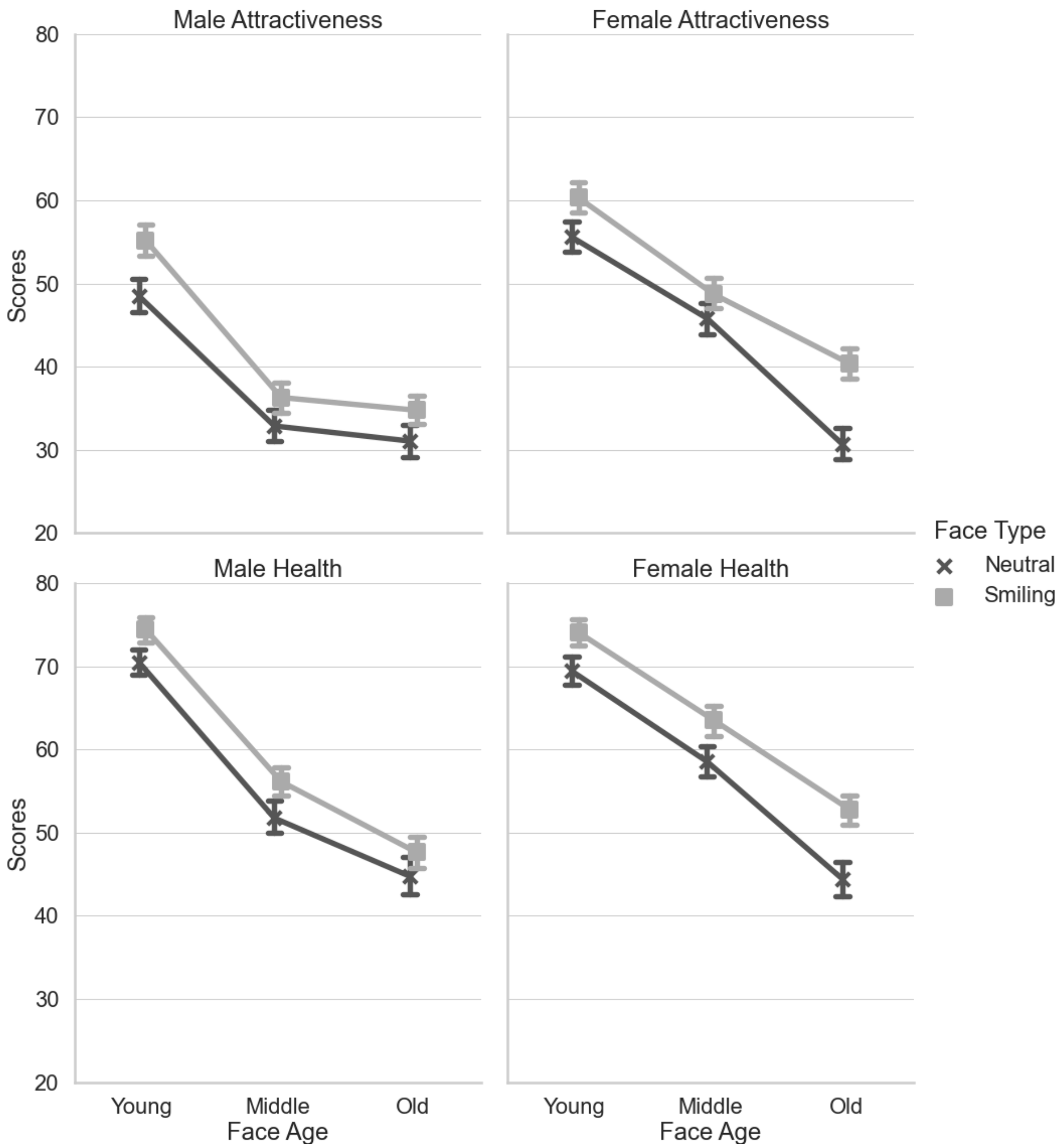
Health. For health evaluations, we found a significant main effect of Emotion, $F(1, 54) = 51.899$, $p < .001$, $\eta p^2 = .490$, where smiling faces (M=61.444, SD=13.592) were rated healthier than neutral faces (M=56.572, SD=13.084); a significant main effect of Age Group,

$F(2, 54) = 46.109$, $p < .001$, $\eta p^2 = .631$, where Young faces ($M = 72.130$, $SD = 6.055$) were rated healthier than Middle ($M = 57.492$, $SD = 9.346$) and Old faces ($M = 47.404$, $SD = 9.023$). We did not find a significant main effect of Sex, $F(1,54) = 1.921$, $p < .171$, $\eta p^2 = .034$. The complementary Bayes factor analysis showed that this is more likely to be under the null hypothesis ($BF_{01} = 2.005 \times 10^{+16}$).

Our analysis did not yield any significant two–way interactions (see Figure 3), Emotion and Sex, $F(1,54) = 2.270$, $p = .105$, $\eta p^2 = .048$; Emotion and Age Group, $F(2,54) = .623$, $p = .540$, $\eta p^2 = .052$; Sex and Age Group, $F(2, 54) = 1.147$, $p = .325$, $\eta p^2 = .041$. Lastly, we found no significant three-way interactions between Emotion, Sex, and Age Group, $F(2,54) = 1.482$, $p = .236$, $\eta p^2 = .052$. The complementary Bayes factor analysis showed that this is more likely to be under the null hypothesis ($BF_{01} = 20.284$). These findings imply that facial health is appraised similarly for both males and females regardless of emotion depicted, in that as one gets older, the less healthy they appear to be.

Figure 3.4

Interactions of Face Sex, Face Age, and Emotion for Attractiveness and Health Ratings.



Note: Ratings were between 0 (Not attractive/healthy at all) and 100 (Very attractive/healthy).

In summary, across three studies (Table 3.5), we found:

1. In Study 1, using static neutral and dynamic smiling faces of different age groups, we found that dynamic smiling faces were rated as more attractive and healthier than static neutral faces. No differences were found for age estimation.
2. In Study 2, we controlled for emotion and used smiling images for dynamic and static conditions for attractiveness and health tasks using different age groups. It was found that dynamic images were rated as more attractive than static images, whereas no differences were found for health perceptions.
3. In Study 3, we looked at how the gender, emotion and age group of the target face influenced perceptions of attractiveness and health using static images. It was found that Young and Middle aged female faces were rated as more attractive than their male counterparts, and that Old faces did not differ in their attractiveness ratings. For health perceptions, smiling faces were rated as healthier than neutral faces, and no significant interactions were found.

Table 3.5.

Summary of Results using Static and Dynamic Images across Perceptions of Facial Attractiveness, Health, and Age.

Study	Study 1			Study 2		Study 3	
Task	Age	Attractiveness	Health	Attractiveness	Health	Attractiveness	Health
Age Group	/	/	/	/	/	/	/
Motion	x	/	/	/	x		
Sex						/	x
Emotion						/	/
Age * Emotion	x	/	/	x	x	x	x
Age * Sex						x	x
Sex * Emotion						x	x
Sex * Age Group * Emotion						x	x

Note: / denotes significant effect. X denotes no significant effect. Highlighted cells were not investigated for that study.

Discussion

This study aimed to investigate whether there would be gender differences on the evaluations of attractiveness and health of neutral and smiling static faces as a function of age. We used a wider age range of male and female target faces and asked participants to judge how attractive and healthy they appear to be.

First, we discuss the findings on attractiveness. As expected, the results support our hypothesis 3a that there will be a main effect of Age Group, where, indeed, Young faces were given the highest attractiveness scores, followed by Middle, and Old faces. This supports previous studies which found that the older the face looks, the less attractive they are perceived to be (Kwart et al., 2012). The current findings further support hypothesis 3b that there will be a main effect of face sex, where we found that female faces had higher attractiveness evaluations than males, and we also found support for hypothesis 3c that there will be main effect of face type, where we found that smiling faces were more attractive than neutral faces. These findings support the conclusions of Penton-Voak and Chang (2008), where they found that overall, female faces, and smiling faces were more attractive.

Interestingly, we only found support for a three-way interaction, where for Females, we found that Young and Middle-aged females had similar attractiveness ratings regardless of emotion depicted, whereas Old aged females had received a ‘boost’ of attractiveness when smiling compared to neutral face. This finding supports the evolutionary psychology perspective which posits that female attractiveness is linked to their reproductive value, where Young and Middle-aged women are more likely to be capable of child-bearing compared to Old faces (Maestriperi et al., 2014). Additionally, the boost of attractiveness in the smiling Old faces highlights the role of positive emotion in eliciting positive evaluations in older population. More practically, this particular finding highlights the *decline* of female attractiveness in older age, and therefore provides an explanation as to why middle-aged females are the highest consumers of appearance enhancing procedures, perhaps the idea is to pre-emptively ‘delay’ the decline in perception of attractiveness (Clarke, 2007).

On the other hand, Middle and Old-aged male faces had similar attractiveness ratings regardless of emotions, whereas Young male faces were more attractive when smiling compared to neutral faces. This is an unexpected result, as we expected that only female faces would benefit from the smiling effect, contradicting the results from Penton-Voak and Chang (2008), where they found that such effect only applied to female faces. Interestingly, we only see this boost in the Old female faces, that is Young and Middle female faces received similar attractiveness ratings for smiling and neutral, but smiling Old female faces received a boost of attractiveness ratings for smiling. The finding that Middle and Old-aged male faces were similar in attractiveness ratings is also in line with the evolutionary perspective, where it was argued that the male reproductive value is closely linked to ability to support and perhaps more experience (Buss, 1989) rather than the facial attractiveness per se. This is in contrast with female reproductive value, which is closely tied to female fertility in younger years.

As for health evaluations, we again supported our hypothesis that there will be a main effect of Age Group, where Young faces were rated to be the healthiest, followed by the Middle and Old group. Furthermore, we also supported the hypothesis that there will be a main effect of face emotion, where smiling faces were rated to be healthier than neutral faces, supporting the findings of Jones et al. (2018). In contrast to the findings from attractiveness ratings, we did not find any significant two-way or three-way interactions between the variables, given the data. This is interesting, as attractiveness and health ratings have been found to be associated with each other and therefore we would expect these two traits to have similar trends. This therefore implies that despite the close relationship between these two variables, perceivers use different facial cues when judging different social traits.

Our findings show that attractiveness and health ratings follow the same trend as a function of age and emotion, where younger and smiling faces received more positive evaluations, however, the gender bias was only observed on attractiveness ratings.

Overall discussion

In sum, the current series of studies explored the role of various elements in perceptions of attractiveness, health, and age estimations. From Study 1, we found that *motion* had a differential effect on the three key dimensions, where dynamic faces were perceived to be more attractive and healthier, but not older than static images. This finding is the first to demonstrate that facial age estimations are not affected by motion using a wider range of facial targets, and thus suggesting that in age estimation studies, the use of static images is sufficient. Interestingly, this contradicts previous studies which found that increased salience of skin texture in manipulated static images increases age estimation (Samson et al., 2010; Samson et al., 2011). However, this supports the finding that by George and Hole (2000) where they found that age estimation is consistent despite manipulations of spatial and surface cues. However, due to some limitations in the stimuli used in this study, we decided to conduct further experiments to disentangle the role of motion further for attractiveness and health tasks.

Study 2 highlighted the differential role of motion on attractiveness and health perceptions, where we found that when emotion was controlled for (smiling), despite the strong correlation between the two traits, only attractiveness ratings were affected by motion. It is important to note, however, that the interaction between face types in Study 1 was not replicated for attractiveness, therefore, although dynamic faces were more attractive, emotion expression also significantly contributes to ratings as demonstrated by Bowdring et al (2020).

Study 3 highlighted the role of gender on attractiveness ratings, but not on health. We found that female faces were rated as more attractive than male faces, but not healthier.

Overall, the current study has demonstrated that perceptions of the three key facial traits were affected differently by motion, emotion, and facial sex. Facial age is reliably estimated regardless of motion (and to some extent, emotion), whereas facial attractiveness and health appear to be influenced by emotion expression. Critically, our findings also demonstrate that despite the high correlation of ratings for attractiveness and health, observers may utilise different cues when making a judgement.

One major limitation of the current set of studies is that we have a female dominated sample, and it had been previously shown that own-gender bias could occur, where perceivers give more positive ratings to same-sex faces, and therefore could explain the gender-bias found in the attractiveness data. Additionally, Penton-Voak and Chang (2008) also have found that male faces were more attractive when moving, regardless of emotion and here we only have static faces. It is therefore imperative to get a more gender-balanced sample, and incorporate both emotions (smiling and neutral) and motion (static and dynamic) in the next study.

Another limitation of the current study is the use of morphed still images to create the dynamic stimuli, instead of using a naturalistic movement, e.g. video of the person. Although the use of morphed images as a dynamic stimuli increases control and allows standardisation of emotion intensity (Dobs et al., 2018), use of morphed images have been shown to reduce correct identification of familiar faces compared to naturally moving faces (Lander et al., 2006). It was found that naturally moving faces were easier to recognise due to the distinctive features that each individual face have when moving (Lander & Chuang, 2005; Butcher & Lander, 2017). Morphing images removes these idiosyncratic features in favour of experimental control and standardisation and therefore could have affected how faces were perceived on social traits explored in this study. Further studies could employ a more naturalistic dynamic stimuli, e.g. use of video images, to investigate how this would affect the perception of facial attractiveness, health, and age.

In sum, the current set of studies show that facial motion, emotion, and sex contribute in varying degrees to perceptions of facial age, attractiveness and health. Despite the close relationships between these three traits, our findings suggest that perceivers use different cues depending on which social trait is being evaluated.

Next steps

To disentangle the effect of target sex, motion, and emotion on evaluations of facial attractiveness and facial healthiness, I decided to investigate the three factors and their interactions simultaneously, using a larger participant and target sample.

Chapter 3.2. To smile or not to smile: Differential effects of gender, emotion, and motion in perception of attractiveness and health across the age range

Abstract

Previous studies have demonstrated certain bias in perceptions of facial attractiveness and health, however, these studies are limited in their generalisability as most use young samples and young observers, as well as being limited in the use of dynamic facial stimuli. Using a wider age range of facial stimuli, we investigated the role of facial age, motion, sex, and emotion on ratings of attractiveness and health. As expected, we found that young faces received the highest attractiveness and health ratings, however, the trends diverged on facial sex and motion. We observed gender bias on attractiveness rating, where female faces were rated as more attractive than males, but not on health ratings. We also found that in contrast with attractiveness ratings where dynamic faces were rated as more attractive, static faces were viewed as healthier. Our findings show that although facial attractiveness and health are closely and positively linked with each other, observers could be using different cues when making evaluations on these traits.

Thank you to Bethan Hughes, Danielle Jones, Emily Cooper, Rozalie Matejkova, and Ondrej Andrew Burysek, for helping with data collection for this chapter.

Research into facial attractiveness has demonstrated that certain facial features such as symmetry, averageness, and sexual dimorphism contribute to observer's perception of an individual's attractiveness (Jones & Jaeger, 2019; Rhodes et al., 2001; Thornhill & Gangestad, 1999) , and healthiness (Foo et al., 2017; Jones et al., 2001). From previous studies, we have learned that facial attractiveness and health are highly related to one another (Fink et al., 2006, 2012; Jones et al., 2018; Matts et al., 2007; Voegeli et al., 2021) and evolutionary psychology perspective posits that people's perceptions of these two traits are highly linked to an individual's *physical* health, where a face conveys signals and cues which convey a person's genetic quality (Buss & Schmitt, 2019; Jones, 2018,although see Kalick et al., 1998 and Foo et al., 2017). Evolutionary psychology further suggests that these facial cues have evolved to signal an individual's good genes and therefore aid in mate selection (Rhodes, 2006; Thornhill & Gangestad, 2006).

Buss (1989) argued that males and females look for different traits in a mate, where females look for males who would be able to provide support, while males appraise females for their reproductive value. For women, this translates to their reproductive value being closely linked with age, as the younger they are, the more fertile and less complications arise for childbearing, while men are not as restricted in this sense (Maestriperi et al., 2014). We could therefore argue that for females, looking younger would imply higher reproductive value (Maestriperi et al., 2014), and hence higher ratings of attractiveness and health, whereas this may not be the case for men, particularly as their ability to reproduce extends further than their female counterparts (McLellan & McKelvie, 1993).

Influence of facial age

Facial attractiveness and health have been found to decline with age (Matts et al., 2007; Samson et al., 2010), where the older a face looks, the less attractive (Foos & Clark, 2011; Kwart et al., 2012; McLellan & McKelvie, 1993) and less healthy they are perceived

(Jones et al., 2018). As we get older, the facial shape changes are brought about by widening of the jaw, increasing cranial strain, drooping of the nose and eye regions and atrophy of facial fat deposits (Henderson et al., 2016), whereas facial texture changes occur from emergence of fine lines, development of wrinkles from sun damage or the repetitive movement from speech and emotion expressions (Mendelson & Wong, 2012). These changes have been cited to influence the desire of middle and older-aged adults to 'slow down' the signs of aging, in a bid to appear younger, and hence more attractive and healthier (Clarke & Griffin, 2007; The American Society for Aesthetic Plastic Surgery, 2018).

However, as research on facial attractiveness have primarily used younger models and perceivers (Koscinski, 2013; Penton-Voak & Chang, 2008; Rhodes et al., 2012; Roberts et al., 2009; Rubenstein, 2005), findings may not necessarily be generalisable to older adults. Two studies were found to have used a wider age range of both target faces and perceivers and have supported the notion that perceived attractiveness decline with age (Foos & Clark, 2011; McLellan & McKelvie, 1993). However, these two studies have used stimuli that were not standardised, e.g. either cut out and rephotographed as black and white images (McLellan & McKelvie, 1993) or taken from the internet (Foos & Clark, 2011), which could have affected how the faces were perceived.

Female bias in perception of attractiveness

Research investigating gender differences in attractiveness ratings have broadly suggested that there occurs a female bias in perception of attractiveness. Morrison et al., (2013) presented male and female static faces to perceivers and found that female faces receive significantly higher attractiveness ratings than male faces, regardless of emotion expression. Similarly, Morrison et al. (2007) used facial models and found that feminised faces were quicker to be identified as attractive compared to masculinised faces. Finally, in a more recent study, Lindeberg et al. (2019) have also found that female faces were faster to be categorised as attractive than male faces, although the same is true for unattractive faces. At present, no study has directly investigated gender differences in facial health perceptions.

These studies therefore suggest that a gender bias occurs in attractiveness ratings, where, in general, female faces receive higher attractiveness ratings than male faces. In line with an evolutionary psychology perspective, this could be due to the more salient function of attractiveness for female reproductive value, e.g., facial attractiveness for women is highly associated with fertility (Maestriperi et al., 2014), whereas male reproductive value could also be judged in other traits, e.g., muscle mass (Foo et al., 2017). However, these studies used young targets and perceivers, therefore the boost in female attractiveness could be attributed to the high reproductive value of women in this group. However, earlier studies which used a wider age range of target faces and participants have also shown a positive bias towards female faces (Korthase & Trenholme, 1982; McLellan & McKelvie, 1993) although these earlier studies have used black and white images and therefore could have affected the perception of attractiveness, as skin colour has been shown to also contribute to facial attractiveness (Matts et al., 2007; Samson et al., 2010).

Emotion bias in attractiveness and health

Another factor that has been shown to affect ratings of attractiveness and health is the emotion expressed by the target face. Studies investigating the effect of emotion on attractiveness have consistently found that faces with positive expressions receive higher ratings compared to neutral (Bowdring et al., 2021; Golle et al., 2014; Penton-Voak & Chang, 2008; Rubenstein, 2005), sadness (Ueda et al., 2016) and other facial expressions (Morrison et al., 2013). Furthermore, Goelle et al. (2014) have demonstrated that it is not only the smile, but the *intensity* of the smiling that influences attractiveness ratings. They created morphed images depicting different intensities of smiling (25% - 100%) of attractive and unattractive faces and found that the more intense the smile, the more attractive the faces appear to be, and, interestingly, this also boosted the attractiveness ratings of ‘unattractive’ faces.

However, most of the studies which looked at the role of emotion in attractiveness ratings have employed young targets and perceivers (Bowdring et al., 2021; Golle et al., 2014; Morrison et al., 2013; Penton-Voak & Chang, 2008; Rubenstein, 2005). This has vital implications, as middle-aged women have been found to be the most likely consumers of appearance enhancing treatments and procedures (The American Society for Aesthetic Plastic Surgery, 2018).

As for health perception, Jones et al. (2018) presented participants with a wider age range of male and female faces depicting either a neutral or a happy expression. They found that young faces received the highest health ratings, followed by middle, and old groups; male and female target faces received similar health ratings, and that smiling faces looked healthier than neutral faces. Additionally, these findings persisted after accounting for facial attractiveness, implying that emotion expression independently influence health perceptions. This study has highlighted the importance of positive affect in evaluation of facial health across the age range, however, they only used static images in this study. The current study therefore also aimed to investigate the role of motion in health perception.

Influence of Motion

One of the main criticisms on attractiveness research is that until recently, static images with neutral expressions were used as target faces. This has been criticised as being low in ecological validity as daily interaction is marked by dynamic movements, either through speech or emotion expression (Penton-Voak & Chang, 2008; Rubenstein, 2005). It was therefore necessary to incorporate motion when investigating attractiveness perceptions. However, studies which incorporated static and dynamic stimuli in their experiments yielded inconsistent results. In general, findings suggest that motion (in form of speech) does not significantly affect attractiveness ratings (Rhodes et al., 2011; Roberts et al., 2009; Rubenstein, 2005), however, these studies have employed young target and perceiver samples and therefore generalisability to older samples are weak. No study has yet explored the influence of motion in health perception.

Rubenstein (2005) was the first to compare attractiveness ratings for static and dynamic conditions and found no significant difference between the two face types. However, they also found that the ratings were not correlated with one another, where the attractiveness ratings for the static condition of a face did not match the attractiveness ratings for the dynamic condition of the same face. This was argued to be due to the between-subjects design used, where participants only saw either a static or dynamic condition. On the other hand, Roberts et al (2009) looked at the correlation of attractiveness ratings in different sampling types, and found that when ratings for static and dynamic faces were given by the same person, i.e. repeated measures design, the ratings had higher correlation compared to when ratings were given by different people. However, Roberts et al (2009) did not investigate whether there was a significant difference in the attractiveness ratings between static and dynamic faces, only that the two are related.

Other studies which incorporated motion have found no significant difference in attractiveness ratings for males (Rhodes et al., 2011), females (Roberts et al., 2009) or both (Koscinski, 2013). However, these studies have used speech as their form of dynamic stimuli. Other studies which have incorporated emotion expression have found contradicting results.

Interactions of Sex, Motion, and Emotion

One study that has investigated the role of motion, emotion expression, and target face sex on attractiveness ratings is that of Penton-Voak and Chang (2008). Here, they utilised male and female faces on static and dynamic conditions depicting neutral or happy expression. In this study they found that in general, dynamic faces were more attractive than static faces and smiling faces were more attractive than neutral faces, whereas no main effects of target face sex were reported. The interactions found that for females, emotion expression has more influence on attractiveness ratings regardless of motion, whereas for males, dynamic faces were rated as more attractive, regardless of emotion. These findings suggest that perception of attractiveness for males and female faces are influenced by different factors, and therefore only employing static conditions on research investigating attractiveness could undermine the evaluations for some target models. However, although this study has highlighted the importance of stimuli conditions, they only used young samples both as targets and perceivers, therefore we are unable to determine whether the same trend occurs for older faces. If female attractiveness is more closely related to age than male attractiveness, we would predict that in older adults, attractiveness ratings for males and females should not deviate from each other, regardless of emotion or motion.

Aims, Hypotheses and Predictions

As there was a strong evidence of positive emotions being rated as more attractive and healthier, the current study therefore aims to investigate whether evaluations of attractiveness and health would be affected by the target face *sex*, *motion*, and *age* for two types of emotions separately. We believe that this would be the first study to investigate the effects and interactions of these variables on facial attractiveness and health.

For this section, therefore, our hypotheses were as follows:

1. There will be a significant effect of target Age Group, where Young faces would receive the most positive evaluations;

2. there will be a significant effect of Face Sex in the evaluations of attractiveness and health in dynamic, smiling faces, where in general, female faces would receive more positive evaluations;
3. there will be a significant effect of Face Motion, where dynamic faces would be rated more positively than static faces,
4. there will be a significant interaction between face Sex and Age Group, where gender differences would emerge in the Young group and Middle group, but not in Old faces;
5. there will be a significant interaction between face Sex and Emotion, where females would receive higher evaluations in the smiling faces, where male evaluations would not be affected; and
6. There will be an interaction between face Sex and Motion, where male faces will receive higher ratings on dynamic than static ratings; and
7. There will be a three-way interaction between target Age Group, Motion, and Sex.

Methods

Participants

Both tasks were open for participants aged 18 and over, with no exclusions for sex, and must have access to a computer and have no or corrected visual impairment to complete the experiments. Due to the nature of stimuli, only tasks completed by Caucasian participants were included in the analysis.

For attractiveness task, 137 participants accessed the anonymous link for the study through recruitment platforms (Prolific and SONA). Data collection ran between 17 January 2021 and 07 June 2021. 11 participants did not identify as Caucasians and therefore were not included in the analysis. No further exclusions were made. The final sample therefore had N = 126 participants (70F, 56M) with age range of 18-78 (M = 30.54, SD = 13.68).

For the health task, 126 participants accessed the study link anonymously using similar platforms. Data collection ran between 27 January 2021 and 29 June 2021. Seventeen participants did not meet the eligibility criteria and therefore were excluded. No further exclusions were made. The final sample therefore had N = 109 participants (59F, 49M, 1 Other), with age range of 18-59 (M = 28.59, SD = 9.57).

All participants were presented with the study aims and consented to participate. Only completed tasks were included in the final analysis. Psychology students were rewarded two credits for their participation and were also given a chance to be entered into a raffle draw to win one of two Amazon vouchers, along with members of the public. Participants from Prolific for both tasks were paid £3.75 for their participation.

Materials

For this experiment, we used all the 171 faces in the FACES database in both the static (Ebner et al., 2010) and dynamic conditions (Holland et al., 2019). For both conditions, we picked a neutral and smiling expressions, as negative expressions were shown to reduce attractiveness (Ebner et al., 2018) and health ratings for older adults (Jones et al., 2018).

To reduce fatigue and carry-on effects, we created four variations of the tasks. First, we allocated each face model a number (1-171). Using Python's random package (Vallat, 2021), we created face groups 1, 2, 3, and 4, with 43, 43, 43, and 42 faces respectively. In the first variation of the task, Group 1 depicted a static neutral face; Group 2 static smiling face; Group 3, dynamic neutral; and lastly Group 4 depicted dynamic smiling faces. For the subsequent variations, the face groups were moved along the conditions, which means that participants saw all faces, but only saw them in one condition. This strategy has been previously used in social perception studies Jones et al., 2018; Morrison et al., 2013; Voegeli et al., 2021).

Design

Using the face as our unit of measurement (Jones et al., 2018; Morrison et al., 2013), this study used a mixed design. Our between variables were *Age Group* (three-levels: Young, Middle, and Old), and *Sex* (two-levels: Male and Female). Our within-measures variables were *Motion* (two-levels: Dynamic and Static). Our dependent variables were the attractiveness and health ratings given by participants for each face, with separate analyses done for each emotion.

Procedure

Participants accessed the tasks using an anonymous link to Gorilla platform (Anwyl-Irvine, 2018). They were presented with Participant Information Sheet, and Consent was confirmed prior to commencing the task. Demographical information was collected, e.g. age, sex, and ethnicity. Prior to each task, participants were given an instruction as to which trait they were to judge each face. To reduce the study length and avoid carry-on effects, participants rated the faces on one trait only. Images were presented in four blocks, with each block depicting one face type, e.g. static neutral, static smiling, dynamic neutral, and dynamic smiling. Blocks were set-up to be presented randomly to each participant, and within each block, trials were also presented randomly. This means that no two participants had the exact same order of faces presented to them.

For each trial, a fixation cross appeared in the middle of the screen for 250ms. The target face appeared in the middle of the screen and was presented until the participant submitted their response. Below the image, participants were asked ‘How attractive/healthy is this face?’ A sliding scale was provided, with values between 0 (Not at all attractive/healthy) and 100 (Very attractive/healthy). After each block, participants had an opportunity to take a break before proceeding to the next block. The study took approximately 20-25 minutes to complete.

Data Analysis

To inspect data quality, individual participant scores were averaged and standard deviations were calculated, where those with SD = 0 were removed. As no participants met this criteria, all responses were kept. Scores were then averaged for each *target face* per condition, creating a composite score for static and dynamic conditions for each trait. As our preliminary studies and previous research have reliably shown that faces depicting a happy emotion were consistently rated as more attractive and healthier than neutral faces, we decided to run a 2 (static and dynamic) x 2 (male and female) x 3 (young, middle, and old) repeated measures ANOVA on happy and neutral faces separately for each trait. We also ran complementary Bayes factor analyses to determine whether the effects were under null hypothesis, given the data.

Results

Descriptive Statistics

Table 3.6 shows the means and standard deviations for attractiveness and health ratings categorised by emotion, age of face, sex of face, and motion. In general, Young faces received the highest trait evaluations, followed by the Middle and Old faces. Female faces also generally received higher trait evaluations than male faces. For attractiveness ratings, happy, dynamic faces generally received higher evaluations than happy, static faces; whereas the opposite trend was observed for health ratings. Neutral faces, on the other hand, received similar trait evaluations between the motion types for both attractiveness and health ratings.

Table 3.6

Means and Standard Deviations for Attractiveness and Health evaluations.

Emotion	Age Group	Sex (n)	Attractiveness Ratings				Health Ratings			
			Dynamic		Static		Dynamic		Static	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Happy	Young	Female (29)	53.91	8.33	54.72	9.30	70.55	5.75	72.57	7.04
		Male (29)	47.98	9.01	46.01	9.88	69.89	6.87	71.74	6.69
	Middle	Female (27)	40.31	7.71	38.06	8.15	55.91	7.07	59.05	6.34
		Male (29)	31.78	5.40	29.73	6.34	53.76	5.73	55.39	6.16
	Old	Female (29)	31.72	5.13	30.26	5.64	45.90	5.81	48.88	7.23
		Male (28)	29.59	4.77	27.39	5.14	46.29	5.38	47.50	7.52
Neutral	Young	Female (29)	51.05	7.62	53.23	8.26	68.75	5.28	69.18	5.70
		Male (29)	44.66	6.43	44.73	8.43	67.04	6.71	67.93	6.53
	Middle	Female (27)	35.35	8.10	35.29	6.50	53.35	6.24	54.35	6.99
		Male (29)	26.48	5.59	26.00	6.44	50.60	7.20	51.12	6.81
	Old	Female (29)	25.30	5.65	24.47	6.54	39.46	6.82	40.53	6.04
		Male (28)	22.96	4.96	22.38	4.30	42.51	5.59	42.19	5.57

Note: Ratings were made between 0-100.

Attractiveness

Happy Faces

Firstly, we found a significant main effect of Age Group as expected, $F(2, 165) = 139.632$, $p < .001$, $\eta p^2 = .629$, where in general, Young faces received the highest attractiveness ratings (M = 50.656, SD = 9.482), followed by the Middle (M = 34.818, SD = 7.899), and Old faces (M = 29.760, SD = 4.938). We found a significant main effect of Face Motion, $F(1, 165) = 23.101$, $p < .001$, $\eta p^2 = .123$, where dynamic faces were rated as more attractive (M = 39.258, SD = 11.449) than static faces (M=37.750, SD = 12.471). We also found a significant main effect of face sex, $F(1, 165) = 32.387$, $p < .001$, $\eta p^2 = .164$, where females received higher attractiveness ratings (M = 41.552, SD = 12.109) than male faces (M = 35.492, SD = 10.693). Figure 3.5 illustrates the attractiveness scores between each sex, age group and motion conditions for happy faces.

We did not find a significant two-way interactions between Face Motion and Age Group, $F(2, 165) = 2.283$, $p = .105$, $\eta p^2 = .027$); between Face Motion and Sex, $F(1, 165) = 3.059$, $p = .082$, $\eta p^2 = .018$; and between Age Group and Sex, $F(2, 165) = 2.890$, $p = .058$, $\eta p^2 = .034$. Lastly, we also did not find a significant three-way interactions between Motion, Age Group, and Sex, $F(2, 165) = 1.947$, $p = .146$, $\eta p^2 = .023$. . The complementary Bayes factor analysis showed that these findings more likely under the null hypothesis ($BF_{01} = 464509.907$, $BF_{01} = 5.393x10^{+32}$, $BF_{01} = 2120.941$, and $BF_{01} = 9.659$, respectively.) .

Neutral Faces

As expected, we found a significant main effect of Age Group, $F(2, 165) = 229.069$, $p < .001$, $\eta p^2 = .735$, where Young faces received the highest attractiveness ratings (M = 48.414, SD = 8.287), followed by Middle faces (M = 30.616, SD = 7.764), and Old faces (M = 23.795, SD = 5.14). We also found a significant main effect of Face Sex, $F(1, 165) = 41.156$, $p < .001$, $\eta p^2 = .200$, where females received higher ratings (M = 37.496, SD =

13.258) than male faces (M = 31.299, SD = 11.327). However, we did not find a significant main effect of Face Motion, $F(1, 165) = .023$, $p = .880$, $\eta p^2 = 000$. The complementary Bayes factor analysis showed that this is more likely under the null hypothesis ($BF_{01} = 1.214 \times 10^{+34}$).

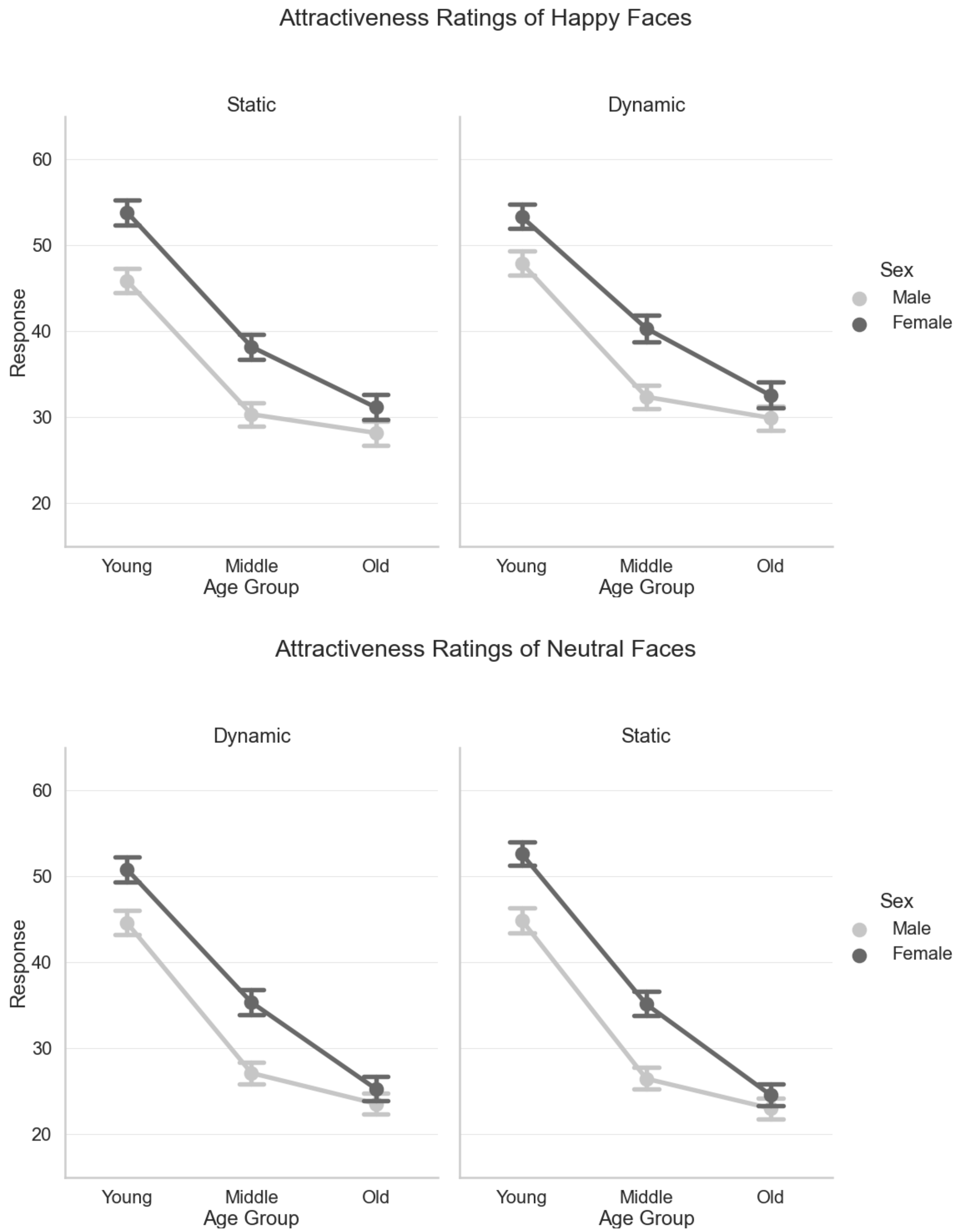
We found a small, but significant two-way interaction between Age Group and Sex, $F(2, 165) = 4.500$, $p = .013$, $\eta p^2 = .052$, where for Young and Middle faces, females received significantly higher attractiveness ratings than males ($p < .001$), but for Old faces, no significant difference was found between sexes ($p = 1.00$). Within the sexes, attractiveness ratings for females decline through the age groups ($p < .001$), whereas Young male faces received significantly higher attractiveness ratings than Middle and Old faces (both $p < .001$), but no significant difference between Middle and Old male faces ($p = .537$).

We also found a small, but significant two-way interaction between Face Motion and Age Group, $F(2, 165) = 3.076$, $p = .049$, $\eta p^2 = .036$; a post-hoc pairwise comparison with Bonferroni corrections revealed that within each age group, there was no significant difference between the face types (all $p > .05$), and within motion, each age group received significantly different attractiveness ratings, Younger faces received higher ratings compared to Middle and Old faces, and the Old faces received the lowest attractiveness ratings for both Motion types.

We found no significant two-way interaction between Face Motion and Sex, $F(1, 165) = 12.403$, $p = .230$, $\eta p^2 = .009$; and no significant three-way interactions between Face Motion, Age Group, and Sex, $F(2, 165) = 10.736$, $p = .287$, $\eta p^2 = .015$. The complementary Bayes factor analysis for these two interactions showed that this is more likely under the null hypothesis ($BF_{01} = 4.559 \times 10^{+32}$; and $BF_{01} = 9.659$ respectively).

Figure 3.5

Attractiveness Ratings for Male and Female Happy Faces on Static and Dynamic Conditions Across the Age Group.



Health

Happy Faces

As expected, we found a large, significant main effect of Age Group, $F(2, 165) = 233.375$, $p < .001$, $\eta p^2 = .739$, where, similar to attractiveness ratings, Young faces received the highest health ratings (M = 71.185, SD = 6.243), followed by Middle faces (M = 55.975, SD = 5.91), and the Old faces (M = 47.146, SD = 5.977). We found a significant main effect of Face Motion, $F(1, 165) = 33.930$, $p < .001$, $\eta p^2 = .171$, where static faces were rated as healthier (M = 59.257, SD = 12.082) than dynamic faces (M = 57.125, SD = 11.746). This was an unexpected finding, as we predicted that dynamic faces would be rated as healthier, following the trend of attractiveness rating. We did not find a significant main effect of face sex, $F(1, 165) = 2.243$, $p = .136$, $\eta p^2 = .013$. The complementary Bayes factor analysis showed that this is more likely under the null hypothesis ($BF_{01} = 2.942 \times 10^{+51}$), implying that males and females received similar health ratings.

Figure 3.6 shows the interactions for health task. All two-way interactions yielded a non-significant effect: Face Motion and Age Group, $F(2, 165) = .130$, $p = .879$, $\eta p^2 = .002$; Face Motion and sex, $F(1, 165) = 2.459$, $p = .110$, $\eta p^2 = .015$; and Age Group and Sex, $F(2, 165) = .680$, $p = .508$, $\eta p^2 = .008$. Lastly, we also did not find a three-way interaction between the three variables, $F(2, 165) = .468$, $p = .627$, $\eta p^2 = .006$. The complementary Bayes factor analysis showed that these findings more likely under the null hypothesis ($BF_{01} = 15.690$, $BF_{01} = 1.665x10+46$, $BF_{01} = 1.753x10+6$, and $BF_{01} = 1202.348$, respectively).

Neutral Faces

We found a large, significant main effect of Age Group, $F(1, 165) = 303.544$, $p < .001$, $\eta p^2 = .786$, where as expected, Young faces received the highest health ratings (M = 68.224, SD = 5.603), followed by the Middle (M = 52. 300, SD = 6.539), and the Old faces (M = 41.149, SD = 5.786). We did not find a main effect of Face Motion, $F(1, 165) = 3.117$,

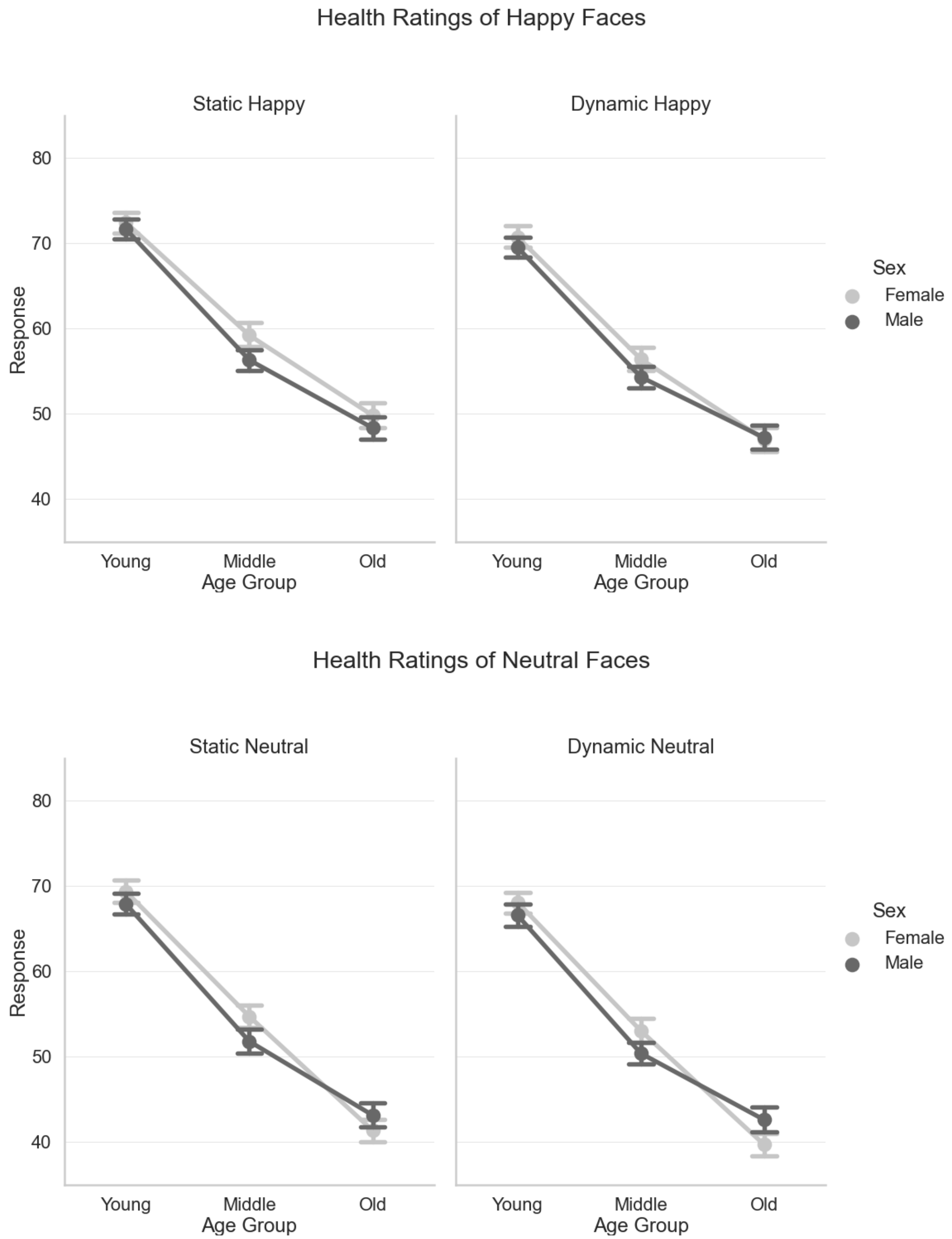
p = .079, ηp^2 = .019, and similarly, we did not find a main effect of Face Sex, $F(1, 165)$ = .607, p = .437, ηp^2 = .004. The complementary Bayes factor analysis showed that this is more likely to be under the null hypothesis ($BF_{01} = 4.279x10^{+52}$; and $BF_{01} = 4.482x10^{+52}$, respectively). This implies that for faces with neutral expressions receive similar health ratings regardless of motion and sex.

We found a small, but significant two-way interaction between Age Group and Sex, $F(2, 165) = 3.069$, p = .049, ηp^2 = .036. Post-hoc comparisons using Bonferroni correction showed that as expected, there is a decline in the health ratings as faces get older, for both sexes (all $p < .001$). Additionally, within the Age Group groups, male and female faces did not receive significantly different health ratings (all p = 1.000). However, female faces received higher health ratings in the Young and Middle Age Groups, whereas this was reversed for Old faces.

We did not find a significant two-way interaction between Face Motion and Age Group, $F(2, 165)$ = .117, p = .890, ηp^2 = .001; and Face Motion and Sex, $F(1, 165)$ = .469, p = .494, ηp^2 = .003. Finally, we also did not find a significant three-way interactions between the three variables, $F(2, 165)$ = .628, p = .535, ηp^2 = .008. The complementary Bayes factor analysis showed that these findings were more likely under the null hypothesis (BF_{01} = 28.928, $BF_{01} = 4.344x10^{+53}$; and BF_{01} =1557.448, respectively).

Figure 3.6

Health Ratings for Male and Female Happy Faces on Static and Dynamic Conditions Across the Age Group.



In summary, we tested Happy and Neutral faces separately using Young, Middle, and Old male and female faces in both static and dynamic conditions (Table 3.7). We found that:

1. For both attractiveness and health task, the ratings decrease as a function of age, that is, Young faces received the highest ratings, followed by the Middle and Old faces.
2. There was a significant effect of motion only for happy faces in both tasks – dynamic faces were more attractive, while static faces were rated as healthier.
3. There was a significant effect of gender only for attractiveness task, where female faces were rated as more attractive.
4. A two-way interaction between Age and Sex were found for Neutral faces for both tasks, where Young and Middle Female faces were rated as more attractive and healthier.
5. Overall, no two-way interactions between Motion and Sex, and Age Group and Motio were found, and no three-way interactions were found in both tasks.

Table 3.7.

Summary of Results for Perception of Attractiveness and Health using Static and Dynamic Images.

Task	Attractiveness		Health	
	Happy	Neutral	Happy	Neutral
Age	/	/	/	/
Sex	/	/	x	x
Motion	/	x	/	x
Age x Sex	x	/	x	/
Age x Motion	x	/*	x	x
Motion x Sex	x	x	x	x
Age x Sex x Motion	x	x	x	x

Note: / denotes significant effect. X denotes no significant effect. *After Bonferroni corrections, no significant pairwise comparisons were found.

Discussion

The current study investigated whether factors such as target age group, motion, and gender would have an effect on perceptions of attractiveness and health of smiling and neutral faces.

First, we will look into the findings for attractiveness. As expected, we found that younger faces were rated to be more attractive than older adults in both smiling and neutral conditions, supporting our hypothesis 1 that there will be a main effect of Age Group on ratings. This supports previous findings where younger faces were found to be more attractive (Kwart et al., 2012), further demonstrating the premium that youth holds in facial perception. Furthermore, we also found support for our hypothesis 2 that there will be a main effect of face sex, where in line with previous studies (Penton-Voak & Chang, 2008), we found a gender bias on attractiveness ratings, that is, female faces were rated to be more attractive than male faces. This further provides support to the evolutionary perspective, which posits that for females, facial attractiveness has more value than in males (Buss & Schmidt, 1989). Taking the age and gender bias found in attractiveness ratings, our findings highlight the main concerns of older women, where as they get older, they are perceived to be less attractive and therefore provides an explanation as to why they feel the need to maintain their youthfulness. In addition, we want to highlight that for female faces, there is a steady decline in attractiveness ratings as the faces move from Young, Middle, and Old, whereas for males, Middle and Old faces were similar in attractiveness ratings, further supporting notion that women's attractiveness value is closely linked with age (Buss, 1989; Schmidt & Buss, 1996)

We also found a partial support to our hypothesis 3, that there will be a main effect of motion in attractiveness ratings, where we only found that this is the case for happy faces, not in neutral faces, the Bayes factor analysis showed that this is likely a null effect, given the data, implying that for neutral faces, the perception of attractiveness remains similar across facial movement. This could be due to the more salient change in the moving happy faces, e.g. the dynamic change from neutral to smiling could have been perceived as more attractive than a simple still shot of a smiling face (Dobs et al., 2015). The neutral faces, however, changed from neutral to neutral and therefore no perceptual change occurred.

We did not find significant two-way interactions for attractiveness ratings of happy faces, but found a small, but significant effect between age group and sex for the neutral faces. Our studies imply that for Young and Middle female faces were perceived to be more attractive compared to their male counterparts, whereas no differences were found for Old faces. This highlights the premium of youth for females, as there is an observable decline in ratings of attractiveness as a function of age that was not observed in males when using neutral stimuli.

Surprisingly, we did not find a three-way interaction between the variables in both neutral and smiling faces – this contradicted the findings from Penton-Voak and Chang (2008) where they found that male faces were more attractive when moving. It could be due to the differences in the age of stimuli used, where the current study employed a wider age range of target faces, whereas Penton-Voak and Chang (2008) only used young male faces. The Bayes factor analysis suggested that this is a null effect, given the data, therefore implying that target sex, age, and facial motion independently contribute to the perception of facial attractiveness.

For health ratings, we again found support for hypothesis 1 that there will be a main effect of face age – as expected, for both emotions, young faces received the highest health ratings, followed by the middle and old faces. This highlights the notion that signs of facial aging (as a consequence of physical aging) contribute to diminishing perception of facial healthiness (Samson et al., 2011).

However, in contrast with the findings from attractiveness ratings, our results did not support our hypothesis 2 that there will be a main effect of face sex, where we found that male and female faces received similar health ratings. This is the first study to demonstrate that for health perceptions, no gender bias was observed. This is an interesting finding, considering the close relationship found between health and attractiveness ratings. Similarly, we also found a partial support for our hypothesis 3 that there will be a main effect of motion – although this was only observed on happy faces. Interestingly, our findings were contradictory to our prediction (and the findings from attractiveness ratings), where we found that static images were rated to be healthier than dynamic images. This is unexpected, as health and attractiveness ratings are highly correlated with each other (Jones et al., 2018). This could imply that different mechanism underlying the perception of facial attractiveness and health. This therefore supports the notion that perhaps perceivers use different facial cues when judging for attractiveness and health – it was suggested that facial shape and facial colour contribute more to facial health (Jones, 2018; Samson et al., 2011), whereas facial texture and sexual dimorphism contribute to facial attractiveness (Jones & Jaeger, 2019). These findings from the current study are the first to demonstrate that health ratings were not affected by the target gender and motion, despite being closely linked with attractiveness ratings.

Similar to the attractiveness data, we did not find significant interactions between variables for the happy faces, but found a small, but significant interaction between Face Age and sex for the neutral faces. Again, we found that Young and Middle female faces were rated as healthier compared to their male counterparts. Interestingly, Old male faces were rated as healthier than Old female faces – this again highlights the premium of youth for females – and worse, being older is more detrimental to perception of female healthiness.

The lack of three-way interactions between the variables given the data, and the support for the main effect of motion only for happy faces implies that for health ratings, perhaps the most important factor is the facial age – how old we are (and by extension, we look) determines how healthy we are perceived. This therefore further highlights the increase in the consumption of anti-aging treatments, as these procedures aim to make the consumer look younger, and by extension, healthier.

One of the limitations of the current study is that although we have incorporated a wider range of perceiver age, the sample is still predominantly young (mean = 30). Previous studies have shown an own-age bias in attractiveness perception, and therefore it could be the case that the results were driven by a predominantly young sample. However, as we were interested as to how the same face would be rated on different motions and emotions, this bias would have been reduced. Another limitation is that due to the restrictions with face-to-face interactions (due to COVID-19), the current study has been run completely online using the participants' own computer. This could mean that different monitors could have presented the faces with subtle differences in colour and size, despite the study being held in one platform. This could be an issue with the interpretations of health ratings, as previous studies have suggested that facial colour contributes to perceptions of health. However, the preliminary studies were also run online, and we found high levels of agreement between participants for the faces. We therefore believe that this is the case for the current study as well.

In sum, the current study has demonstrated that the target's age, sex, and motion contribute independently to ratings of attractiveness, but only the target's age had a significant effect on health ratings. Such findings are intriguing, as facial attractiveness and health have been known to be highly associated with each other, that is, the more attractive someone is, the healthier they were perceived to be. This therefore supports the notion that different facial cues and signals may be used by perceivers depending on which facial trait is being judged.

Next steps

Studies have demonstrated that people are good at categorising faces into various groups, e.g. age, sex, and ethnicity. As we have seen in this chapter, a target face's age, sex, and emotion all play a role in how targets are perceived based on their facial age and sex. In the next chapter, I explore whether a target's ethnicity affects other's evaluations of their attractiveness, health and age.

Chapter 4.

Other-ethnicity effects

Chapter 4.1 Do individual differences in face recognition ability moderate the other ethnicity effect?[1]

Abstract

Individuals are better at recognizing faces from their own ethnic group as compared to other ethnicity faces – the *other-ethnicity effect* (OEE). This finding is said to reflect differences in experience and familiarity to faces from other ethnicities relative to faces corresponding with the viewers' ethnicity. However, *own-ethnicity* face recognition performance ranges considerably within a population, from very poor to extremely good. In addition, within-population recognition performance on other-ethnicity faces can also vary considerably with some individuals being classed as '*other ethnicity face blind'* (Wan et al., 2017). Despite evidence for considerable variation in performance within population for faces of both types, it is currently unclear whether the magnitude of the OEE changes as a function of this variability. By recruiting large-scale multinational samples, we investigated the size of the OEE across the full range of own and other ethnicity face performance whilst considering measures of social contact. We find that the magnitude of the OEE is remarkably consistent across all levels of within-population own- and other-ethnicity face recognition ability, and this pattern was unaffected by social contact measures. These findings suggest that the OEE is a persistent feature of face recognition performance, with consequences for models built around very poor, and very good face recognisers.

[1]For the published article, please see:
Childs, M. J., Jones, A., Thwaites, P., Zdravković, S., Thorley, C., Suzuki, A., Shen, R., Ding, Q., Burns, E., Xu, H., & Tree, J. J. (2021). Do individual differences in face recognition ability moderate the other ethnicity effect? *Journal of Experimental Psychology: Human Perception and Performance, 47*(7), 893–907. https://doi.org/10.1037/xhp0000762
Supplementary Material: https://osf.io/bwhtg/

The face plays a central role in human social interaction. Typically, from a young age, we are able to identify familiar faces which aids in survival and attachment (Barrera & Maurer, 1981), and as we age, our ability to recognise faces in different contexts allows us to distinguish between familiar and unfamiliar faces which has an impact on our interpersonal relationships (Gobbini et al., 2004). A consistently reported phenomenon in facial recognition is that typically developing samples are generally better at recognizing faces from their own ethnicity compared to other ethnicities; also known as the other-ethnicity effect (OEE; Malpass & Kravitz, 1969; McKone et al., 2012). A well-known theoretical account of this effect is posited by *perceptual expertise theory*, which suggests that the OEE reflects a lack of experience in seeing and encoding other-ethnicity faces. Supporting evidence comes from infant studies, where 6-9 month-old infants were shown to be able to discriminate between own-ethnicity and other-ethnicity faces (Anzures et al., 2013; Kelly et al., 2007; Sangrigoli & de Schonen, 2004). Training studies, where participants show reduced OEE after training with other-ethnicity faces (Lebrecht et al., 2009) also support this notion.

The OEE and contact

A key factor that is claimed to impact one's performance with faces of different ethnicities relates to the amount of social contact they have with certain groups. The *contact hypothesis* posits that the higher the contact an individual has with faces of a particular ethnicity, the more accurate they are at recognizing members of that group (Goldstein & Chance, 1985). For example, Zhou et al. (2019) demonstrated that Whites and East Asians born and raised in the wider Toronto area had comparable face recognition abilities for Whites and East Asian faces (i.e. East Asians born in the Toronto area did not display an OEE for White faces). In addition, length of exposure to White faces moderated the OEE for East Asians (i.e., the longer they had lived in Toronto, the smaller the OEE). In general (although see Harvey, 2014; MacLin et al., 2004; Ng & Lindsay, 1994), studies investigating the role of contact (both in geographical and self-report) in face recognition show that as contact increases so the magnitude of the OEE can diminish (see Table 4.1 for summarised findings). However, it is particularly noteworthy that although contact can diminish the magnitude of the OEE, it often does not eliminate this effect completely (De Heering et al., 2010; although see Estudillo et al., 2020)

Variations in Individual Face Processing Performance for Own- and Other-Ethnicity Faces

In Table 4.1, we have provided a summary of several studies of the OEE that explored the degree to which the effect is impacted by social contact (e.g., high 'contact' group versus low 'contact' group), and a number have shown that across groups, the magnitude of the OEE can indeed vary. But this often masks the fact that *within* groups there is often considerable variance in individual ability with own-ethnicity faces – where it is often implicitly assumed that *own-ethnicity* face performance (i.e., baseline face recognition ability) across two samples of the same population (e.g., two UK White populations) is quite homogeneous, such that between group differences are driven by other variables (such as social contact).

For own-ethnicity faces, there are two sub-groups of facial recognisers – very good (super-recognisers; SRs) and very poor (developmental prosopagnosics). People with *developmental prosopagnosia* (DP) have impairments in recognizing own-ethnicity faces despite having normal intelligence and an absence of brain injury (Bate et al., 2019; Burns, Bennetts, et al., 2017; Burns et al., 2014; Burns, Martin, et al., 2017; Jackson et al., 2017). Interestingly, in our experience, DPs often report anecdotally *"all faces look the same to me…."* and *"I often confuse two different people who I know that look similar…"* (Bate & Tree, 2017).[1] Conversely, people dubbed *super recognisers* (SR), are reported to do extremely

[1] In addition, very recently a DP volunteer in our lab mentioned that he had confused his girlfriend with his best friend's girlfriend because they had superficial physical similarities (similar height, build, hair colour/style and clothing), despite the fact that one was Asian and the other White.

well with own-ethnicity faces (Ramon et al., 2019). What makes the two perform differently? Qualitative differences in face processing between controls and the two groups have been identified, e.g. lack of inversion effect (de Gelder & Rouw, 2000; B. C. Duchaine et al., 2006; Rouw & Gelder, 2002)on DPs, improved face recognition in DPs following holistic face processing training (DeGutis, et al., 2014), and significantly stronger use of holistic face processing by SRs compared to controls (Belanova et al., 2018). These suggest that individuals in the either end of the face recognition ability may look at a face stimuli differently than an average person, thus the difference in their performance.

However, more studies have shown support for quantitative differences, that is, DPs and SRs are simply on the low and the high end of the face recognition ability performance respectively (Russell et al., 2009). Using eye-tracking technique, Bobak et al. (2017) looked at how DPs, SRs, and the typical group look at a static face by comparing their eye-movements while viewing static images. They found that generally, those with more severe DP spent significantly *less* time looking at internal features overall, some DPs looked at eyes and nose less and spent more time looking at their mouth areas compared to controls; whereas SRs spent *more* time looking at the nose, particularly when they were asked to remember the faces compared to controls. This study showed that DPs were more varied and had a qualitatively different pattern of viewing the internal facial features compared to controls, suggesting a different configural processing, whereas SRs were similar to controls on their viewing pattern, albeit paying more attention to the nose area. This study therefore suggests that at least some DPs' face recognition ability differ qualitatively to controls, whereas SRs differ in a more quantitative manner.

There is now a great deal of evidence that suggests that the range of own-ethnicity face recognition accuracy across individuals for a particular population can be substantial (i.e., several standard deviations), and thus raises an important question – might the magnitude of the OEE change as a function of this variability? One approach to exploring this question is to focus on the performance of sub-populations linked to the 'extremes' of this distribution of own-ethnicity (baseline) face recognition ability – namely, on those who are performing very poorly (developmental prosopagnosia) or those performing extremely well (super recognisers). The logic being, if individual variability in baseline face recognition does impact on the emerging OEE, one might expect differences between sub-populations – and we will discuss this work now. Our study takes a novel approach, however, by exploring the degree to which the magnitude of the OEE varies across *the full* distribution of base level face recognition ability, and thus considers this issue in the widest possible sense (more below).

As mentioned previously, DPs show significant impairments in recognising own-ethnicity faces and have mentioned that faces appear similar to them, hence, the poor performance. This raises an interesting question – perhaps poor base-level face recognition ability emerges because of a general inability to draw from one's visual experience when learning faces? - that is, *despite* high familiarity/experience with own ethnicity faces, performance remains poor. If this is true, then we might expect poor face recognisers to do equivalently (with no OEE) across all ethnicities of faces ("*all faces look the same…"),* since high visual experience gives them little benefit at all. However, perceptual studies (Cenac et al., 2019; DeGutis et al., 2013) have found that DPs as a group demonstrated an OEE. A recent study by Cenac et al. (2019) looked at facial recognition abilities of White controls and DP participants using a sequential matching task (with White, East Asian, and Black ethnicities). All participants were matched on measures of social contact with other-ethnicity faces (i.e., minimal contact with people from East Asia and Black backgrounds).

Cenac et al. (2019) concluded that DPs in their sample did not have disproportionately poorer performance for other-ethnicity faces relative to controls. However, their findings could not speak to the issue of whether the OEE was present across both groups because this study did not find an overall OEE for either group (*despite* the low degree of contact). Indeed, the data reported by Cenac et al. (2019) illustrated a trend towards an inverted OEE – with controls and DPs better at matching other-ethnicity faces. It remains unclear why this occurred, but might reflect the deliberate increased variability of the other-ethnicity faces in their stimuli (all computer generated), which may have made the other-ethnicity faces easier to discriminate. In any case, no typical OEE was reported using their paradigm, which may be problematic with respect to interpreting their findings. Putting this issue aside, their findings suggest that DP cases are largely worse than controls for *both* own and other ethnicity faces on testing of face perceptual matching.

SRs on the other hand, have superior face recognition ability, and may thus show a general 'boost' to recognition performance for faces of a variety of ethnicities (outside their own), such that for them the OEE may be relatively diminished. Alternatively, SRs may still show an own-ethnicity face advantage despite their generally excellent face recognition abilities. Similar findings from Bate, Bennetts, Hasshim, et al., 2019 and Robertson et al., 2020 independently provide evidence for the latter pattern using various face memory and face matching tasks, which show that while SRs outperformed a matched sample on respective tests (i.e., better performance with *both* own- and other-ethnicity faces), a similar OEE size was found across the two groups. This suggests that even when base-level face recognition performance is extremely good, an advantage remains for own-ethnicity faces.

Table 4.1

Summary of Key Studies' Findings Relating to the Other-Ethnicity Effect in Relation to Geographical and Self-Report Contact

Study	Test comparisons	Samples	OEE
Chiroro & Valentine, 1995	Old-new	Africans and Whites living in Harare, Zimbabwe (high contact) Whites in UK (low contact) Africans in South Zimbabwe (low contact)	Hits: High contact group had similar levels of hits for both African and White faces compared to low contact groups. False Positives: High contact Africans had lower FP compared to the other groups.
De Heering et al., 2010 (Geographical)	Old-new	Adopted Asian children in Belgium and White children	White children showed OEE; Asian children showed similar recognition of Asian and White faces.
Hancock & Rhodes, 2008 (Self-report)	Recognition tasks using upright and inverted images	Chinese and Whites living in Australia (varied arrival times)	Increased contact with the other-ethnicity predicted lower OEE in recognition of upright faces, and reduced inversion effects.
Harvey, 2014 (Self-report)	Old-new	White students tested on White and Indian faces	No significant effect of contact levels on recognition performance of Indian faces.
MacLin et al., 2004 (Self-report)	Recognition tasks using upright and inverted images	White students who were categorised as either Novices/Experts in African-American basketball players	No inversion effects found in both groups.
Ng & Lindsay, 1994 (Self-report)	Old-new	Study 1: Whites and Asians living in Canada (Asians reported *high* contact with Whites) Study 2: Whites and Asians living in Singapore (Whites reported *low* contact with Asians)	Study 1: Asians showed similar FA rates for both White and Asian faces. Self-report contact was not significantly related to recognition performance. Study 2: Whites recognized both types of faces equally.

			Whites in Singapore did not have a significantly different recognition performance compared to Whites in Canada.
(Rhodes et al., 2009 (Geographical and self-report)	White and Asian faces - blurred faces and scrambled faces	Chinese students living in Australia (varied in arrival time)	Hits and false alarm rates (d’) had a negative correlation with duration of stay in Australia. Self-report contact did not reach significance.
Tanaka et al., 2004 (Geographical and self-report)	Part-whole task	Whites and Asians living in Germany (Asians reported *high* contact with Whites)	Whites = high recognition of whole face for White faces, low recognition of whole and part faces for Asian faces. Asians = no significant difference in the recognition of part or whole faces for both face types.
Wright et al., 2003 (Geographical and self-report)	Old-new	Blacks and Whites living in South Africa (high contact) Whites in UK (low contact)	Hits and false alarm rates (d’) of Black African population were significantly negatively correlated with self-report contact.
Zhou et al., 2019 (Geographical and self-report)	Cambridge Face Memory Test (CFMT) Australia and Chinese	Chinese individuals living in Australia (varied arrival time) Whites	Higher contact (longer time spent in Toronto and higher self-report contact) with Whites
Zhao et al., 2014 (Self-report)	Part-whole task, blurred and scrambled task, CFMTs	Chinese and Germans	Higher contact predicted smaller OEE in CFMTs, whole condition, and blurred condition, compared to part and scrambled conditions.

Thus, there is preliminary evidence that the OEE persists at the ‘extremes’ of *own-ethnicity* recognition performance within a given population – when this is considered via a comparison of performance between population sub-groups. However, there remains an additional pattern of ‘extreme’ within-population individual performance to be considered; namely, extremely poor *other-ethnicity* performance. Given the fact that within a population there is a distribution of performance with own-ethnicity faces, an assumption is that a similar distribution exists for individuals with other-ethnicity faces, and that these distributions are correlated, moving together. However, it may also be possible that there are individuals who have very poor performance with other-ethnicity faces despite good own-ethnicity face performance, akin to an exaggerated form of OEE – a pattern dubbed ‘*other ethnicity blindness’*.

To explore this issue, Wan et al. (2017) tested samples of both White and Asian participants on the Australian and Asian-CFMTs – in order to identify such ‘extreme’ poor performers they used absolute cut-off scores for each test (i.e. mean accuracy minus 2 standard deviations; SD). Participants who scored lower than 2SDs below the mean on their own-ethnicity face memory test were excluded to rule out the influence of general poor facial recognition ability (i.e., developmental prosopagnosia). White participants who met the criteria for ‘other ethnicity blindness’ were thus identified using a cut-off from the Asian participants’ sample on Asian-CFMT (and vice versa for Asian participants) – and under this criteria, it was found that 8% (*N*=36) of the sample performed lower than 2SDs below the mean. It was further argued that this selectively extremely poor facial recognition for other-ethnicity faces was neither due to lack of effort, nor poor general facial recognition ability, and that the level of contact may influence such cases. However, we would point out that this study only used one CFMT test to ‘diagnose’ participants who were other-ethnicity face blind. Typically, two or more tests are used to ‘diagnose’ DP (i.e. own-ethnicity face blindness); thus it is unclear whether the cases identified would continue to meet criteria for 'other-ethnicity face blindness' if other tests had been used - given the possibility of regression to the mean (discussed below). Nonetheless, this work suggests if we consider within population individual variance on *other-ethnicity* face recognition, it may be the case that the magnitude of the OEE varies across the distribution, where it may be being magnified at one extreme.

However, in all these studies of the OEE with individuals who are in the 'extremes' of own- or other-ethnicity face recognition ability, the approach has been to compare a (often quite small) sample of their 'extreme' group with another sample that comprised the rest of the population. A key criticism of this practice is that it involves the use of an arbitrary cut-off criteria score (2 SDs below average on a key test, as described above) for group categorisation, which likely does not reflect qualitative differences in performance. In other words, participants with performance either side of such a cut-off (i.e., 2.02 SD below average versus 1.98 SD below average) may artificially imply key group differences even when the performance between individuals may not be significantly different. This is a key motivation for the current study's novel approach – since it ensures explicitly that we did not group the participants into categories, but rather considered performance across the *entire* distribution (i.e., at all levels of performance from extremely poor to extremely good) – and thus we can ask (for the first time) whether the magnitude of the OEE remains equivalent across *all* levels of performance in a given population. It is important to note that although there are studies which looked at all levels of recognition performance in a population using both own- and other-ethnicity face recognition tasks, e.g. Horry et al. (2015) and Robertson et al. (2020), they do not explicitly measure the magnitude of OEE across the whole of the population – in Robertson et al.'s case, they only made comparisons of OEE magnitude for super-recognisers and controls, and in Horry et al.'s case, they only reported the correlation of own- and other-ethnicity face recognition performance – and therefore do not necessarily touch upon this matter.

Furthermore, not only do we consider the question of the size of the OEE across *own-ethnicity* face performance, we also explore the same issue from the position of *other-ethnicity* face performance. To our knowledge, this is the first study to use an out-group face ability measure as a predictor of OEE, which opens another avenue for us to understand this effect further.

Finally, given we are interested in the *universality* of the magnitude of the OEE across individuals in a population, we also sought to explore this issue across a number of different nations with populations that were either largely White (UK, Australia, and Serbia) or largely Asian (China, Japan, S. Korea, and Singapore), and thus the multi-national nature of our sample would allow us to investigate OEE in a more extensive manner.

Exploring within population individual variation in face recognition

It is noteworthy that a potential criticism of some of the previously discussed research on sub-groups of 'extremes' of individual performance is that they largely studied face *perception* (i.e., face matching), rather than face recognition. This is despite the fact that group-based studies of the OEE (see Table 4.1) have often focused on face *recognition*. To address this issue, another key motivation for the current study was that it sought to focus on individual variation within a population on measures of face recognition performance. In order for us to achieve this objective, it was important for us to use a well-validated measure of face recognition ability – and so we selected the Cambridge Face Memory Test (CFMT). In this case, we used three well-established versions: Boston (Duchaine & Nakayama, 2006), Australian (McKone et al., 2011), and Asian (McKone et al., 2012). In Table 4.2, we summarise a number of studies that used versions of the CFMT to investigate the OEE – importantly, in all cases the studies report a robust OEE (Cohen's *d* effect sizes between 0.5 –

1.24). In addition, because of its established validity and reliability, the CFMT has been used in a great range of individual differences work relating to face recognition over the last fifteen years (see Wilmer, 2017 for a comprehensive review). Thus, we have confidence that the CFMT is a robust tool for our current purposes.

An impetus for using the CFMT and the final motivation for the current study relates to the fact that it has three different versions (mentioned above) – and thus we would be able to utilise a CFMT test (e.g., the CFMT Boston) as an *independent* measure of individual face recognition memory performance from those used to traditionally capture and calculate the OEE (e.g., CFMT Australian versus CFMT Asian). This enables us to consider an important potential confound - *regression to the mean* (that is, individual performance can vary around its “true mean”, such that an extreme high or low score may naturally move on its second measurement). Put simply, if a key group of interest (DPs, super- recognisers, or cases of other-ethnicity blindness) is initially selected via ‘extremely’ poor scores on one measure (own-ethnicity face recognition), it is likely these same participants might be less poor on a second measure of face recognition because of regression to the mean. Therefore, the observed differences between two tests could be simply due to this phenomenon when the same test is used as the classifier *and* a comparator. Having a third face recognition memory measure that would provide an independent measure of face recognition memory from that used to compute the OEE was thus extremely useful, and the three well-established variants of the CFMT made it ideal for our purposes.

Table 4.2

Summary of Key Studies' Findings Relating to the Other Ethnicity Effect Using CFMT

Study	Test comparisons	Samples	OEE
Zhou et al., 2019	Boston-Asian	White	d = .64
DeGutis et al., 2013	Boston - Asian	White	d = .5
Crookes & Rhodes, 2017	Australian - Asian	White	d = 1.04 (standard) d = 1.24 (self-paced)
Horry et al., 2015	Australian - Asian	White	d = .91
		Asian	d = 1.14
Wan et al., 2015	Australian - Asian	White	% difference = 7.25
		Asian	% difference = 8.84
McKone et al., 2012	Australian - Asian	White	d = .76
		Asian	d = .84

Note: List of studies which used two versions of CFMT (own- versus other-race) to measure OEE. Note that all studies reported a robust effect, thus implying that in normative population, individuals are better at recognizing faces from their own- compared to those from other-ethnicities.

The fact that the CFMT has three variants also made it ideal for the current study given we sought to recruit large samples of both White and Asian participants. The current study aims to use these three CFMT variants in testing these different populations in order for our analyses to ask two different, but related questions. Firstly, for the White sample, our independent measure of face recognition memory is a White stimulus set (the 'Boston' CFMT), and so we will be determining whether the magnitude of the OEE varies as a function of individual ability for *own-ethnicity* faces. For the Asian sample, our independent measure of face recognition memory is the same White stimulus set ('Boston' CFMT), and so in this case we will be determining whether the magnitude of the OEE varies as a function of individual ability for *other-ethnicity* faces. Thus this work will consider the OEE in a manner never yet attempted – it will ask does the size of the OEE vary across a given population when considered either across the distribution of own-ethnicity performance (in three different large White samples) or across the distribution of other-ethnicity face recognition performance (in four different large Asian samples).

Study Aims and Implications

In summary, the primary aim of this study was to investigate the OEE across within population distributions of own- and other-ethnicity face recognition performance. For the most part, previous work has often focused on 'extremes' of performance with either own-ethnicity (i.e., very poor performers – DP or very good performers – super recognisers) or other-ethnicity faces (i.e., other-ethnicity blindness), and we have raised various methodological issues with several previous studies. Instead of (somewhat arbitrary) comparisons of performance across sub-groups of a given population, we have taken the approach of considering the pattern and magnitude of the OEE across *all levels* of face recognition ability. Thus allowing us for the first time to determine whether this OEE pattern might in some way vary in size as a function of within population individual variance for *own-ethnicity* faces on the one hand and for *other-ethnicity* faces on the other hand (whilst also controlling for social contact).

Our findings will have interesting implications – if it is determined that the magnitude of the OEE for individuals in a given population is in fact impacted by their relative performance as indexed at baseline by an own- or other-ethnicity face measure, this has consequences for future studies of the OEE going forward (since they must take this into account). However, if it is determined that the magnitude of the OEE remains constant across both distributions of performance, this would provide interesting evidence of the universality of the OEE in face recognition performance. Thus we believe that understanding the degree to which the OEE is impacted by within population individual variation will speak both to previous work on the OEE that has been undertaken (see Tables 1 & 2) and to studies of group comparisons of the OEE that have focused on comparisons with participants who perform at the ‘extremes’ of these distributions.

Method

Participants

Eight hundred and fifty-two participants (largely undergraduate students - see Table 4.3) were recruited from universities in their respective countries. Participants were recruited in their respective universities as part of their Psychology course requirement. 28 participants did not complete the study and were therefore their data were excluded from analysis (N=824). Informed consent was acquired prior to the start of the experiment. All participants had normal or corrected to normal vision during test completion. As we sought to consider the OEE and influence of contact, recruiting solely from one country could mean that we are not able to capture differences in the level of contact. We therefore sought to recruit across nations for which we may assume there are varying levels of contact with other ethnicities (e.g. UK has more diverse population than Serbia, and a rural University in China would have less diverse population than South Korea and Japan). Additionally, recruiting from different countries of similar ethnic groups would give us a more diverse sample and increase the generalisability of the findings.

Statement of Ethics

All participants gave written consent forms and were compensated with study credits for participating. This study was approved by the Swansea University Ethics Committee and followed the Declaration of Helsinki (Williams, 2015).

Table 4.3

Participant Count, Age Means and Standard Deviations in the Sample

Sample	Country	Age Mean	Age SD	Female, Male	Total sample
White	Australia	19.54	1.99	71, 31	102
	Britain	18.67	0.93	159, 36	195
	Serbia	20.26	1.49	56, 47	103
Asian	China	19.05	0.95	61, 42	103
	Japan	19.77	1.58	62, 58	120
	South Korea	20.37	1.18	53, 56	109
	Singapore	20.49	1.33	68, 24	92
Grand		19.61	1.51	530, 294	824

Note: Descriptive statistics of the sample cohort shown for each country, ethnic group, and grand total.

Materials

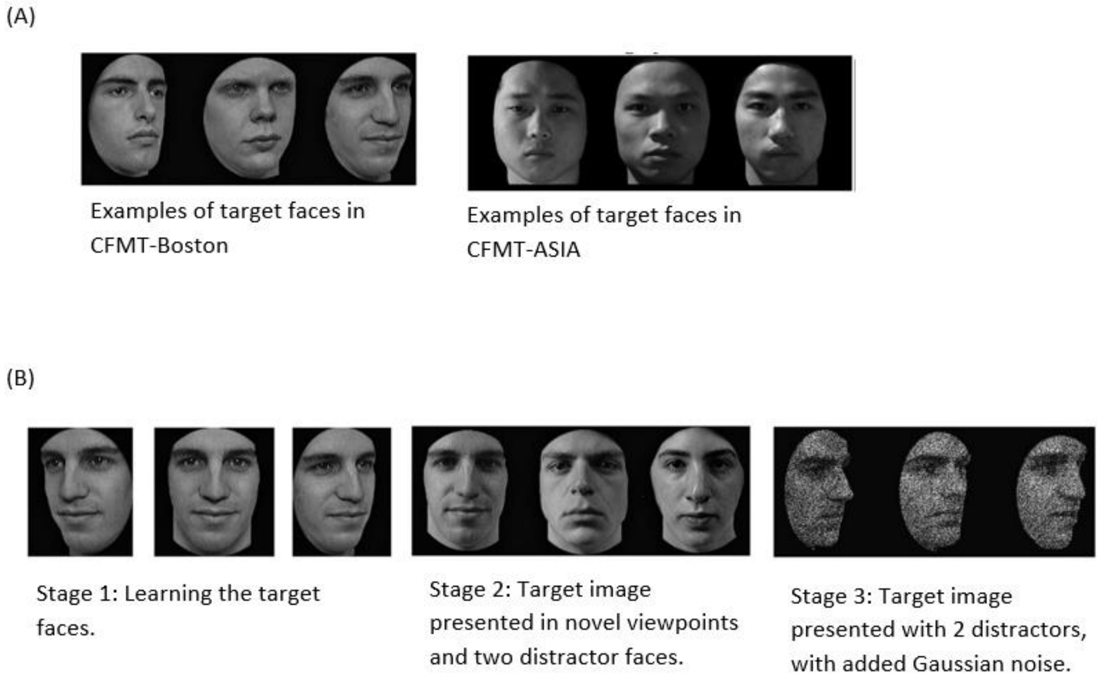
Cambridge Face Memory Test (CFMT) Versions

To estimate the OEE, we employed face recognition tasks that utilise faces from different ethnicities. In this case, we used three well-established versions. First was the original 'Boston' task, which primarily has faces from Harvard University with South European or Middle Eastern features (Duchaine & Nakayama, 2006). Internal reliability (IR) for this version was reported to be between .86-.90 for White participants (Bowles et al., 2009; DeGutis et al., 2013; McKone et al., 2012; Wilmer et al., 2010) and .94 for Asian participants (McKone et al., 2012). Second was the 'Australian', which has a combination of primarily White British-ethnicity faces from Australia, New Zealand, and Scotland (McKone et al., 2011). IR for this version was reported to be between .88-.89 for Whites (McKone et al., 2011; Horry et al., 2015) and .85 for Asians (Horry et al., 2015). Finally, the 'Asian', which primarily has Han-Chinese faces (McKone et al., 2012). IR for this version was reported to be .88-.90 for Asian participants (McKone et al., 2017; Horry et al., 2015) and between .77-.89 for White participants (Horry et al., 2015; McKone et al., 2012; DeGutis et al., 2013). Overall, these studies demonstrate that the different versions of the CFMT are reliable in detecting OEE, as given by the high internal reliability found from the tasks as well as the similarity in difficulty levels across the tests (McKone et al., 2011; McKone et al., 2012).

All CFMTs followed the original procedure outlined by Duchaine and Nakayama (2006), shown in Figure 4.1. All faces were greyscale images of males, with hair cut-out. All versions had three phases. (1) Learn (18 trials; three target faces) – participants were shown the target faces in three views (left, front, right) and were asked to identify the target in a triad (one target and two distractors). (2) Novel (30 trials, six target faces) – participants were shown the target faces in different lighting or viewpoint in a triad with two distractors. Finally, (3) Noise (24 trials) was similar to the Novel phase, but with Gaussian noise added to increase the difficulty of the task. Between each phase, all six target images were presented in front view to the participants for 20 seconds as a reminder. For each test version, accuracy of identifying the target faces was recorded for every phase and they were summed to obtain total accuracy (72 trials). Therefore, the higher the score, the better one's facial recognition ability. Each of the CFMTs was presented to participants in a set of three different orders (balancing which CFMT was seen first), and in line with previous findings (McKone et al., 2012), no significant differences between presentation orders was found (see Supplementary Materials).

Figure 4.1

Phases of Cambridge Face Memory Test



Note: (A) Examples of target faces in CFMT-Boston (Duchaine & Nakayama, 2006) and CFMT-Asia (McKone et al, 2012). (B) Illustrative images for all CFMT procedures. For full details of the procedures, see Duchaine and Nakayama (2006).

Social Contact Scale (Walker & Hewstone, 2006)

To measure self-reported contact, we used a ten item, 5-point Likert questionnaire. Item 1 asked how many people from the other ethnicity participants knew - Up to 2, Up to 5, Up to 8, Up to 10, Up to 12. Items 2-5 pertain to the social component of the questionnaire, which asked how much contact participants have with the other ethnicity, e.g. 'I often spend time with East Asian (White) people, using the following scale: *strongly agree, sort of agree, not sure, sort of disagree, strongly disagree*. Items 6-10 pertain to the individuation component, which asked participants how often they engaged with the other ethnicity, e.g. 'I have looked after or helped a South Asian (White) friend when someone was causing them trouble or being mean to them', using the following scale: *very often, quite often, sometimes,* *hardly ever and never.* The latter two subscales were scored so that lower values indicate higher levels of the measure, while the first subscale simply counts the number of people from other ethnicity group the person knows. To make analyses more straightforward, we reverse scored Social and Individuation components.

Table 4.4 presents the contact scores for this study; it is clear that average contact scores for both measures were largely quite low (perhaps surprising given our sampling across different countries), and variability in contact within populations was also reasonably small (social and individuation contact – see Walker & Hewstone, 2006). Therefore our contact measure was collapsed – and we used overall mean contact scores for the subsequent analyses, with higher scores representing more contact (individual components are more fully explored in the Supplementary Materials).

Procedure

Participants were recruited in their respective Universities as part of their Psychology course and completed the study in the laboratory. Participants were provided with a Participant Information Sheet, and informed consent was acquired prior to commencing of the study. All participants completed the Social Contact questionnaire (Walker & Hewstone, 2006) before starting the battery of CFMTs.

The computer tasks were presented using a bespoke programme constructed by the department's software technician, following the methods outlined for the CFMT (Duchaine & Nakayama, 2006). The order of CFMTs was counterbalanced for each participant to reduce order effects (see Supplementary Materials for further analysis). Following completion, participants were thanked for their time and awarded course credits.

Table 4.4

Mean Scores and Standard Deviations of Contact Scores

	Know		Social		Individuation		Mean Contact	
Country	M	SD	M	SD	M	SD	M	SD
Australia	1.66	0.97	1.91	1.02	2.34	1.16	1.97	0.83
Britain	1.5	0.76	1.56	0.86	1.85	0.99	1.63	0.67
China	1.07	0.25	1.38	0.81	2.13	0.35	1.53	0.33
Japan	1.51	0.84	1.16	0.4	1.24	0.53	1.3	0.46
Korea	1.85	1.5	1.28	0.57	1.47	0.76	1.53	0.61
Serbia	1.05	0.26	1.12	0.44	1.16	0.49	1.11	0.3
Singapore	1.3	0.72	1.3	0.47	1.7	0.76	1.43	0.45

Note: The means for the number of people known (Q1), social (Q2-5) and individuation (Q6-10) components of the Social Contact Scale (SCS, Walker & Hewstone, 2006) used in this study did not show significant variance, allowing the authors to collapse the scores to create a composite contact measure which was used in the subsequent analyses. The original scores for social and individuation components of the SCS were inversed, i.e. higher scores mean lower contact, however, for the linear model analysis, we needed the scores across all variables to be in the same direction, e.g. higher scores mean better recognition skills and higher contact with other-ethnicity group. Therefore, items 2-10 in the SCS were reverse scored, and the three scales were averaged together to create a Mean Contact score where higher scores reflect higher contact.

Design and Analytic Strategy

To address our questions, we built a statistical model that allows us to simultaneously estimate the size of the OEE, the effect of social contact, and independent own- or other-ethnicity recognition performance on CFMT scores. Importantly, it allows us to estimate the interactions between these variables, revealing how the magnitude of the OEE is affected by other variables. For example, it is possible that the size of an individual's OEE depends on their own-ethnicity or other-ethnicity recognition ability, their amount of social contact, or both. Here, we build two separate models to test these effects in our White (*n* = 400) and Asian (*n* = 424) sample of participants, respectively.

To estimate these effects, we utilised a linear mixed regression model, with three main predictors and the full set of interactions between them. Our model structure is as follows, with exposition on the predictors and their interpretation:

$$Y_{si} = (\beta_0 + S_0) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1 X_2 X_3$$

Where B_0 represents the own-ethnicity test scores, S_0 pertains to participant error, X_1 represents the difference between own-ethnicity and other-ethnicity test score, i.e. OEE, X_2 represents Boston scores, and X_3 represents average social contact scores.

For both models, we z-scored standardised both the Boston CFMT and Social Contact scores across all the available data (separately for Asian and White participants). This meant that our models are easily interpretable. The intercept, β_0, represents the average score on the reference-coded CFMT task (for the White model, the Australia CFMT, and for the Asian model, the Asia CFMT). The random intercept, S_0, is estimated per-participant, and accounts for the fact that the Asian and Australia CFMT scores are sampled from the same individual. They thus represent the offset from the overall intercept. Models were estimated using lme4 in R (Bates et al., 2015).

The dependent measure here are the scores on the Australia and Asia CFMTs, collapsed into a single vector of scores, nested within participants as a repeated measure. For example, the *i*th score may represent the score on the Australia CFMT for participant *s*. We aimed to predict these scores as a function of the following inputs.

The coefficient $\beta_1 X_1$ is the effect of a categorical variable that coded the CFMT task that a given score was taken from – that is, the Australia or the Asia CFMT. For our model fitted to White data, the Australia CFMT was coded with zero (i.e., was designated the reference category) and the Asia CFMT coded as one. For the model fitted to Asian data, this was reversed. This has the effect of making the own-ethnicity CFMT task the baseline or reference measure. We labelled this the Face Memory Test coefficient (FMT). Importantly, when estimated, this coefficient represents the OEE, measuring the differences between the scores of the Australia and Asia CFMTs. A useful conceptualisation of this coefficient, which is the crux of our model, is that it allows us to fit two slopes simultaneously to the data – one for the Australian CFMT scores, and one for the Asian CFMT scores. For example, these two slopes can run parallel to one another or move in different directions, if an interaction is present. This allows us to negate issues of difference scores or the use of residuals that are common, as they have undesirable statistical properties and bias estimates of effects ((DeGutis et al., 2013a; Freckleton, 2002; McElreath, 2020). It also ensures the difference between the CFMT tasks is estimated simultaneously with other predictors, and thus is not the same as simply subtracting one CFMT from the other.

The coefficient $\beta_2 X_2$ represents the scores on the Boston CFMT. For our White participants, this is taken as an independent own-ethnicity performance measure that may predict the dependent measure, and conversely for our Asian participants, this coefficient represents an independent other-ethnicity performance measure. The coefficient $\beta_3 X_3$

represents the average scores on the Social Contact scale, with higher values representing more contact with individuals of different ethnicities.

It follows that the coefficient of $\beta_4 X_1 X_2$ represents the interaction between FMT and scores on the Boston CFMT. Thus, this coefficient can represent a different slope between the Australia and Asia CFMTs. If, for example, individuals with higher own-ethnicity recognition (or other-ethnicity, for Asian participants) ability exhibit a smaller OEE, this coefficient would represent such an effect, with the slopes for the Australia and Asia initially being far apart but coming closer together as Boston scores increase. Very similarly, the coefficient of $\beta_5 X_1 X_3$ represents the same effect but with Social Contact scores – if individuals with higher contact exhibit a smaller OEE, this coefficient would represent this difference. The coefficient of $\beta_6 X_2 X_3$ allows individuals with higher scores on the Boston and Social Contact measures to have different scores on *either* the Asia or Australia CFMTs, which is of less theoretical interest. However, this term is included in the model, as we wish to test the three-way interaction ($\beta_7 X_1 X_2 X_3$) between FMT, Boston, and Social Contact – that is, whether individuals with high or low scores on *both* the Boston CFMT and Social Contact measure exhibit a larger or smaller OEE. Interaction variables are taken as the multiplication of their components.

We conducted a power analysis via simulation to estimate the smallest effect we could detect with our design, which was between .20 and .25 for each coefficient (i.e., a one unit change in the predictor equates to a .20-.25 unit change in Asia or Australia CFMT scores) at 80% power, which is a very small effect (Appendix I; also see Supplementary Materials for full details).

Results

Descriptive Statistics

Table 4.5 presents the mean scores and standard deviations for each of the CFMT versions for each country cohort and collapsed by ethnicity. Overall, White participants scored higher than Asian participants in the two White versions of the test, while Asian participants scored higher in the Asian version of the test.

Table 4.5

Descriptive Statistics for the Country Cohort on CFMT Measures.

Country		Asian		Australian		Boston	
	N	Mean	SD	Mean	SD	Mean	SD
Australia	102	50.9	7.88	55.15	7.47	55.94	7.87
Serbia	103	51.04	8.22	57.69	7.35	58.14	8.61
UK	195	52.2	8.61	54.37	7.66	55.19	8.45
Overall White	400	51.57	8.33	55.42	7.64	56.14	8.41
China	103	56.59	8.31	48.73	7.52	47.32	8.9
Japan	120	56.73	7.5	48.78	7.41	51.82	7.43
South Korea	109	55.5	8.98	52.72	8.29	51.44	8.01
Singapore	93	55.11	7.55	50.65	8	49.08	8
Overall Asian	424	56.03	8.11	50.19	7.95	50.04	8.26
Total	824	53.87	8.51	52.73	8.22	53	8.87

Note: Mean correct scores (over 72 items; chance performance is ≤ 24) and standard deviations for the three CFMT versions used in this study for each country cohort and ethnic groups.

Reliability Analysis

To determine the internal reliabilities of our measurements we undertook several analyses. Firstly, our selection of the CFMT tests was (as we established earlier) largely motivated by previous work that has established their high measurement reliability. Nonetheless, we checked the internal reliabilities for each of the CFMT versions across our sample, and determined Cronbach’s alpha values: Boston CFMT a = .917, Australia CFMT a

= .873; and Asia CFMT *a* = .846. Split into the two ethnicity groups, our analysis yielded similar *a* values, for Whites: Boston CFMT a = .933, Australia CFMT *a* = .863, and Asia CFMT *a* = .820; and for Asians: Boston CFMT *a* = .883, Australia CFMT *a* = .851, and Asia CFMT *a* =.843. These correspond with the reports of internal reliability values in other studies (see Horry et al., 2015; McKone et al., 2012), confirming that the use of these CFMT versions was appropriate.

However, although each independent test shows high internal reliability, the OEE, which is derived in our models as a covariate-adjusted difference between the two measures, may not be (Sunday et al., 2017, Ross et al., 2015). No study has yet investigated the internal reliability of the OEE itself, and thus it remains an open question as to whether this measurement may in fact be far noisier than has previously been assumed, and thus throwing doubt on findings focused on individual performance (e.g., the lack of interactions found between OEEs and other variables may be due to the noise in the measurement). However, it is also important to note that the linear mixed model approach used in our analysis can closely incorporate individual performances on the CFMTs by estimating individual offsets from the global intercept, which was both a major motivation and advantage of choosing the analytical approach we presented here.

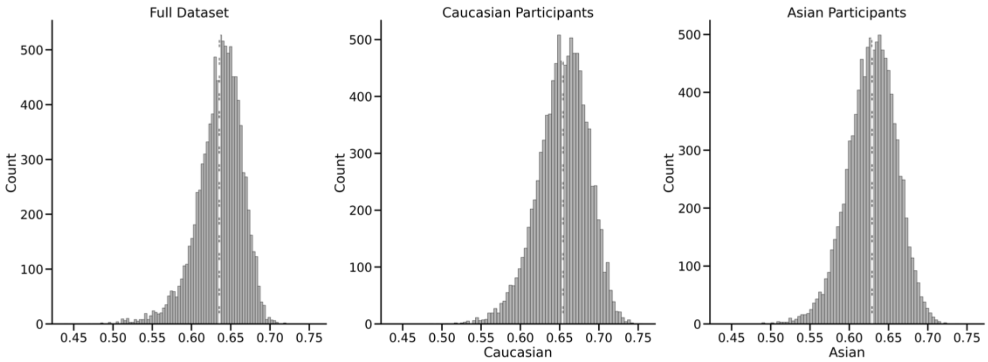
In order to explore the internal consistency of the OEE, we first divided the items into the phases as described by Duchaine and Nakayama (2006), i.e. Learn (items 1-18), Novel, (items 19-48), and Noise (items 48-72). Within these phases, we randomly split the items into two equal size groups – e.g the first nine random items from Learn phase were labelled *Learn 1*, the first fifteen random items from Novel phase were labelled *Novel 1*, and the first twelve random items from Noise phase were labelled *Noise 1,* and so forth. Using a bootstrap resampling approach, we created these random splits 9,999 times, and summed the scores within the each split across the different phases, which created composite scores for the half of the test, i.e. Learn 1, Novel 1, and Noise 1 were collapsed together to make a composite score - Split 1.

Using the split-halves mentioned above, we took the difference for each of the test halves between the corresponding own-ethnicity and other-ethnicity score for our samples, e.g. for Asian samples, we used Asia Split 1 – Australia Split 1 and for White samples, we used Australia Split 1 – Asia Split 1 to create an OEE 1 score, and so on. Using Spearman-Brown correction, we analysed the reliability of the OEE scores for each of the split pairs, generating a distribution of split-half reliability coefficients. We tested this within the full sample, and within each participant ethnicity subsample. The means were highly similar, 0.64 for the full sample, 0.65 for the White sample, and 0.63 for the Asian sample. The distributions are shown in Figure 4.2.

Although the mean *a* values for the OEEs are lower than that of the CFMT measures on their own, they are still within acceptable levels (Ursachi et al., 2015). Nonetheless, it is striking that the OEE measure is indeed lower in internal reliability, and this demonstrates that although our individual measures did have very high reliability, the *difference* between these measures (the reported OEE) was lower. This indicates for the first time that work exploring individual differences and the OEE, must utilise very reliable face recognition measures across ethnicity and report internal reliability scores for the OEE they have determined.

Figure 4.2

Reliability analysis for OEE scores



Note: Distributions of the reliability of the OEE generated by bootstrap resampling. The average of each distribution is marked by the dashed white line.

White Model

The estimated coefficients for the model fit to the data from White participants are shown in Table 4.6. Only two predictors were statistically significant. The first was the FMT, which estimated the difference between the Australia (coded zero) and Asia CFMTs, b = -3.97, $t(395.99) = 9.76$, $p < .001$, thus representing a significant OEE effect. This is directly interpretable as the Asia CFMT having, on average, a lower score than the Australia CFMT by 3.97 points. Second was the Boston CFMT predictor, which here represented an independent measure of own-ethnicity performance, $b = 5.03$, $t(731.19) = 14.98$, $p < .001$. Thus, as individual scores on the Boston CFMT increased by one standard deviation, on average, scores on Australia CFMT increased by 5.03 points. There was no significant effect of social contact, and notably, we observed no significant interactions between the FMT predictor or the Boston predictor. This indicates that while the scores on the Asia CFMT are lower than the Australia CFMT, the slope changes by more or less the same amount for each with increasing own-ethnicity recognition ability (measured by the interaction between the

FMT and Boston coefficient, *b* = 0.16) or social contact (measured by the interaction between FMT and social contact coefficient, *b* = 0.55). The interaction between all three predictors was also not significant. Despite this, the variance explained by the fixed effects alone was relatively high, marginal R^2 = .40 (Nakagawa & Schielzeth, 2013).

Table 4.6

Parameter Estimates for the White Participants' Model

Parameter	*b* [95% CI]	*SE*	*t*-value	*p*-value
Intercept	53.64 [52.97, 54.31]	0.34	157.09	< .001
FMT				< .001
(0 = Australia)	-3.98 [-4.77, -3.18]	0.41	-9.76	
Boston	5.03 [4.37, 5.69]	0.34	14.98	< .001
Contact	0.03 [-0.53, 0.6]	0.29	0.11	0.911
Boston * FMT	0.16 [-0.63, 0.94]	0.4	0.39	0.698
Contact * FMT	0.55 [-0.12, 1.23]	0.34	1.61	0.107
Boston * Contact	-0.14 [-0.73, 0.46]	0.31	-0.45	0.654
FMT * Boston * Contact	0.24 [-0.48, 0.95]	0.36	0.65	0.516

Note: Estimates for the White model showing FMT scores significantly influence the variability in the scores. Boston-CFMT scores was used as own-ethnicity measure. Contact scores do not show significant contribution in the FMT scores, indicating that level of contact in this study do not influence other-ethnicity face recognition.

Asian Participants

The coefficients for the model fit to the data from Asian participants are displayed in Table 4.7. Again, only two predictors were significant – the FMT, which here estimated the difference between the Asian (this time coded as zero) and the Australia CFMTs, b = -5.86, $t(420) = 15.23$, $p < .001$, demonstrating a significant OEE effect. This means that for Asian participants, scores on the Australia CFMT were on average 5.85 points lower than for the Asia CFMT. Additionally, there was a significant coefficient for the Boston CFMT score, which here represented a measure of independent other-ethnicity performance, b = 5.31, $t(757.82) = 15.83$, $p < .001$. Here, this represents the pattern that a one standard deviation increase in *other*-ethnicity recognition performance is associated with, on average, a change of 5.31 units in own-ethnicity performance as measured by the Asia CFMT. The lack of significant interaction between the FMT and Boston predictor here (b = -0.16) indicates that this relationship is practically equivalent between the Boston and the Australia CFMT scores. The variance explained in the Australia and Asia CFMT scores was as similarly high as the model built on White data, marginal $R^2 = .45$.

Table 4.7

Parameter Estimates for the Asian Participants' Model

Parameter	*b* [95% CI]	*SE*	*t*-value	*p*-value
Intercept	57.77 [57.11, 58.42]	0.33	173.91	< .001
FMT				< .001
(0 = Asia)	-5.86 [-6.61, -5.1]	0.38	-15.23	
Boston	5.31 [4.65, 5.97]	0.34	15.77	< .001
Contact	-0.4 [-1.2, 0.4]	0.41	-0.97	0.331
FMT * Boston	-0.17 [-0.93, 0.6]	0.39	-0.42	0.671
FMT * Contact	0.23 [-0.7, 1.16]	0.47	0.49	0.622
Boston * Contact	-0.1 [-0.96, 0.76]	0.44	-0.23	0.822
FMT * Boston * Contact	-0.44 [-1.44, 0.55]	0.51	-0.87	0.382

Note: Estimates for the Asian model showing FMT scores significantly influence the variability in facial recognition scores. Boston-CFMT scores were used as other-ethnicity measure. Similar to the White model, contact scores do not show significant contribution in the FMT scores, indicating that level of contact in this study does not influence other-ethnicity face recognition.

Examining Model Predictions

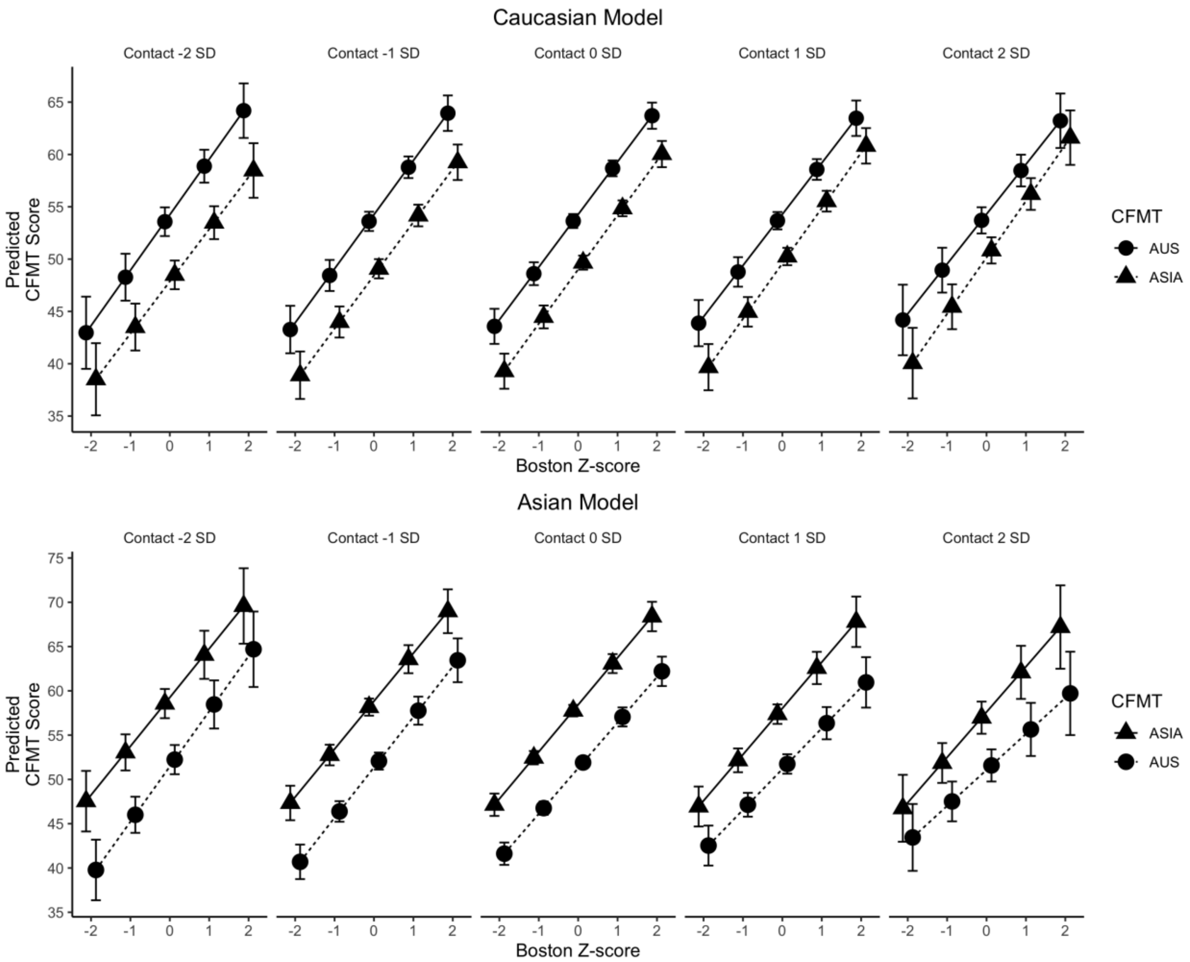
The estimated statistical models thus far demonstrate significant OEEs, and an influence of the Boston CFMT scores on the Australia and Asian CFMTs, whether that represents an own- or other-ethnicity measure of recognition performance. Examining the predictions made by the models is key to their interpretation. As the models essentially fit a separate slope for the Australia and Asia CFMTs simultaneously (coded by the FMT coefficient), and by allowing these separate slopes to interact with the other predictors, we are able to examine the likely OEE at high and low levels of the Boston CFMT and social contact scores. Figure 4.3 demonstrates the predictions of each model, derived by using the models to predict scores separately for the Australia and Asia CFMTs for hypothetical participants with varying scores on the Boston and social contact measures. The figure makes it clear that the OEE – the difference between the slopes of the Asia and Australia CFMTs – is consistent at all various combinations of low and high Boston and social contact measures, evaluated here at scores ranging from ±2SDs on the predictors. Indeed, this consistency is clear from the lack of interactions in the model. These predictions thus allow us to examine how individuals with excellent or very poor own- or other-ethnicity performance and high or low levels of contact might do on tests of own- or other-ethnicity performance, but with information estimated from a full range of data as opposed to smaller samples.

Further Considerations and Robustness Checks

An additional possible source of variability we have not considered so far is that participants were sampled from different countries within our models – that is, not all White and Asian participants were from the same countries, as described in the method. It is therefore possible that variation within those countries in terms of face processing ability or otherwise could have an impact on our results.

Figure 4.3

OEE magnitude in White (top) and Asian (bottom) Participant Models



Note: Predictions of the White participants model top, by varying levels of contact (separate axes) and the Boston CFMT (X-axis). Error bars represent 95% confidence intervals of the FMT scores. For the top axis, the Boston represents an independent measure of own-ethnicity performance, and for the bottom axis, it represents a measure of other-ethnicity performance.

To test this, we recreated our two models, but this time included an additional random intercept for country alongside that of participants representing their country of origin (i.e. whether Asian participants were from South Korea, China, Japan, or Singapore, and White participants were from the UK, Australia, or Serbia). Treating country of origin as a random factor is appropriate as we wish to make inferences about countries that are generally Asian or White, and our data represents only a sample of the possible countries that fit this profile. We compared these new models to the original models used in the analyses without the additional random intercept using a likelihood ratio test, to confirm whether the more complex model had a better fit to the data. For both the White model, the likelihood ratio test was not significant - χ^2(1) = 0.00, p = .999. For the Asian model, this test was significant, χ^2(1) = 6.97, p = .008 – indicating that country of origin did improve the fit to the data. Examining the AIC of the model showed a small change between models (without = 5511.5, with = 5506.5), and the marginal R^2 of the model increased by 0.01%, from 0.446 to 0.456. The overall pattern of results were unchanged.

We also estimated our models by swapping the positions of the Australia and Boston CFMTs, by using Australia scores as the independent measure of performance and Boston scores being predicted alongside the Asian CFMT. No differences in the overall conclusions were found. We also sought to examine the stability of the OEE effect by using random split-half resampling techniques, which showed the magnitude of the OEE was very consistent. See the Supplementary Materials for details.

General Discussion

The current study represents the largest ever undertaken investigating the OEE across within population distributions of own-ethnicity and other-ethnicity face recognition performance. Our results demonstrated the following key findings:

1. Our study finds a robust OEE effect in both Asian and White samples, replicating previous studies of the OEE using the CFMT paradigm.
2. Our modelling approach allowed us to test whether the magnitude of the OEE varied in relation to individual levels of own ethnicity OR other ethnicity ability. It did not. Our model therefore shows a remarkably consistent impact of the OEE across the entire range of the populations investigated.
3. Our approach also allows us to test whether social contact impacts the OEE – and we found no evidence for this. But, it is of note that in our case the range of scores on our measure of social contact was not substantial (with contact scores being relatively low), despite the fact that we sampled across a number of different countries. In any case, a meta-analysis of OEE research articles demonstrated that self-report assessments of other-ethnicity contact explained less than 3% of the total variance in the OEE (Meissner & Brigham, 2001), indicating that factors beyond the kind of measures we have implemented on this issue may be more key to modulating the OEE in individual performance (e.g., such as bilingualism; Burns et al., 2019).
4. Our model also indicates no combination of these factors appear to impact scores on CFMTs of own or other ethnicity (i.e., no evidence of a two or three-way interaction).

In summary, this work demonstrates that an OEE is a consistent feature of face recognition performance for participants sampled across a variety of nations and cultures –

and in addition this differentiation in performance, which could be characterised as either an own-ethnicity advantage or an other-ethnicity disadvantage, is consistent in magnitude across all individuals. Our finding that individuals at ‘extremes’ of own ethnicity performance show an equivalent OEE is consistent with previous work undertaken with groups of individuals classified as *developmental prosopagnosia* (Cenac et al., 2019; DeGutis, et al., 2011) and *super-recognisers* (Bate et al., 2018; Robertson et al., 2019). In that in both cases, the evidence emerging from testing of such populations suggests both groups show OEEs; our work builds on this by further indicating that the quality of this OEE is indeed no different from that of individuals at any other points on the distribution of own-ethnicity recognition ability. This work therefore supports the view that the two groups differ from controls in a quantitative manner.

However, our findings, at least initially, may be seen to be contrary to those of Wan et al., (2017) and their reports of individuals with putative *other-ethnicity blindness*, in that we found no evidence that the quality of the OEE differed even with individuals who performed at the lowest end of the distribution of other-ethnicity face recognition accuracy. It should be noted that an advantage of our work is that by considering this issue across the full distribution of population performance, we avoided issues around classification ‘cut-off’ (i.e., 2 SDs) discussed earlier. In the Wan et al. (2017) study, poor performers were selected on the basis of a somewhat arbitrary statistical distinction, albeit an approach often used by others – and this classification was not confirmed with any further testing. Thus, it remains possible that in their work, the observed differences between two tests could be simply due to the fact that the same test was used as the classifier *and* as the comparator. It is therefore likely that these differences in our approaches may explain the potentially contrary findings.

However, it should be noted that social contact was quite limited in variability in all our participant cohorts – and we suggest this may explain why no effect of social contact was seen despite previous reports indicating an influence (e.g., Zhou et al., 2019). This key issue may also explain our initially contrary findings to Wan et al. (2017); where it is possible that if across a test group there is considerable variability in social contact, a small sub-group may have much lower social contact than the rest of the group in general. If so, that sub-group might perform much worse relative to the rest of the group, and thus reach the classification of 2SDs below the mean. Importantly, Wan et al., (2017) reported that of the 37 participants who met criteria for being very poor with other-ethnicity faces (i.e., 2SDs below mean accuracy), 36/37 had reported low contact with individuals of the other ethnicity, and thus it's likely what is driving the presence of very poor other-ethnicity face accuracy is a process linked to social contact rather than face processing in general. Given the low variability of contact in our samples, we would suggest this could explain why we found no evidence of any individuals with a relative *other-ethnicity blindness.* We therefore agree with the conclusions of Wan et al. (2017) that the presence of individuals who would meet such a criteria is likely dependent on the relative individual variability of social contact for the group tested and not to do with the base level of face recognition performance generally. As a consequence, we are reluctant to draw the more general conclusion that social contact does not influence the magnitude of OEE and it would be interesting to test this in a sample with a much more varied pattern of social contact than we were able to obtain.

Furthermore, it is important to stress that our modern society allows for more varied types of social contact than the face-to-face interactions that traditionally defined 'social contact', as measured by the questionnaire used in the current study. For example, East-Asian pop bands have been increasing in popularity in the Western media through films, music videos, and advertisements, among others, and vice versa. This type of cultural contact is not covered in the contact measure that was used in this study, but could potentially have a considerable impact on individuals' ability to recognise and discriminate between faces of other-ethnicities – simply because they can provide many more opportunities to increase exposure to faces from other ethnicities beyond contact in the traditional sense. We would therefore suggest that this needs to be incorporated in future studies that aim to measure contact with other ethnic groups.

With this in mind, consider the case of individuals who appear to perform very poorly with own-ethnicity faces; what is striking from our work is that *despite* the issues with own-ethnicity faces such individuals have, they manifest an OEE commensurate with the rest of the population. This clearly indicates that whatever unpins the challenges faced by such individuals with faces of their own ethnicity, this is independent of the OEE. We would speculate that this reflects the fact that all individuals, independent of natural face recognition ability, can still gain *some* visual learnt experience from own-ethnicity faces. This learnt experience underpins a remaining advantage for own-ethnicity faces (or disadvantage for other-ethnicity faces) and hence an OEE is consistently present. We therefore interpret our findings in a similar manner to that of Cenac et al. (2019) – namely, that face processing is underpinned by two key factors: on the one hand there is a form of *inherited susceptibility* to generally poor face processing ability and on the other hand there is a *visual learnt experience* factor that can drive differential performance across types of face ethnicity. What our work clearly demonstrates is that variability on the first of these factors has no impact on the magnitude of the OEE in face recognition memory – regardless of an inherited susceptibility to being generally poor or very good with faces, all other things being equal, there is always a consistent and universal ‘fixed’ benefit/cost to recognition memory across faces of differing ethnicities. The consistent nature of this OEE effect also implies that if inherited susceptibility to generally very good face processing ability is the case for a given individual, although that person will perform more poorly with other-ethnicity faces, they will still be largely superior to all other individuals in that same population. Thus making the case that in practical terms, the best persons to employ for passport control will always be superior face recognisers in a given population.

Earlier in the introduction we mentioned that previous work has demonstrated that the OEE can be moderated through participant training with other-ethnicity faces (e.g., Lebrecht et al., 2009). An account for this training effect has been linked to the suggestion that differential performance across face ethnicities may be underpinned by the degree of configural or featural processing being used. That is, it is likely that own-ethnicity faces, given their high degree of familiarity, implicate a different 'bias' toward configural/holistic versus featural/part-based processing, (Hayward et al., 2008; Zhao et al., 2014; Rhodes et al., 2010). With such an explanation in mind, DeGutis et al. (2011) has suggested that training can mediate attentional 'bias' across own/other ethnicity faces such that it 'boosts' configural processing of other ethnicity faces. Although we did not examine the issue of configural/featural processing, we might speculate that the consistent and 'fixed' OEE pattern we see across all levels of individual ability in our work, is the consequence of this consistent attentional 'bias' across faces of different types.

This raises an interesting future avenue of research regarding the effects of training on the OEE – previous studies have largely considered such effects at the group level (e.g., Tanaka & Pierce, 2009), and we suggest rather taking an individual differences approach – thus exploring the consequences of training across individual variability in own ethnicity face recognition. For example, although the work by DeGutis et al., (2011) speaks to the question of the impact of training for individuals at the lowest end of performance (i.e., developmental prosopagnosia), it would be interesting to explore the consequences of training across all levels of individual ability using a similar approach to that undertaken here. If the OEE reflects a fixed 'cost' of a strategic 'bias' in attentional resource allocation for configural processing across faces of different ethnicities, and training can reduce this 'bias', the prediction should be that all levels of ability would see the same relative reduction in OEE magnitude. Put simply, if the OEE reflects the consistent impact of a strategic attentional 'bias', then it should be possible via training for all types of faces to reach optimal performance commensurate with own-ethnicity face testing for a given individual.

A final consideration is the issue of statistical power and potential measurement error. Our large sample and use of linear mixed models afford greater power, and our power analysis (see SM) indicated that we can comfortably detect changes in CFMT scores as small as .20 - .25 across our predictors. Notably, some of the coefficients in our two models were estimated to be below this threshold, and as such, we cannot explicitly rule out the absence of an effect here (i.e., there may be an interaction between the OEE and own-ethnicity performance) that is too small to detect with our sample size. An important factor that may contribute to increasing 'noise' in our pattern of results is that the OEE itself has low measurement reliability. As a means of mitigating the potential contribution of poor measurement, we selected three face recognition measures with well-established reliability (see Horry et al., 2015; McKone et al., 2012), and this was also confirmed in our own analyses. However, just because an individual measure is reliable, does not therefore entail that the *product* of two such measures (that is the difference between the two, which is how the OEE is defined) is necessarily also reliable (see Ross et al., 2015; Sunday et al., 2017). As a consequence we undertook reliability analyses on our OEE effect, and report that indeed reliability levels are lower than is seen for the individual tests themselves, but still sufficiently high for us to have some confidence in our interpretation of the lack of interactions seen in our analyses. This is in fact the first time such reliability analyses have been undertaken and they provide an important caveat to the findings of OEE studies both past, present and future – since if one assumes that our pattern is often the case (that the OEE is less reliable than the individual tests from which it is computed), the individual tests used *must* be very high in reliability in the first place and it would be good practice for OEE reliability to be reported if not.

A final point is that, for the interaction terms that did not reach statistical significance, the coefficient estimates were very small. Since estimates of coefficients using least squares are unbiased (i.e., on average, the coefficients will represent the effect in the population), and our sample is large enough to provide a stable estimate, we would tentatively conclude that any interaction terms between the OEE and other factors are likely to be small in practical terms. For example, for White participants, the three-way interaction coefficient was .24 units, which is much less than a single unit on a given CFMT, and thus unlikely to translate into a qualitative ‘real-world’ difference in recognizing faces of another ethnicity. However, we also recognize that the issue of measurement error is at play here, and this difference could be larger than this. We did however build our statistical models for our analysis to mitigate these limitations, as the inclusion of the random intercept term means the fixed effect of the OEE is scaffolded by individual level intercepts, and therefore we are confident that such issues were minimal for our current data.

In sum, the current work is the first to consider the OEE from the perspective of individual variability across a variety of nations and cultures; our message is that the magnitude of the OEE is of a consistent quality across all levels of ability seen both from the perspective of variance on own ethnicity face recognition performance and other ethnicity face recognition performance. These findings are consistent with studies that have focused their attention on sub-groups of individuals at both the bottom (i.e., developmental prosopagnosia, DeGutis et al., 2011) and top (i.e., super-face recognisers, Bate et al., 2018) of the population distribution, in that OEE patterns were also reported in their samples – our work builds on this by demonstrating such effects are by no means qualitatively different.

Intriguingly, given the OEE we found across individuals was consistent in magnitude, we speculate that this is compatible with an attentional ‘bias’ account for the OEE (as suggested by DeGutis et al., 2011) – essentially, the OEE reflects the utilisation of a face-based strategic attentional processing ‘bias’, which incurs a benefit/cost to recognition memory across own/other ethnicity faces. This impact is independent of the general level of face recognition memory for any given individual and thus the OEE remains of consistent magnitude across all levels of ability. It would be interesting for future work to explore this issue further, perhaps by considering the impact of training through the lens of individual variability.

Next steps

From here, we established that there is a consistency of the magnitude of OEE regardless of how good or bad one is at recognising faces from their own ethnic group. Next, I was interested to see whether an OEE will also be observed when individuals evaluate faces on facial attractiveness, health, and age. To capture which areas of the face are being used when individuals make the social judgments, I incorporated eye tracking techniques in the next study.

Chapter 4.2 Looking at a face differently: Other-ethnicity effects and eye movements on evaluation of socially salient traits

Abstract

A consistent finding within facial recognition is better recognition of own-ethnicity-group as compared to others, known as the other-ethnicity effect (OEE). The literature is not clear whether this also happens when we evaluate facial age, attractiveness, and health. Previous studies have shown that as faces appear younger the healthier and attractive they look, but studies on which facial areas are used when judging faces from different ethnicities are scarce. The current study recruited participants from two ethnic groups and presented faces from three ethnic groups to investigate which facial areas are used when we make these judgements and whether an OEE exists. We found a consistent significant effect of facial areas across all tasks, where in general, the Nose and Eye regions receive most attention (duration and relative fixation counts) across tasks. No significant effects were found for Participant Ethnicity nor Target Face Ethnicity. Furthermore, we did not find an OEE, although we did find that the Lips and Glabella areas received different fixation proportions depending on the Target Face Ethnicity and task. These findings suggest that distinct facial areas play a role depending on the task, and that people look differently at faces as a function of their ethnicity, regardless of the observers' *own* ethnicity.

Thank you to Alex Green, Jasper Lok, Alex Woodfield, and Joseph Hutchinson for their contribution to data collection on this chapter.

Within a few seconds of looking at someone's face, a perceiver is able to make judgements on various traits of the target, such as identity (Zebrowitz & Montepare, 2008), emotions and intention (Horstmann, 2003), personality (Fink et al., 2005), attractiveness (Langlois et al., 2000) and apparent health (Jones et al., 2018). These social cues in turn influence our perception about the target, and affects how we behave towards them – that is, do we approach someone or avoid them?

Perceptions of facial attractiveness have received increasing attention from researchers over the years, and literature has consistently found that more attractive people have more favourable social outcomes, such as more success in intimate relationships, better career opportunities, and more lenient sentences compared to less attractive individuals (Langlois et al., 2000; Rhodes, 2006). Facial attractiveness is also highly associated with apparent facial health and age, where attractive faces are also rated as healthier and younger looking compared to less attractive faces; and older faces are rated less attractive and less healthy compared to younger faces (Kwart et al., 2012). Perceptions of facial attractiveness therefore have significant social implications, and it is argued that this may drive the increasing demand in the cosmetic industry, where more people, particularly those aged 40-54, engage in treatments that would make them look younger, and hence, more attractive (American Society of Plastic Surgeons, 2020).

Evolutionary psychology research has demonstrated that there are traits which are consistently associated with attractive and healthy faces, e.g. facial symmetry, averageness, sexual dimorphism (Langlois et al., 2000; Rhodes, 2006; Thornhill & Gangestad, 1999), and more recently, facial adiposity (Tinlin et al., 2013) and colour (Jones et al., 2016). It was argued that these traits are associated with 'good genes' and that expression of these traits in an individual reflects one's genetic quality (Andersson, 1989; Hamilton & Zuk, 1982), and therefore have evolutionary functions.

However, although there may be universal cues to perceptions of facial attractiveness, (and by extension facial health and age), little research has been done to investigate which specific facial areas people look at when they make these judgments on people from their own- and other-ethnicity faces. Research on facial recognition using Cambridge Face Memory Tests (CFMT) of own- and other-ethnicity faces show that individuals are better at remembering faces from their own-ethnicity compared to other-ethnicity, also known as the other-ethnicity effect (OEE, Blais et al., 2008; Cenac et al., 2019; Childs et al., 2021; DeGutis et al., 2013; Duchaine & Nakayama, 2006; Horry et al., 2015; McKone et al., 2012; Wan et al., 2017). Although the CFMT is a standardised measure, this design does not inform us how participants look at the target faces, only that there are significant differences in recognition ability. A method which could give us an insight into the mechanisms that affect recognition performance is using eye-tracking (ET) techniques. ET methods capture eye movements such as *fixation*s, i.e., where we look, and *dwell* times, i.e., how long we look at certain areas, which could inform us which areas of the face are being utilised by participants for specific tasks.

Recently, Arizpe and colleagues (2016) used eye-tracking methods on facial recognition tasks. They asked Caucasian males to study faces from Asian, Black and Caucasian ethnicities and investigated their recognition performance, as well as exploring whether the participants looked differently at the faces depending on the target ethnicity. In this study, they found a degree of own-ethnicity faces bias, where participants looked at the eye areas of Caucasians *more* compared to Asian faces, and that they looked at the Nose areas of Caucasian faces *less* than both Asian and Black faces, and looked *less* at mouth areas for Black faces. That is, Caucasian participants looked at eye regions more for own-ethnicity faces, and mouth and nose fixations more for other-ethnicities. Similar findings were found by Stelter, Rommel and Degner (2021), where they presented Caucasian participants with Middle Eastern, Black, and Asian faces. They found that White participants fixated more on eye regions of White faces compared to other ethnicities, but no significant differences in the durations for target faces overall. These two studies demonstrate that Caucasian participants look at different facial areas when looking at own- and other-ethnicity faces.

These findings find support in those of Blais et al. (2008), where they presented Asian and Caucasian faces to Asian and Caucasian participants and asked them to study the faces, identify which faces were presented previously, and to categorize the faces based on race. Here, they found that Asian participants looked at noses more and Caucasian participants looked more on eye regions for both target faces. They argued that the difference in areas of interest is due to the difference in culture between the two ethnicities, i.e. it is considered rude to make excessive eye-contact in Asian culture, whereas this is more acceptable in the Caucasian culture.

The current study therefore aims to add to the literature by incorporating three target face ethnicities (Asian, Black, and White) and two participant ethnicities (Asian and Caucasian) and extending the task to social judgments, e.g. age estimation, attractiveness, and health perceptions.

OEE in social perception

Research on face perception has shown a variation in how people perceive different faces. As mentioned previously, OEE is a phenomenon where individuals are better at recognizing faces from their own-ethnicity compared to those from other-ethnicity (Blais et al., 2008; Cenac et al., 2019; Childs et al., 2021; DeGutis et al., 2013; Duchaine & Nakayama, 2006; Horry et al., 2015; McKone et al., 2012; Wan et al., 2017), and this has been demonstrated to extend to judgments of attractiveness (Darrach et al., 2019), and age perception (Dehon & Brédart, 2001; Lick & Johnson, 2018), where people give more positive ratings to faces from their own-ethnicity compared to members of other-ethnicity.

The OEE in identity using recognition tasks has been widely supported using face memory tasks (Childs et al., 2021; McKone et al., 2012; Wan et al., 2017), perceptual tasks (Cenac et al, 2019; Horry et al., 2015), and eye-tracking tasks (Arizpe et al., 2016; Blais et al., 2008). The literature on attractiveness and health have mixed conclusions, where on the one hand, some studies found that people give higher attractiveness ratings to those who are closer to their ethnicity (Darrach et al., 2019; Potter & Corneille, 2008; Rhodes et al., 2005), however, there also exists an agreement as to which faces are attractive (Coetzee et al., 2014; Kleisner et al., 2017; Rhodes, 2006; Rhodes et al., 2001) and healthy looking (Jones et al., 2018). It is also important to note that although there is a high degree of agreement between cultures as to which faces are attractive, this could also be influenced by familiarity, and some studies have shown that groups differ from which cues they utilise when making these judgements, e.g. Caucasian use face shape cues while Africans use facial colour cues when making judgments of attractiveness for African faces (Coetzee et al, 2014).

Age Perception

Correlational studies have shown that signs of aging, e.g. increased wrinkling in the eye areas and uneven skin tones, were associated with higher age estimations in French (Nkengne et al., 2008), Russian (Merinville et al., 2015), Indian (Merinville et al., 2018), and Chinese samples (Mayes et al., 2010). This implies that there may be universal signs of aging, and that there may be a certain pattern of how aging is manifested in faces, e.g., sagging of upper facial areas to fill lower facial areas (Michaud et al., 2015). However, the rate of this process differs between ethnic groups due to variations in skin properties such as melanin content (Rawlings, 2006), and lifestyle choices such as diet and sun exposure Flament et al., 2015). These differences, therefore, could drive the OEE in age estimation, where people from the same ethnic group may be more familiar with the lifestyle common to their own-group and could therefore adjust their estimates accordingly.

Dehon and Bredart (2002) were the first to demonstrate an OEE in age perception, where they found that Caucasians were more accurate at estimating the age of Caucasian faces compared to African faces. African participants in their sample did not show an OEE, however, and this was attributed to the participants' exposure to Caucasian faces. These results were consistent with the contact hypothesis of OEE, where the more exposure one has to the other-ethnicity, the better they get at recognizing them (Goldstein & Chance, 1985).

Lick and Johnson (2018) on the other hand, proposed that racial categorization has a significant effect in how observers perceive one's age. They argued that certain ethnic groups are associated with more masculine or more feminine traits, and this affects age estimations. Using images from 200 participants, they created morphs of varying traits, from having 'very African' features, to Caucasian and 'very Asian features,' as well as masculine and feminine traits. Overall, they created 40 facial images (20 males, 20 females). Participants were presented the faces individually and asked to estimate their age. They found that faces with more Asian and Caucasian traits were rated younger compared to those with African traits. This study therefore shows that perception of race categorization could have an effect on how we perceive one's age, and as age perception is closely linked with other socially salient traits, this could also have an effect on perception of attractiveness and health. Although this study showed that there are differences on how observers evaluate age of faces from different ethnic groups, they do not provide information as to whether observers are using the same strategy for each face, i.e., are they looking at the same areas of the face across all ethnic groups? The current study aim to address this gap.

Attractiveness Perception

OEE on perception of attractiveness using ET was partially demonstrated by Darrach et al. (2019) where they investigated how Asian, African-American, Caucasian, and Latin-American participants would perceive Caucasian and Latin-American faces before or after they have undergone rhinoplasty. Participants were presented post-surgery images of Caucasian and Latin-American targets, paired with a pre-surgery image of a target with a different race for each trial, e.g., one Caucasian and one Latin-American faces per trial. Their main findings suggest that Asian, African-American, and Caucasian participants looked at Caucasian faces longer overall than they looked at Latin-American faces, whereas Latin-American participants did not exhibit preferences towards neither targets. Interestingly, no significant differences were found in the actual attractiveness ratings between the target faces before surgery, however, post-surgery images of Caucasian targets were rated to be more attractive. This study demonstrated that when making attractiveness judgments, target face ethnicity and participant ethnicity could influence how individuals look at a face, however, we are not able to make conclusions as to which particular areas of the face was attended to when making attractiveness judgements.

Health Perception

One study that looked at OEE on perception of health is Rhodes et al. (2005), where they showed participants from Caucasian and Asian background, and mixed-race composites. Interestingly, they found that Eurasian faces received the highest health ratings. This could be because the more diverse the genetic make-up of an individual, the healthier they tend to be (Lie et al., 2009). The current study therefore aims to expand on this by exploring whether observers look at faces from different ethnic background when judging for facial health.

Which Areas do People Look at When Making Socially Salient Judgments?

Previous studies using behavioural tasks have demonstrated that certain areas of the face were used by observers when making evaluations. Burt and Perrett (1997) averaged faces on different traits, e.g. young/old, unattractive/attractive, female/male, and whether they were saying ‘ee’ or ‘ss’, and made a composite target image, e.g., one side of the face is an average of young faces and the other side is an average of old faces. These composites were presented to the participants and were asked to judge whether the face is young or old, attractive or unattractive, and so forth. They found that participants used the left side (target's right side) when making age, gender, and attractiveness judgements, i.e., when the younger or more attractive face blends was presented on the left side, the face was perceived to be younger or more attractive, and the right side (target's left side) for speech, demonstrating that different facial areas are being used when making social judgments. They argued that this could be due to the influence of the right hemisphere in processing of facial stimuli. However, this study did not have observers from different ethnicity, and therefore does not inform us whether this also applies when people judge faces from their own- and other-ethnicities.

As described earlier, Arizpe and colleagues (2016) found support that Caucasian participants viewed Asian, Black, and Caucasian faces differently during recognition tasks. This study demonstrated that the ethnicity of the target face has an effect on viewing patterns during a recognition task, however, this study only used one group of observers. We are therefore unable to draw conclusions whether this pattern was only applicable to Caucasians, or whether this difference is universal, that is, individuals have a certain pattern when looking at faces from their in-group compared to when they look at faces from the out-group. To expand on this work, the current study recruited participants from two ethnic groups - Caucasians and Asians. This could demonstrate whether different ethnic groups also employ different scanpaths when looking at other-ethnicity faces.

In relation to socially salient traits, previous studies using ET have also shown that there are biases in how we perceive faces when making judgements on age and attractiveness. Nguyen et al. (2009) investigated how observers look at faces of varying ages when making judgments of age and fatigue. They found that although the eye regions received more fixations and saccades, i.e., rapid movement of the eyes in between fixations overall in both tasks, observers looked at the glabella (area between the eyebrows) and the brows more for faces rated as older and looked at the cheeks more for faces rated as higher in fatigue. This study was the first to show that different areas of the face were utilised by observers when judging different traits.

Additionally, Kwart et al. (2012) followed Nguyen et al's procedure (2009) and investigated how observers look at faces when judging for attractiveness and age. Similar to Nguyen's findings (2009), they found that the eye regions received the most attention when observers make judgments on how attractive and how old the image is, followed by fixations to the nose and mouth. These two studies show that there are specific facial areas which are paid more attention when making judgements of attractiveness and age.

Fink et al. (2008) also used ET method to investigate how much observers would look at younger or older female faces. It was argued that skin colour distribution is a significant cue for age, where female faces with even skin tone were rated younger than their actual age, and those with uneven skin tones were rated older (Fink, Grammer, et al., 2006). Participants were therefore presented with five unmodified images with differing distribution. Their findings showed that faces with more even skin tone were rated more attractive and younger, as well as receiving higher visual attention, compared to faces who were rated less attractive and older. Such findings demonstrate that certain facial attributes receive more visual attention than others, and that facial skin colour contributed to people's judgments of attractiveness and age estimation. However, although this study did not investigate the specific areas used by the observers to make these judgments. Additionally, as Asian and African populations report more concerns with uneven skin tone as signs of aging, it will be interesting to see how observers from these groups view facial images.

So far, we have seen that different facial areas and attributes affect facial recognition, attractiveness and age. As far as we know, there are no studies yet which investigated which facial areas are looked at when participants evaluate facial health. With the increasing demand in the cosmetic industry to enhance facial features, it is important to understand whether specific areas of the face contribute to the perception of attractiveness, health and age. This is where eye-tracking techniques provide answers. Eye movements comprise of saccades (rapid, small movements) and fixations (longer gaze) (Arizpe et al., 2016; Kanan et al., 2015). These movements create a pattern, called scanpaths, which inform us of how observers look at a face when making certain judgements. It has been demonstrated that different tasks have idiosyncratic scanpaths (Kanan et al., 2015), and the current study aims to explore whether such difference could also be influenced by participant ethnicity and target face ethnicity.

Aims and hypotheses

The overall aim of the current study, therefore, is to replicate and extend the work of Arizpe et al. (2016) to determine whether the pattern of how we look at faces differ as a function of target face ethnicity and participant ethnicity, and whether this also applies to other task, e.g. ratings of attractiveness, healthiness, and age. The hypotheses for this study were as follows:

1. There will be a main effect of facial areas of interest (AOI) on fixation and duration for evaluations of facial age, attractiveness, and health, where the eye regions would be looked at more than other areas (Kwart et al., 2012; Nguyen et al., 2009);
2. There will be a main effect of Target Face Ethnicity on fixation and duration for evaluations of facial age, attractiveness and health (Stelter et al., 2021);
3. There will be a main effect of Participant Ethnicity on fixation and duration for evaluations of facial age, attractiveness and health;
4. There will be an interaction between AOI and Participant Ethnicity, where Asian participants would look at Noses more and Caucasians looking at eye regions more (Blais et al., 2008);
5. There will be an interaction between AOI and Target Face Ethnicity, though we have no specific predictions about the pattern that may emerge

6. There will be an interaction between the Target Face Ethnicity and Participant Ethnicity on evaluations of facial age, attractiveness, and health, where we expect that participants will look more on own-ethnicity faces than other-ethnicity faces.

7. There will be a three-way interaction between AOI, Target Face Ethnicity, and Participant Ethnicity on fixation and duration for evaluations of facial age, attractiveness, and health, where we expect that there will be higher Nose fixations (and by extension shorter duration) when participants look at own-ethnicity faces, and there will be more mouth and nose fixations when participants look at other-ethnicity faces (Arizpe et al., 2016).

Note that we do not have predictions for interactions between AOI and Target Face Ethnicity as this has not been explored previously. We hope that the current study will provide information regarding these.

Methods

Participants

45 Swansea University students completed the study. Due to the nature of the stimuli, we primarily recruited for participants from Caucasian and Asian ethnic background. All participants must be over 18, with no upper limit for age. One participant reported being mixed Asian and Caucasian and had to be excluded on the final analysis, and one participant's eye tracking data was not saved, therefore leaving 43 participants for the final analysis (26 Caucasians, 18 Asians; 28 M, 15 F) with age range of 18-25 (M = 20.186, SD = 1.402). Undergraduate students from the Psychology department were recruited using the Psychology SONA system (https://psychology-swansea.sona-systems.com/) and received four research credits as part of their course requirement, where other students were recruited through convenience sampling and did not receive compensation. Participants had normal or corrected vision.

Ethics Statement

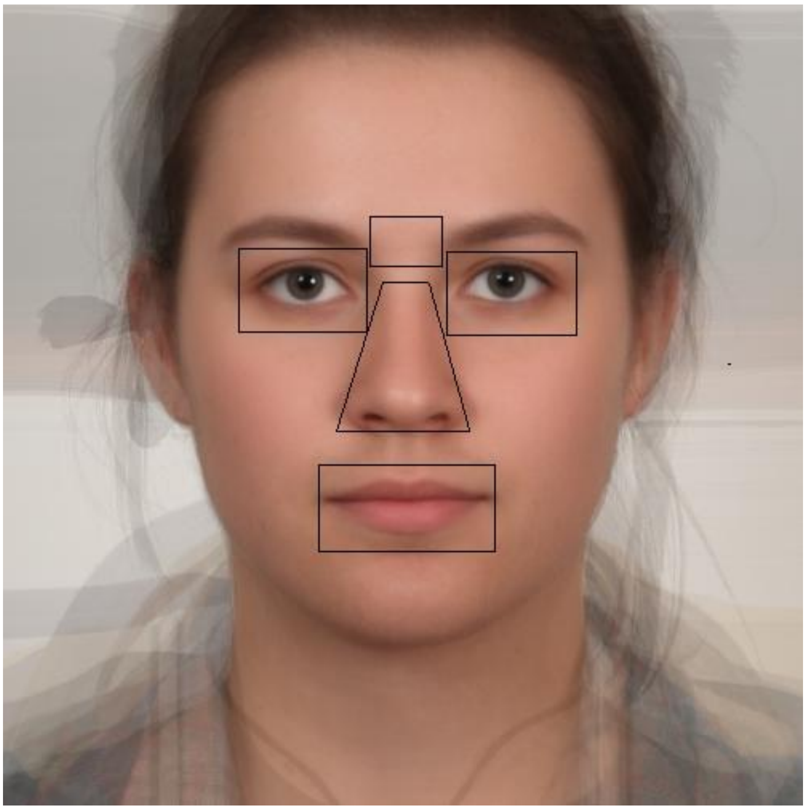
Participants were invited to participate in the study either using SONA or word of mouth. Written participant informed consent was obtained prior to starting the experiment. Participants were shown the eye-tracking apparatus and was informed on how data was being collected in the study. Participants were informed that they are allowed to pause during the experiment if needed, and that they are allowed to withdraw from the study at any point during the experiment. Only fully completed tasks were included in the analysis. This study was approved by the Swansea University Ethics Committee and followed the Declaration of Helsinki (Williams, 2015).

Apparatus

The eye movement behaviour was recorded using the SR Research EyeLink 1000 eye tracker. Using its default algorithm, the EyeLink recorded saccades and fixations made in each trial using the distance between pupil image and corneal reflection at 1000 Hz. Participants sat across a computer monitor, approximately 60 inches away from the 381mm x 306mm screen. The screen has a visual angle of 30° and 21.35°, and image size were 1080x760 pixels. Participants were asked to place their heads on a chinrest to reduce movement. Areas of interest were defined using the area selection tool, using regions based on Arizpe et al.'s study which used left and right eye regions, glabella, nose, mouth and other (outside the pre-defined regions). Each AOI was manually adjusted for each face to fit the facial feature as needed. Number of fixations and dwell times were recorded for each area of interest (Figure 4.4), as well as for each individual trials. We included the full trial duration and all fixations for each trial in our analysis.

Figure 4.4

Demonstration of Areas of Interest Used in the Study.



Note: Each AOI was adjusted to fit the features for each face. 'Other' areas were regions outside the pre-defined AOIs. This image is a composite from Scott et al., 2013 and was only used for demonstrative purposes. AOIs were not visible to participants during the experiment.

Materials

The facial images used in this study were randomly selected from the Asian, African, and Caucasian samples of the Chicago Face Database (CFD, Ma et al., 2015). 20 images were selected from each ethnic group which were compiled to create one block of the experiment (see Github repository for the images). There were three blocks overall, one for each rating - age estimation, health and attractiveness perceptions. The order of the tasks was counterbalanced among the participants to minimize repetition effects.

Procedure

Participants attended the eye tracking laboratory in person. Upon arrival, they received the Participant Information Sheet, as well as a verbal explanation of the study rationale. Following provision of consent to participate, demographic information and questionnaires were completed. Prior to the experiment, participants were asked to sit in front of the eye-tracker and adjustments were made to ensure correct positioning. The experiment consisted of three blocks - age estimation, attractiveness, and health tasks. The order of facial images in each block appeared randomly, and the block order was counterbalanced between participants.

Age estimation

Prior to each trial, a drift correct procedure was made to ensure that participants are fixating at the centre of the screen. A facial image then appeared for six seconds, followed by a sliding scale. Participants were instructed to use the mouse to estimate the age of the face, ranging from 20 to 70 years old. Responses were self-paced, and participants clicked on the ‘Continue’ button to record their answers and proceed with the experiment.

Attractiveness and health perceptions

These blocks followed the layout of the age estimation task, with the response choices for the sliding scale changed to ‘Unattractive - Attractive’ and ‘Unhealthy - Healthy’ respectively. These were recorded as 0-100.

Design and analytic strategy

Duration was defined as the duration of all indices for each trial. We fitted a linear mixed-effects model in R (R Core Team, 2021) using lme4 (Bates et al., 2014) on duration, with fixed effects of areas of interest, participant ethnicity, target face ethnicity and their interactions. Participants and target faces were used as random effects, to reflect the ratings came from different individuals and they were rating different images. Using LME allowed us to investigate the differences in dwell times when participants were evaluating different types of faces. The model for each task was as follows:

$$\mathbf{Duration = AOI \times P_{ethnicity} \times T_{ethnicity} + (1|P) + (1|\ T)}$$

Note: P = Participant, T = Target stimuli.

We also ran a Bayes Factor analysis on each model where appropriate to determine whether the effects found were more likely to be under the null or alternative hypothesis using the 'BayesFactor' package (version 0.9.2, Morey, 2024) in R. Furthermore, we also ran a 'recompute' function where appropriate – this involves re-running the model 1,000,000 times to increase the precision of the estimate.

Following Arizpe et al. (2016) relative fixation counts (RFC) were calculated by the proportion of possible fixations for that trial, e.g. 20 trials for Asian target faces multiplied by all fixations per trial. When added, all RFCs for each trial sum up to one. Due to the nature of RFC calculations, we were not able to use linear mixed models on this variable therefore a Repeated Measures ANOVA was employed and where appropriate, a complementary Bayes Factor Analyses were ran to determine whether the effects were likely to be under the null using JASP (JASP Team, 2024). Greenhouse-Geisser corrections were applied when sphericity assumption was violated. Our independent variables for this study were the *Areas of Interest* (six levels: Nose, Left eye, Right eye, Glabella, Lips, and Other), *Target Face Ethnicity* (three levels: Asian, Black, Caucasian), *Participant Ethnicity* (two levels: Asian and Caucasian). The dependent variables for this study were categorised as evaluations (age estimations, and ratings for attractiveness and health perceptions) and eye movements for each trait (duration and relative fixation counts). Although our main analysis will only focus on eye movements.

Results

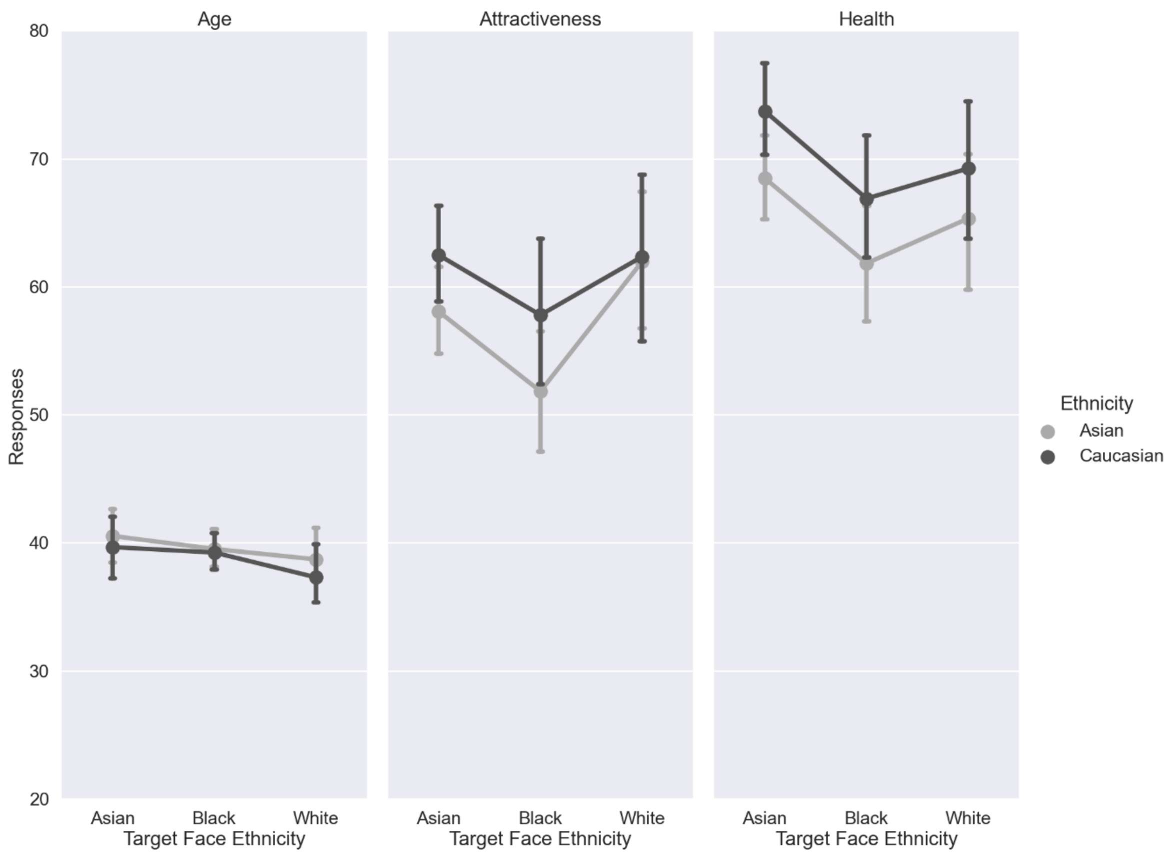
Descriptive Statistics

Evaluations on social traits

Figure 4.5 shows the evaluations given by participants to the target faces. In sum, for age evaluations, Asian (M = 39.58, SD = 4.50) and Caucasian (M = 38.73, SD = 4.99) participants gave similar age estimates to the target faces. For attractiveness and health evaluations, however, Caucasian participants gave higher ratings (M = 60.87, SD = 12.83; M = 69.95, SD = 11.34, respectively), compared to Asian participants (M = 57.29, SD = 11.29; M = 65.21, SD = 10.58).

Figure 4.5

Evaluations Given by Asian and Caucasian Participants to Target Faces.



Note: For age estimations, participants rated the faces between 20 and 70 years old. For attractiveness and health evaluations, participants rated the faces between 0-100.

Durations

Table 4.8 shows the mean and standard deviations of duration for age, attractiveness, and health tasks grouped by trait, participant ethnicity, target ethnicity, and areas of interest.

Overall, Caucasians looked at the faces longer than Asian participants did for age (M = 342.25, SD = 338.05, M = 314.35, SD = 154.28, respectively), attractiveness (M = 334.34, SD = 248.40, M = 311.38, SD = 176.53, respectively), and health (M = 328.44, SD = 218.94, M = 305.81, SD = 144.77, respectively) than Asian participants.

For target face ethnicity, Black faces received the longest duration for age (M = 333.52, SD = 245.21), followed by Caucasian faces (M = 330.40, SD = 248.10) and Asian faces (M = 329.02, SD = 333.11). Similar trends were observed for attractiveness (M = 328.57, SD = 261.44, M = 325.23, SD = 201.09, M = 320.99, SD = 197.60, respectively), and health tasks (M = 323.39, SD = 230.00, M = 320.46, SD = 186.81, M = 313.74, SD = 151.47).

Finally, for areas of interest, the Nose received the longest duration, while the Glabella received the shortest durations across all three tasks.

Table 4.8

Means and Standard Deviations for Dwell Times Across Tasks, Participant Ethnicity, Target Ethnicity, and Areas of Interest.

Decision	Participant Ethnicity	Target Ethnicity	Areas of Interest											
			Glabella		Left eye		Lips		Nose		Other		Right eye	
			M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Age	Asian	Asian	221.64	132.65	329.88	133.31	267.69	166.73	349.06	169.98	329.89	152.09	321.75	161.21
		Black	216.29	133.05	337.58	160.37	265.52	127.35	334.06	122.77	325.93	145.45	307.51	127.84
		White	206.85	155.26	339.89	162.31	268.82	152.10	341.40	185.17	329.91	149.07	311.09	136.35
	White	Asian	245.87	169.66	318.88	158.08	257.34	133.18	388.04	195.01	330.50	197.82	394.58	836.88
		Black	262.70	323.89	356.34	256.87	251.56	134.82	391.57	229.77	345.97	331.57	386.07	374.30
		White	224.32	163.12	346.50	323.95	224.14	103.62	386.31	385.05	338.73	199.64	362.05	285.19
Attractiveness	Asian	Asian	250.65	257.94	329.62	148.03	279.73	141.15	370.26	163.74	289.56	137.94	302.93	122.33
		Black	223.03	174.77	325.39	167.99	301.73	276.82	362.06	157.60	279.55	131.88	299.70	125.20
		White	216.53	191.40	331.52	180.48	260.24	159.03	361.62	224.90	305.43	205.46	314.11	128.91
	White	Asian	247.67	162.27	322.72	176.37	261.75	238.88	417.01	199.42	310.54	194.87	359.08	289.98
		Black	269.20	164.84	338.95	246.89	288.50	300.92	426.14	461.37	305.54	189.07	354.46	261.36
		White	279.38	188.73	308.85	154.48	268.37	319.48	404.59	198.65	285.62	128.06	333.91	213.85

Health	Asian	Asian	230.97	177.34	310.02	159.36	274.43	129.90	372.44	152.53	286.83	110.77	311.13	123.26
		Black	242.28	165.61	324.48	141.36	268.64	128.43	362.44	157.92	285.43	116.74	309.85	130.88
		White	223.88	153.30	305.34	134.22	260.42	120.13	362.62	155.60	286.19	140.75	313.02	134.49
	White	Asian	261.89	156.29	345.87	297.77	266.81	145.24	419.13	224.47	295.40	139.62	329.67	166.38
		Black	301.20	266.83	326.52	215.20	284.80	149.15	399.14	187.86	303.18	302.30	357.00	401.23
		White	278.19	161.90	326.75	161.98	259.76	124.39	384.70	171.13	288.78	116.99	328.13	154.49

Note: Maximum viewing time for each face was six seconds. Durations were measured in ms.

Relative Fixation Counts

Table 4.9 shows the mean and standard deviations for relative fixation counts (RFCs) across traits, participant ethnicity, target ethnicity, and areas of interest. In general, the Nose received the highest RFC (M = .22, SD = .02), and the Glabella received the lowest RFC (M = .09, SD = .04). Due to the nature of how RFC was calculated, i.e. a proportion of fixations each facial area received, when aggregated, target faces (Asian, Black and Caucasian) received similar average fixations, that is, 1.

Table 4.9

Means and Standard Deviations for Relative Fixation Counts Across Tasks, Participant Ethnicity, Target Ethnicity, and Areas of Interest.

Decision	Participant Ethnicity	Target Ethnicity	Area of Interest											
			Glabella		L Eye		Lips		Nose		Other		R Eye	
			M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Age	Asian	Asian	0.10	0.04	0.20	0.02	0.10	0.04	0.21	0.02	0.21	0.02	0.19	0.04
		Black	0.07	0.03	0.20	0.01	0.14	0.03	0.21	0.01	0.20	0.01	0.19	0.03
		White	0.06	0.04	0.21	0.02	0.10	0.04	0.22	0.01	0.22	0.02	0.20	0.03
	White	Asian	0.09	0.07	0.19	0.05	0.10	0.05	0.22	0.03	0.21	0.02	0.20	0.01
		Black	0.08	0.07	0.19	0.05	0.11	0.05	0.22	0.03	0.20	0.02	0.21	0.02
		White	0.05	0.05	0.20	0.05	0.09	0.05	0.23	0.04	0.21	0.02	0.21	0.02
Attractiveness	Asian	Asian	0.10	0.03	0.20	0.02	0.10	0.03	0.21	0.01	0.20	0.01	0.18	0.04
		Black	0.08	0.04	0.20	0.02	0.13	0.04	0.21	0.01	0.20	0.01	0.18	0.05
		White	0.05	0.04	0.21	0.02	0.11	0.04	0.22	0.01	0.21	0.01	0.20	0.04
	White	Asian	0.10	0.07	0.19	0.05	0.09	0.05	0.22	0.03	0.21	0.03	0.20	0.03
		Black	0.08	0.08	0.19	0.05	0.11	0.06	0.22	0.03	0.20	0.04	0.20	0.03
		White	0.06	0.07	0.20	0.04	0.10	0.05	0.24	0.06	0.21	0.04	0.20	0.03
Health	Asian	Asian	0.09	0.03	0.19	0.02	0.12	0.04	0.21	0.01	0.20	0.01	0.20	0.02
		Black	0.09	0.04	0.19	0.02	0.14	0.03	0.20	0.01	0.20	0.01	0.19	0.02
		White	0.06	0.04	0.20	0.02	0.13	0.04	0.21	0.01	0.21	0.01	0.20	0.02
	White	Asian	0.11	0.07	0.18	0.04	0.11	0.05	0.21	0.03	0.21	0.03	0.18	0.04
		Black	0.10	0.07	0.19	0.03	0.12	0.06	0.21	0.04	0.19	0.04	0.18	0.03
		White	0.08	0.06	0.19	0.03	0.11	0.05	0.22	0.05	0.21	0.04	0.19	0.03

Note: RFC was calculated as a proportion of all fixations for a face and sum up to 1.

Inferential statistics

Age

Durations. We found that several AOIs were significant predictors for age estimation: left eye (b = 106.58, t(11877.34) = 4.22, p < 001), nose (b = 128.08, t(11878.20), p < .001, other areas (b = 108.92, t(11878.17), p < .001), and right eye (b = 95.44, t(11873.81), p < .001). We also found a significant interaction between the right eye and White participants (b = 73.59, t(11872.48), p = .03).

To further investigate the significance of coefficients in our model, we conducted an ANOVA (using Type III sums of squares) on the fitted linear mixed model in R (R Core Team, 2012). We found a small, significant effect of Areas of Interest, F(5, 11878.8) = 47.89, $\eta p^2 < .02$, $p < .001$ (Figure 4.6). Post-hoc pairwise comparisons showed that the Nose was significantly looked at longer than the Glabella, Lips and Other areas ($p < .001$), but was looked at with similar duration as the Left and Right eye areas ($p > .05$). The Left and Right eye areas were significantly looked at longer than the Glabella and Lips area (both $p < .001$) and were looked at similarly as the Nose and Other, and each other ($p > .05$).

We did not find significant main effects of Participant Ethnicity, F(1, 41.1) = 1.15, $\eta p^2 = .03$, $p = .23$, the complementary Bayes factor analysis showed that this result was not likely to be under the null hypothesis, given the data (BF_{01} = .012, recomputed); nor of Face Ethnicity, F(2, 81.8) = .81.8, $\eta p^2 < .01$, $p = .65$, the complementary Bayes factor analysis showed that this was more likely under the null hypothesis (BF_{01} = 436.405).

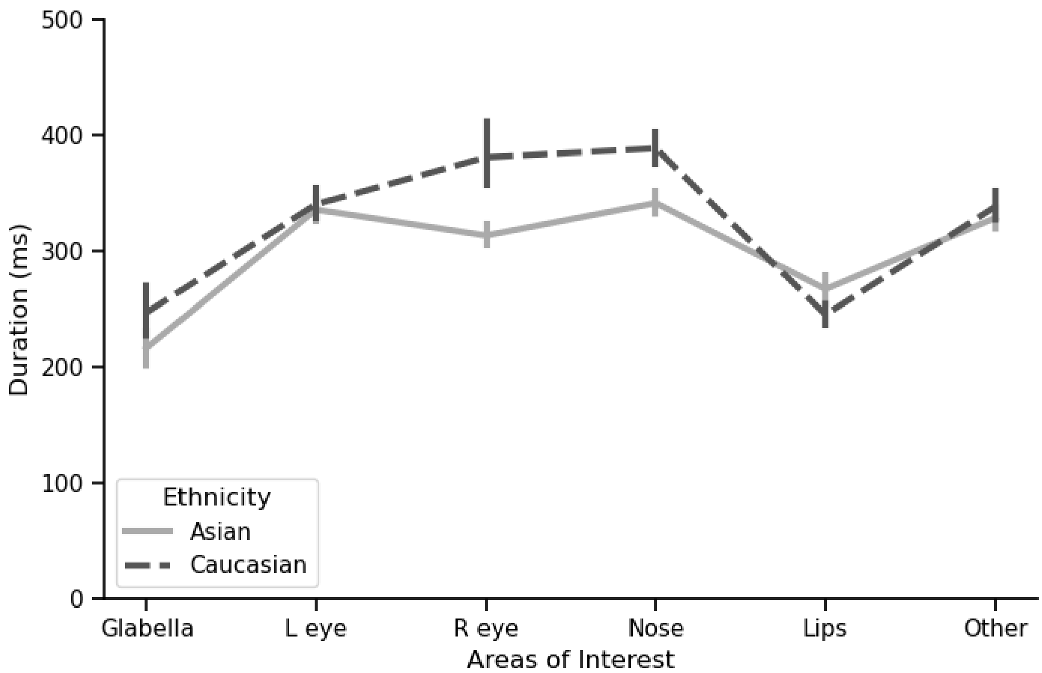
We did find a small, but significant interaction between AOI and Participant Ethnicity, F(5, 11870.0) = 4.77, $\eta p^2 < .01$ $p < .001$ (Figure 4.6). Post-hoc analysis showed that Asian participants looked at the Glabella in shorter durations than the other AOIs except the Lips (all $p < .001$), and that the Lips were also looked at in shorter duration than the Nose area (p <

.001). For White participants, the Glabella and the Lips areas received shorter durations than all the other AOIs. No other significant comparisons were found.

We found no significant interaction between Participant Ethnicity and Face Ethnicity, F(2, 11870.3) = 1.20, $\eta p^2 < .01$, p = .30; between AOI and Face Ethnicity, F(10, 111872.3) = .63, $\eta p^2 < .01$, p =.78; nor a significant interaction between all three variables, F(10, 11863.7) = .16, $\eta p^2 < .01$, p =.99. The complementary Bayes factor analysis showed that these findings were more likely under the null hypothesis (BF_{01} = 117.639, BF_{01} = 4565483, and BF_{01} =1886498, respectively).

Figure 4.6

Interactions of AOI and Participant Ethnicity for Durations on Age Task.



Relative Fixation Counts. We found a large, significant main effect of AOI, F(1.908, 78.237) = 112.780 $p < .001$, $\eta p^2 = .733$. Post-hoc comparisons with Bonferroni corrections showed that the Nose received significantly more fixations compared to other areas (all $p < .001$), whereas the Left and Right eye, as well as Glabella and Lips received similar fixation counts ($p = 1.00$ for both).

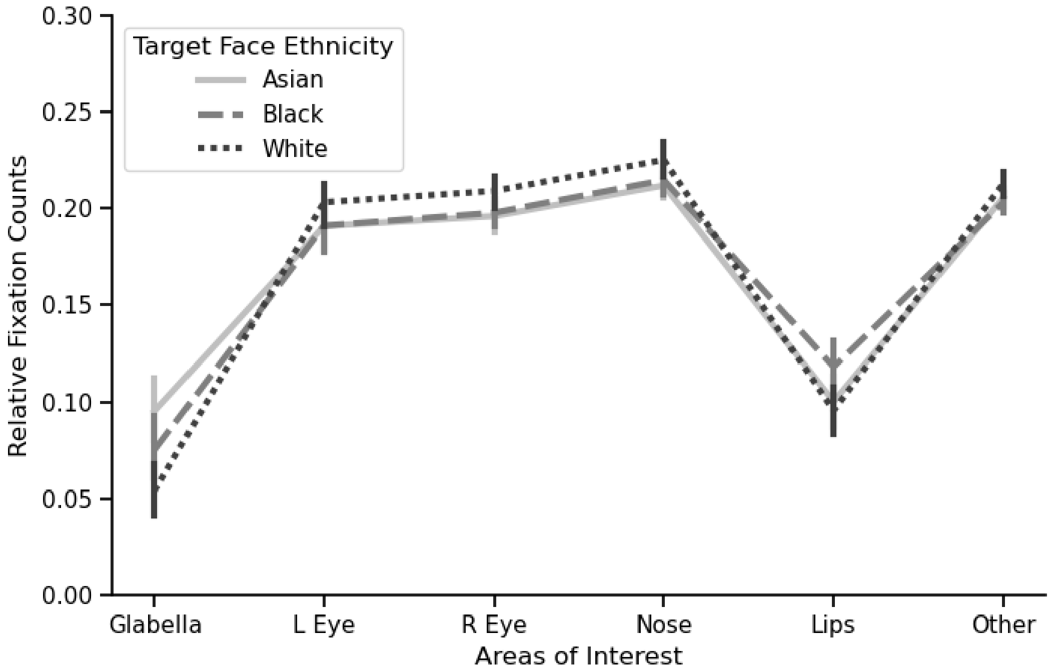
We did not find a significant main effect of Face Ethnicity, F(2, 82) = .274, $p = .761$, $\eta p^2 = .007$, nor of Participant Ethnicity F(1,41) = .558, $p = .459$, $\eta p^2 = .013$. The complementary Bayes factor analysis showed that this is more likely under the null hypothesis ($BF_{01} = 8.564x10^{+94}$, $BF_{01} = 1.444x10^{+95}$, respectively).

We found a significant interaction between AOI and Face Ethnicity (Figure 4.7), F(4.763, 195.283) = 19.740, $p < .001$, $\eta p^2 = .325$. Within AOIs, Glabella regions were looked at significantly different from each other (all $p < .001$), and the Lips area of White and Asian faces received similar RFCs ($p = 1.000$), while the Lips area of Black faces were looked at more than Asian and White faces (both $p < .001$). No significant difference for L eye, Nose, Other, and R eye regions were found. Within each face ethnicity, the Nose, Other, and both eye regions received similar RFCs ($p = 1.000$) and were looked at more than Glabella and Lip regions ($p < .001$).

We did not find a significant interaction between AOI and Participant Ethnicity, F(1.908, 78.237)= 1.321, $p = .256$, $\eta p^2 = .031$; and Participant Ethnicity and Face Ethnicity, F(2, 82) = .558, $p = .575$, $\eta p^2 = .013$. Finally, we also did not find an interaction between the three variables, F(4.763, 195.283) = 1.695, $p = .141$, $\eta p^2 = .040$. The complementary Bayes factor analysis showed that these findings are more likely under the null hypothesis ($BF_{01} = 1.804x10^{+40}$, $BF_{01} = 1.790x10^{+97}$, and $BF_{01} = 11.490$, respectively) .

Figure 4.7

Interactions of AOI and Target Face Ethnicity for Relative Fixation Counts on Age Task.



Attractiveness

Durations. We found that several AOIs were significant predictors for attractiveness perception: left eye (*b* = 78.44, t(11752.47) = 4.16, $p < 001$), nose (b = 120.43, t(11751.87) = 6.43, $p < .001$, other areas (*b* = 40.46, t(11754.09) = 2.16, $p = .03$), and right eye (b = 56.42, t(11752.70) = 2.95, $p = .003$). We also found a significant two-way interaction between the nose and White participants (b = 57.20, t(11754.50) = 2.32, $p = .02$) and right eye and White participants (b = 71.63, t(11755.11) = 2.86, $p = .004$). Finally, we found small, but significant three-way interactions between Other area, White participants and White target faces (b = -94.9, t(11745.35) = -2.4, $p = .02$) and Right eye, White participants and White target faces (b = -88.3, t(11746.38) = -2.19, $p = .03$).

To further investigate our model for attractiveness, we ran an ANOVA (using Type III sums of squares) on the fitted linear mixed model in R (R Core Team, 2012).

We found a small, but significant main effect of AOI (Figure 4.6), (F(5, 1747.4) = 95.12, $\eta p^2 = .04$, $p < .001$). Post-hoc pairwise comparisons showed that the Left and Right eye areas received similar durations ($p = 1.0$), as well as Lips and Other areas ($p = .48$). All other comparisons were significant, where in general, the Nose, Left and Right areas received longer durations than the other areas of interests.

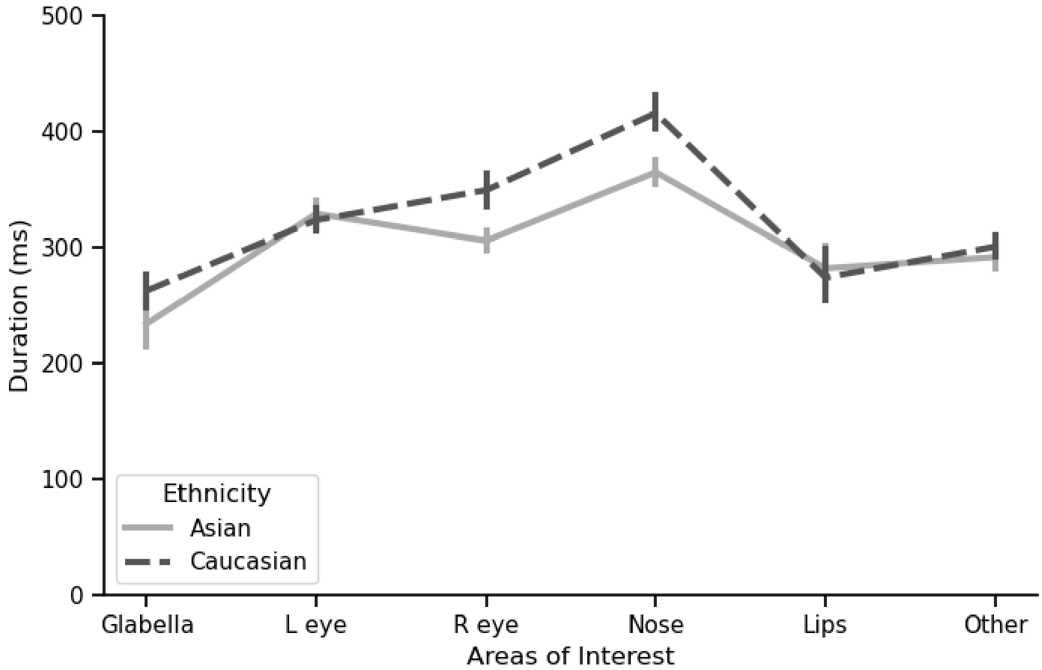
We did not find a significant main effect of Participant Ethnicity, F(1, 41.0) = .52, $\eta p^2 < .02$, $p = .34$. The complementary Bayes factor analysis showed that this was not likely under the null hypothesis (BF_{01} = .001, recomputed). We also did not find a significant main effect of Face Ethnicity, F(2, 76.1) = .52, $\eta p^2 < .01$, $p = .60$. The complementary Bayes factor analysis showed that this was more likely under the null hypothesis (BF_{01} = 221.027).

We found a small, but significant interaction between AOI and Participant Ethnicity, $F(5, 11739.1) = 3.69$, , $\eta p^2 < .01$, $p = .002$ (Figure 4.8). Post-hoc analysis has shown that for Asian participants, the Nose received the most durations, Right eye received similar durations to Left eye ($p = .31$) and Other ($p = .46$), and the Lips area received similar durations to Other and Right eye (both $p > .06$). The Glabella received the least durations (all $p < .005$). For Caucasian participants, the Nose also received the most durations, followed by the Left and Right eye ($p = .09$), and Lips and Other ($p = .81$). The Glabella also received the shortest durations (all $p < .001$).

We did not find any two-way interactions between AOI and Face Ethnicity, $F(10, 11742.3) = .57$, $\eta p^2 < .00$, $p = .84$; nor between Participant Ethnicity and Face Ethnicity, $F(2, 11740.8) = 1.33$, $\eta p^2 < .00$, $p = .27$. Finally, we also did not find significant interactions between the three variables, $F(10, 11734.0) = 1.15$, $\eta p^2 < .00$, $p = .32$. The complementary Bayes factor analysis showed that these findings were likely under the null hypothesis ($BF_{01} = 8794201$, $BF_{01} = 79.127$, and $BF_{01} = 37731.67$, respectively).

Figure 4.8

Interactions of AOI and Participant Ethnicity for Durations on Attractiveness Task.



***Relative Fixation Counts*.** We found a large, significant main effect of AOI on frequency counts, F(5, 205) = 74.086, $p < .001$, $\eta p^2 = .644$. Following post-hoc comparisons using Bonferroni corrections, we found that the Nose, Other, and both eye regions received similar RFCs ($p > .05$) and were higher than Glabella and Lips, which received similar RFCs ($p > .05$).

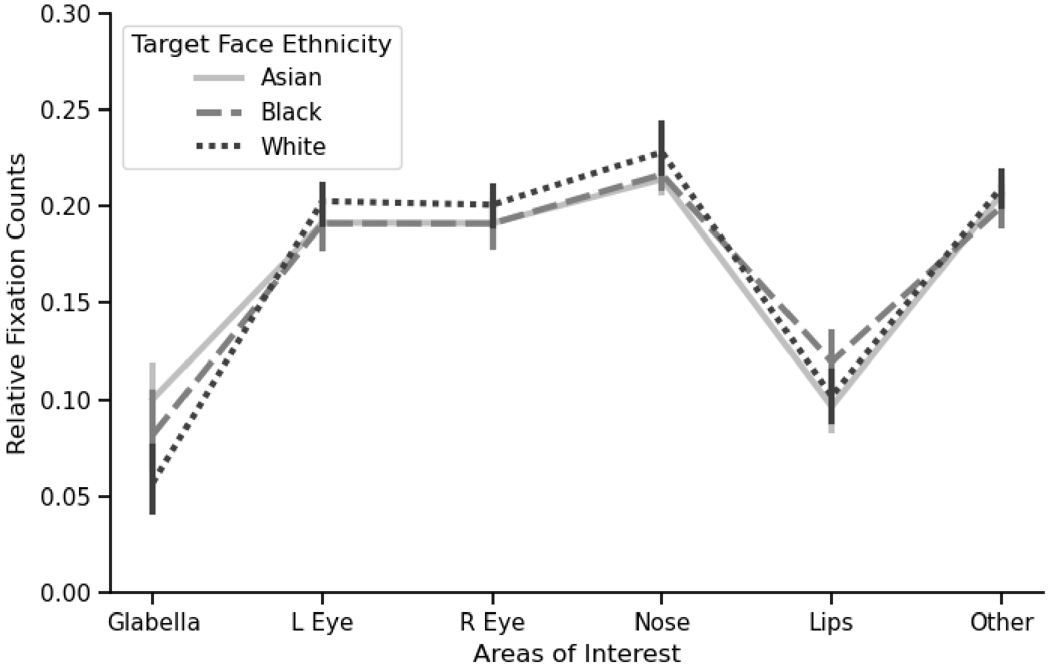
We did not find a significant main effect of Face Ethnicity, F(2, 82) = .015, $p < .985$, $\eta p^2 = .000$, nor of Participant Ethnicity, F(1, 41) = .010, $p = .920$, $\eta p^2 = .000$. The complementary Bayes factor analysis showed that these findings were more likely to be under the null hypothesis (BF_{01} = 4.624x10+72 and $BF_{01} = 8.033\text{x}10^{+72}$, respectively).

We found a significant interaction between AOI and Target Face Ethnicity, F(10, 410) = 15.701, $p < .001$, $\eta p^2 = .277$ (Figure 4.9). Post-hoc comparisons using Bonferonni corrections found that within AOIs, the Glabella of Asian faces were looked at *more* than those of Black and White faces (both $p < .001$), Black Lips were looked at *more* than White and Asian Lips ($p < .01$). All other comparisons were not significant ($p > .05$). Within Face Ethnicity, Asian L eye, Nose, Other, and R eye received similar RFCs ($p = 1.000$), and were looked at more than the Glabella and Lips ($p < .05$), which received similar RFCs ($p = 1.000$). For Black and White targets, L eye, Nose, Other, and R eye regions also received similar RFCs ($p = 1.000$) and were looked at more than Glabella and Lips (both $p < .05$), however, the Lips region received higher RFCs than Glabella in both target faces ($p < .001$).

We did not find a significant interaction between AOI and Participant Ethnicity, F(5, 205) = .750, $p = .587$, $\eta p^2 = .018$; and between Participant Ethnicity and Target Face Ethnicity, F(2, 82) =.517, $p = .598$, $\eta p^2 = .012$. Finally, we did not find a significant interaction between all three variables, F(10, 410) = .458, $p = .916$, $\eta p^2 = .011$. The complementary Bayes factor analysis showed that these findings were more likely to be under the null hypothesis ($BF_{01} = 3.362\text{x}10^{+29}$, $BF_{01} = 1.054\text{x}10^{+75}$, and $BF_{01} = 62733.796$, respectively).

Figure 4.9

Interactions of AOI and Target Face Ethnicity for Relative Fixation Counts on Attractiveness Task.



Health

Durations. We found significant predictors for health: Left eye (b = 74.66, t(12183.55) = 4.45, $p < .001$), Nose (b = 135.76, t(12183.70) = 8.20, $p < .001$, Other (b = 50.23, t(12182.80) = 3.03, $p = .002$), and Right eye (b = 76.154, t(12186.23) = 4.56, $p < .001$). We also found three-way interactions between Left eye, White participants, and Black target face (b = -66.93, t(12172.67) = -2.20, $p = .03$).

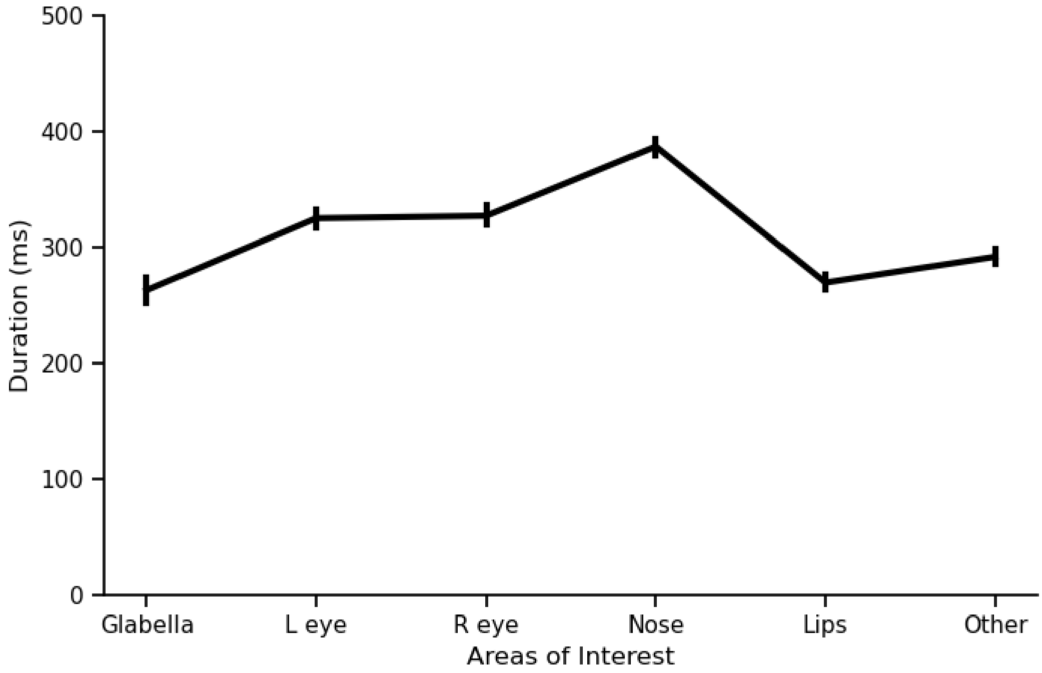
For further investigation, we also ran an ANOVA (using Type III sums of squares) on the fitted linear mixed model in R (R Core Team, 2012). Here, we found a small, but significant main effect of AOI (Figure 4.10), F(5, 12179.5) = 125.60, $p < .001$, $\eta p^2 = .05$; where a post-hoc pairwise comparison showed that the Nose, Left and Right eye areas were looked at significantly longer than all the other facial regions ($p < .001$). The Left and Right eye areas received similar durations ($p = .99$).

We did not find a significant main effect of Participant Ethnicity, F(1, 40.9) = 1.08, $\eta p^2 = .03$, $p = .31$. The complementary Bayes factor analysis showed that this was not likely under the null hypothesis ($BF_{01} = 4.382x10^{-08}$, recomputed). We also did not find a significant main effect of Target Face Ethnicity, F(2, 68.0) = 2.93, $\eta p^2 = .08$, $p = .06$. The complementary Bayes factor analysis showed that this was more likely under the null hypothesis ($BF_{01} = 24.839$).

We did not find significant two-way interactions between AOI and Participant Ethnicity, F(5, 12168.4) = 78, $p = .56$, $\eta p^2 < .01$; AOI and Face Ethnicity, F(10, 12176.3) = .88, $p = .55$, $\eta p^2 < .01$; and Participant Ethnicity and Target Face Ethnicity, F(2, 12167.0) = 1.31, $\eta p^2 < .01$, $p = .27$. We also did not find a significant three-way interaction between the three variables, F(10, 12165.2) = .1.07, $\eta p^2 < .01$, $p = .38$. The complementary Bayes factor analysis showed that these findings were more likely under the null hypothesis ($BF_{01} = 127.695$, $BF_{01} =1627780$, $BF_{01} = 216.499$ and $BF_{01} = 80864$, respectively).

Figure 4.10

Main Effects of AOI for Health Task.



Relative Fixation Counts. We found a large, significant main effect of AOI, F(5, 205) = 62.6777) = 112.623, $p < .001$, $\eta p^2 = .605$. Post-hoc comparisons with Bonferroni corrections showed similar trends to the Age and Attractiveness tasks, where the Nose, Other, and both eye regions received similar RFCs ($p > .05$), and were looked at more than the Lips, and Glabella the least.

We did not find a significant main effect of Target Face Ethnicity, F(2, 82) < .001, p = 1.000, np2 < .001, nor of Participant Ethnicity, F(1, 41) < .001, p = 1.000, $\eta p^2 < .0071$. The complementary Bayes factor analysis showed that these are likely to be under the null hypothesis ($BF_{01} = 5.331x10^{+58}$, and $BF_{01} = 8.796x10^{+57}$, respectively), suggesting that both Asian and Caucasian participants looked at the three target groups similarly.

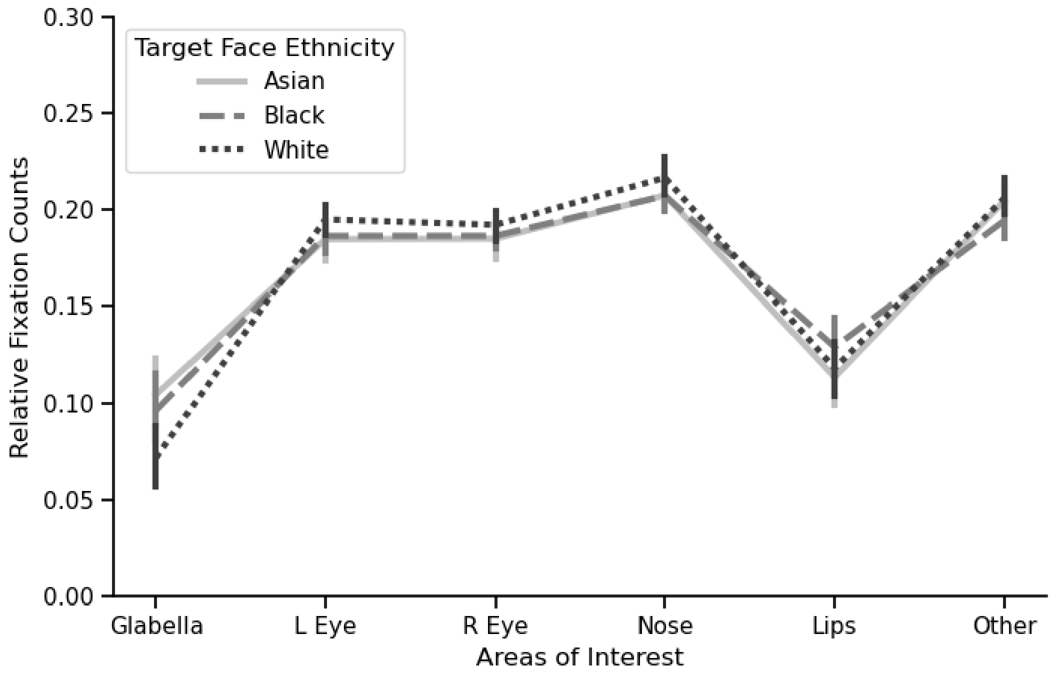
We found a small, but significant interaction between Face Ethnicity and AOI (Figure 4.11), F(10, 410) = 11.735, $p < .001$, $\eta p^2 = .223$, where we found similar trends from the Attractiveness task, where all three target faces received similar RFCs for the Nose, Other and both eye regions ($p > .05$). However, the Glabella and Lips of Asian faces did not significantly differ in RFCs received ($p = .648$), while the Lips of Black and White faces were looked at more than their Glabella regions ($p < .05$).

Within AOIs, the Glabella of White faces received less fixations compared to Glabella of Asian and Black faces (both $p < .001$), which received similar fixations ($p = 1.00$). The Lips of Black faces received more fixations than Asian Lips ($p = .001$) than Lips of Asian faces. Lips of White faces received similar fixations to both Black and Asian faces (both $p > .05$). No other comparisons were significant.

No other significant two-way interactions were found, where Participant Ethnicity and AOI, F(5, 205) = .1.400, p = .226, ηp^2 = ..033, and Face Ethnicity and Participant Ethnicity, F(2, 82) < 001, p = 1.000, $\eta p^2 < .001$. Finally, no three-way interactions were found between the variables, F(10, 410) = .785, $p = .644$, np2 = 0.19 The complementary Bayes factor analysis showed that these are likely to be under the null hypothesis (BF_{01} =1.895x10+21, BF_{01} = $1.282x10^{+61}$, and BF_{01} =11589.807), respectively.

Figure 4.11

Interactions of AOI and Target Face Ethnicity for Relative Fixation Counts on Health Task.



To summarise the findings (Table 4.10):

1. There was a significant main effect of AOIs, whereas no significant main effects were found for Face Ethnicity and Target Face Ethnicity for both durations and RFCs across all tasks. For the RFCs, the Nose, both eye regions, and Other areas received similar proportions of fixations across all tasks. For duration, some differences in the durations received by the AOIs were observed for different tasks.
    2. Looking at the duration variable, we saw a significant two-way interaction between AOI and Participant ethnicity for Age and Attractiveness tasks only. There were some differences in the durations for each AOI from participant groups.
    3. Looking at the RFCs variable, we saw a significant two-way interaction between AOI and Face Ethnicity across all tasks. The Nose, Other, and both eye regions received similar RFCs for all face ethnicities. For Asian faces, the Glabella and Lips received similar RFCs, whereas for Black and White faces, the Lips received slightly more RFCs than the Glabella across all tasks.
    4. We did not find evidence of a significant three-way interaction across all tasks.

Table 4.10.

Summary of Findings for Eye-tracking Study.

Variables	Age Durations	Age RFCs	Attractiveness Durations	Attractiveness RFCs	Health Durations	Health RFCs
AOI	/ *Nose, L, R eyes > Other > Glabella, Lips*	/	/ *Nose > L, R eyes > Lips, Other > Glabella*	/	/ *Nose > L, R eyes > Lips, Glabella, Other*	/
Face Ethnicity	x	x	x	x	x	x
Participant Ethnicity	x	x	x	x	x	x
AOI x Face Ethnicity	x	/	x	/	x	/
AOI x Participant Ethnicity	/ *Caucasians: Nose, R eye > L eye, Other > Glabella, Lips; Asians: Nose, L, R eye, Other > Lips, Glabella*	x	/ *Caucasians: Nose, L, R eye > Other, Lips > Glabella; Asians: Nose, L, R eye > Other, Lips > Glabella*	x	x	x
Face Ethnicity x Participant Ethnicity	x	x	x	x	x	x
AOI x Face Ethnicity x Participant Ethnicity	x	x	x	x	x	x

Note: / denotes a significant effect. x denotes non-significant effect.

Discussion

The current study explored whether an OEE will be observed when Asian and Caucasian participants make judgments on facial age, attractiveness, and health of Asian, Black, and Caucasian target faces by examining how long participants looked (durations) and which areas of the face they looked at most (relative fixation counts). As there are broad trends in findings between the three tasks, we will discuss them together.

Main effects

Across all tasks, we found support for hypothesis 1 where we found a main effect of AOI on durations and RFCs. For RFCs, there was a consistent trend where the Nose, Other, and both eye regions received similar proportion of fixations. However, there are some differences in how the AOIs received durations. For the age task, the Nose and both eye areas received similar durations, followed by Other and then Glabella and Lips. For attractiveness, the Nose received the highest duration, followed by both eye areas, Lips and Other, and finally the Glabella. For the health task, the Nose also received the highest duration, followed by both eye areas, and the rest of the face received similar durations. This is an interesting finding as considering that the variables were the same for the three social perception tasks, and the high associations between these three traits, the attention given to each facial area in terms of duration differed as a function of tasks. The general trend, however, remains that the Nose and eye regions received the longest durations – similar to the allocation of fixations overall.

These findings provide partial support to previous studies which found that the eye regions were looked at more compared to other regions for attractiveness (Kwart et al., 2012) and age estimation tasks (Nguyen et al.., 2009). The findings for the Nose being looked at with the same proportion was new – this could be an artefact of the placement of the fixation cross prior to each trial the current study. The eye-tracking device needed to be calibrated by looking at a fixation cross before each trial, which was placed in the middle of the screen, a similar location to the Nose area, which could account for the higher duration and fixation counts attributed to the Nose. However, previous studies that changed the start position of the fixation cross showed that the starting position did not affect their results (Arizpe et al., 2016). Furthermore, we aimed to reduce this effect by not including the first fixation index of each trial for our analysis; however, this could still have had a carry-over effect, as participants were only given six seconds to look at the target face. For future studies, researchers could consider varying the position of the fixation cross across trials to reduce the influence of starting position on duration and fixation counts.

We did not find support for hypothesis 2 and 3 for both variables, where we did not find a significant main effect of Participant Ethnicity (hypothesis 2) and Target Face Ethnicity (hypothesis 3). The Bayes Factor analysis further showed that we have strong evidence for the *RFC* data that the findings were more likely under the null effect. This implies that Asian and Caucasian participants had a similar trend in how they viewed the faces, and that Asian, Black, and White faces were looked at similarly.

Similarly, the Bayes Factor analysis also showed strong evidence that findings for Target Face Ethnicity for *duration* variable were more likely to be a null effect; that is, we could conclude that given our data, the target faces received similar durations across the three tasks. On the other hand, despite the non-significant main effect found, our data showed support that the Participant Ethnicity factor was more likely under the alternative hypothesis, even when we increased the precision of the estimation for this main effect across the three tasks. This could be an example of Jeffrey-Lindley's Paradox (Wagenmakers & Ly, 2023) when Bayesian and Frequentist tests disagree. In this case, we found a non-significant effect from our ANOVA, whereas the Bayes factor analysis indicated that, given the data, the effect was more likely under a non-zero alternative hypothesis. In addition, we do not have clear prior information that could be incorporated here to adjudicate the result – for example, while prior information around the effect of Participant Ethnicity can be found in recognition tasks where robust OEE's are often observed (Arizpe et al., 2016; Cenac et al., 2019, Childs et al., 2022), the focus here was on social perception tasks (Kwart et al., 2012; Nguyen et al., 2009), where no obvious prior information exists on varying participant ethnicity differences. We can only conclude therefore that this effect is small in size, given our estimated effect sizes, and its definitive presence or absence is unclear.

Furthermore, this could also be due to the time limit set for the task – all tasks were viewed for six seconds, therefore participants could only allocate six seconds worth of durations, regardless of their ethnicity. Perhaps future studies could incorporate a maximum time to view each face, but participants would be able to move on to rate the face whenever they were ready rather than having a set time for viewing.

Two-way and three-way interactions

RFCs. We did not find support for hypothesis 4, 6, and 7, where there were no significant two-way interactions between AOI and Participant Ethnicity (hypothesis 4); between Participant Ethnicity and Target Face Ethnicity (hypothesis 5), and all three variables respectively (hypothesis 7).

We only found support for hypothesis 5, where we found an interaction between AOI and Target Face Ethnicity. Overall, the Nose, Other and both eye regions were looked at similarly across all faces – the differences were observed on Glabella and Lips areas between the target faces on each task.

For age, the Lip area of Black faces were looked at more than Asian and White faces. For attractiveness, the Glabella area of Asian faces were looked at more than Black and White faces; and similar to age, the Lip area of Black faces were also looked at more than Asian and White faces. Finally, for health task, the Glabella of White faces were looked at less than Asian and Black faces; and interestingly, the Lip area of White faces received similar fixations to both Asian and Black faces, but the Lip area of Black faces were looked at more than Asian faces.

These findings were interesting as these suggest that how observers look at facial areas differ as a function of the target face ethnicity, regardless of their own ethnicity. This partially supports findings by Arizpe et al. (2016) where they found that Caucasian participants looked at target faces differently, that is, they looked at the eye regions more for Caucasian faces, and the nose and mouth areas of Asian and Black faces. However, our findings extend to this by incorporating another observer ethnicity and suggest that there was no other-ethnicity effect on how we perceive faces on judgments of attractiveness, health, and age, given that both Asian and Caucasian participants looked at the three target faces similarly. Perhaps the differences could be due to the task on-hand, where Arizpe et al. (2016) study looked at facial recognition, rather than judgments of facial traits. It could be that face recognition tasks use different processes due to the element of remembering the faces later on, compared to simply studying and judging faces on facial attractiveness, health, and age.

Durations. We found support for hypothesis 4 for age and attractiveness tasks, where we found a significant interaction between AOI and Participant Ethnicity. In both tasks, the Nose received the longest and the Glabella received the shortest durations. There were some differences in how the participant groups looked at the other facial areas. For age, the Left eye were looked at more than the R eye, Other and Lips areas, whereas for attractiveness, the Left and Right eyes received more durations than Other and Lips. This could highlight the idea that both eye regions are important when judging for attractiveness and therefore received similar durations, compared to age estimation.

We did not find support for hypothesis 5, 6, and 7, where there were no significant two-way interactions between AOI and Target Face Ethnicity (hypothesis 5); between

Participant Ethnicity and Target Face Ethnicity (hypothesis 5), and all three variables respectively (hypothesis 7) across all tasks. The Bayes Factor analysis that these effects were likely under the null hypothesis, therefore, we could conclude that no OEE is observed when making judgements on facial attractiveness, health, and age.

Overall discussion

A consistent finding of the current study was that the AOIs play a significant role when making judgments on facial age, attractiveness and health for both duration and RFCs. Across three tasks, the Nose received the highest durations and fixations, followed generally by the Eyes and Other regions, and lastly Lips and Glabella regions. We also consistently found no main effect of Target Face Ethnicity and Participant ethnicity, which suggests that overall, participants look at faces similarly, and target faces receive similar durations and RFCs. However, this could be because of the set time for the observation prior to giving an evaluation. For future studies, researchers could consider giving the participant control of how long they would like to look at a face prior to making an evaluation. This could give more insight whether own- or other-ethnicity faces were looked at more before giving a response.

One interesting point is that although we first established that these three traits behaviourally seem to be associated with each other, i.e., the more attractive a face is the healthier and younger it appears to be, there were some differences in how the faces were looked at, that is, some facial regions received longer durations or more fixations depending on the task. Generally, the Nose, both eye regions and Other were the ones receiving the most attention across all tasks. The eye regions receiving a lot of attention supports previous eye-tracking studies (Kwart et al., 2012; Nguyen et al., 2009). The incorporation of Nose and Other facial areas were a novel finding for perception of social judgments. These four areas generally receiving similar attention suggests that these areas play a significant role in perception of social judgment.

The main aim of the current study is to determine whether an OEE is present when observers make social judgments on own- and other-ethnicity faces. Although we did not find support for this, that is, we did not find a significant interaction between the three variables across the tasks, we did find an interaction between AOI and Target Face Ethnicity for the RFC variable. We found that the Glabella and Lips – the two areas that received the least RFCs, differed in the overall RFCs received as a function of Target Face Ethnicity. It could therefore be that faces are looked at differently as a function of their ethnicity, regardless of the observer's *own* ethnicity. Future studies could expand on these findings and focus on the aspects between these features across ethnic groups to gain insight on the mechanisms behind these differences. It is important to note, however, that we did not find this interaction for the duration variable, although changing the maximum duration for each face and allowing participants to provide their answers at any point of the trial could yield more interesting results.

It is important to note that as far as we know, the current set of studies were the first to demonstrate that no OEE is observed when perceiving facial attractiveness, health, and age. This implies that observers look at faces in a similar way when judging social traits, contradicting findings from recognition tasks.

A limitation of the current study is that we did not have data regarding participants' contact to other ethnicities, and therefore we are not able to make conclusions as to whether the interaction effects we found for AOI and Target Face Ethnicity on RFC variable were due to differences in exposure to the other ethnic groups. For future studies, it could be beneficial to include a contact questionnaire to establish whether contact has an influence on how faces from different ethnic groups are viewed. Another limitation of the current study is that we did not use the actual age of target faces – Nguyen et al. (2009) found that participants looked at the glabella area more when faces were judged older, and therefore the target's face could have an influence on how participants looked at the stimuli. In future, we could use an alternative age variable, e.g. age estimates from a sample prior to the experiment itself, and use target faces with a similar age group.

To conclude, the current study aimed to establish whether an OEE would be observed when people make judgments on facial age, attractiveness, and health. This is also the first study to investigate which facial areas are used when people make judgments on facial health, and also expanded on previous studies looking at evaluations of facial attractiveness and age by incorporating two participant ethnicities and three target face ethnicities. Our data suggest that unlike facial recognition tasks, no OEE is observed when people make judgments on facial attractiveness, health and age. Furthermore, we found a general trend where the Nose, Left and Right eyes, and Other areas received the most attention when making social judgment. This could have practical implications on which facial areas to target for enhancement of facial appearance.

Chapter 5.

General Discussion

Discussion

This thesis examined used various methodologies to explore a range of topics and contributed to areas of evolutionary psychology and face perception.

Using survey and vignettes, I explored the role of intrasexual competition (IC) in perception of types of anti-ageing treatment and individuals who undergo them, and its relationship with skincare behaviour and other personality traits. I employed experimental studies to investigate the role of motion, emotion, and gender on perception of strongly associated facial social traits, i.e., attractiveness, health, and age. Finally, I investigated whether the magnitude of other-ethnicity effects (OEE) on facial recognition using three facial recognition tasks, and explored the role of OEE on perception of facial social traits as a function of participant ethnicity and target face ethnicity using eye tracking methods.

In this section, I provide a recap of the key findings from each experimental chapter and relate the findings to the wider literature in turn.

5.1 Women's competitive traits are related to their view of aesthetic dermatology.

In Chapters 2.1 and 2.2, I looked at how intrasexual competitiveness trait was related to skincare behaviours, attitudes towards aesthetic behaviour, and perception of individuals who engage in age concealment.

Given that previous research have shown that behaviours associated with beautification strategies such as the use of make-up and cosmetic procedures could be predicted by an individual's competitiveness towards the same-sex (Mafra et al., 2020; Wang et al., 2021), in chapter 2.1, I sought to determine whether this could also be observed on behaviours relating to a simple everyday skincare routine. The use of skincare products such as cleansers, moisturisers, toners, and serums have been shown to improve the appearance of facial skin and even reverse the effects of ageing (Mangal et al., 2021; Messaraa et al., 2020)– similar to the benefits promised by make-up and cosmetic procedures. However, I did not find support for this hypothesis. This chapter provided an unexpected and novel finding, where skincare behaviour (expenditure and use) did not predict a woman's level of intrasexual competitive traits.

A reason for this could be that make-up and cosmetic procedures are used as a means-to-an-end. That is, they are used for the purpose of enhancing one's appearance either temporarily or more long-term, while use of skincare does not provide such an impact. Although the use of skincare products on its own could be considered an enhancement strategy, their results tend to be more subtle than use of make-up and cosmetic surgery, and perhaps using skincare products could be seen as a self-care regime, that is, something that women do for themselves (Chao, 2021), rather than for other people. In support of this, a recent study has shown that individuals who used skincare products for 28 days generally showed significant improvement in their feelings of empowerment, happiness and self-esteem (Zhang et al., 2020).

Next, we found that among the other traits explored; only Perceived Results from Aesthetic Dermatology (PRAD) and Anxiety Against Ageing (AA) had more than 95% probability of having a positive effect on the IC trait. This means that the better one thinks of the outcome from cosmetic procedures and the more anxious they are of appearing older, the higher their competitive traits are. Additionally, self-esteem and Knowledge About Aesthetic Dermatology (KAAD) had more than 95% probability of having a negative effect on the IC trait, that is, the less confident someone is and the less they know about cosmetic procedures, they are likely to be less competitive. This is in line with the idea that women that are more competitive would want to know more about and were more positive about the outcomes of enhancing their appearance. I also want to highlight here that individuals worrying about their appearance have higher probability of being more competitive, which could explain their more positive outlook towards the aesthetic dermatology results. This supports previous studies that indicated that those who were more likely to engage in aesthetic dermatology treatments were higher in competitiveness (Arnocky & Piché, 2014; Wang et al., 2021).

In Chapter 2.2, I investigated how an observer's competitiveness would influence their evaluations of targets who used age concealment techniques. Additionally, I was interested to see whether the previous negative schema towards individuals who engage in age concealment has changed, considering the shift in the prevalence of more affordable, and less invasive age concealment techniques.

I found that generally, the perception remains the same – more invasive procedures received more negative evaluations compared to milder, less invasive procedures, in line with previous studies (Chasteen et al., 2011; Harris, 1994), despite the increased prevalence and accessibility of anti-ageing procedures. It is important to note, however, that the newer anti-ageing procedures and devices that are coming out are still considered less invasive than Botox and full face-lift, which could be influencing the perceptions of those engaging in more invasive techniques.

Most importantly, this study found that an observer's competitive trait plays a significant role in their evaluations – those with higher competitiveness gave a lower evaluation overall, compared to those with lower competitiveness. This highlights the idea that other women engaging in appearance enhancement strategies could be seen as a threat by those who have higher competitiveness trait, that is, there would be more women to compete against (Wang et al., 2021). This could therefore lead to more competitive women to engage in another strategy available to them – degradation of the competitor, hence the less positive evaluations overall. One of the limitations of this chapter however, is that we did not include a competitive context in

Finally, the motivation behind the engagement also plays a role on how the person was perceived – targets received the most positive evaluation when they wanted to do it for self-esteem purposes, followed by employment and finding a partner the least (Harris, 1994). In today's society, and in particular following the lockdown restrictions during COVID-19 pandemic (Gill & Orgad, 2022) there is a drive towards self-care and self-empowerment for women. This could be seen in social media where campaigns of positivity and pursuit of individual happiness is encouraged to be the norm. For example, Dove Self Esteem Project (Dove, 2023) was launched in 2004 and they collaborated with Unicef with the aim to empower young girls to build their self-esteem and confidence (Unicef, n.d.). Another example is Body Shop's Rise Up With Self Love campaign with the aim to help people believe and accept themselves (The Body Shop, 2023). This general shift in societal perspective could have a role as to why motivations of self-esteem and employment to enhance one's appearance were viewed more positively compared to motivations of romantic nature. The role of social media and by extension, advertisement of anti-ageing products on consumption of age concealment products such as skin care products to more professional treatments is not within the remit of this thesis, however, the pervasive nature of social media is something that needs to be considered in future studies.

Taken together, chapters 2.1 and 2.2 show two, linked findings – the characteristics of individuals who would be likely to engage in age concealment, and how others view the individual. On the one hand, this could have applications for marketing products that have beautification elements, informing on which audience would be more responsive to their products. On the other hand, this could guide consumers of age concealment on how to interact with people before and after their procedures, where having an understanding of the social implications of engaging in anti-ageing procedures could prepare individuals on how they navigate their social interactions.

These two chapters also have interesting implications for the evolutionary psychology perspective on intrasexual competition and mating. As the technology used in aesthetic dermatology becomes more prevalent and their results advance, the concept of an *'aging face'* could change in the future. Females may be able to maintain their youthful looks for longer, say a 40 year old passing as a 30 year old (or even younger), but the biological, reproductive function is still restricted, i.e., the older a woman gets, the less fertile she becomes (Maestriperi et al., 2014). This means that women will be able to use self-enhancement to make themselves more attractive to the opposite sex, however, female attractiveness may not be a valid cue for reproductive value in this instance. In line with this, some evidence suggest that facial attractiveness do not signal health in general (Cai et al., 2019; Foo et al., 2017; Kalick et al., 1998). It will therefore be interesting to see the implications of increased attractiveness in older age may have on seeking mates.

Additionally, it would be interesting to see if other strategies of intrasexual competition emerge due to the prevalence of self-enhancement techniques or whether other strategies such as degradation would be used more often and more aggressively.

5.2 Facial attractiveness and health were perceived differently despite high correlations.

The literature, and the current set of studies, demonstrated that facial attractiveness and facial health have a strong, positive relationship – where as a face is perceived as more attractive, the healthier they also appear to be. These traits even share similar facial cues and characteristics overall. However, in Chapter 3, we demonstrate that there are some divergence in how these two traits were perceived.

First, it is important to mention that I found that age estimates remained stable across emotion, motion, and age group. I therefore decided to focus on attractiveness and health and aimed to disentangle the effects of these factors in the subsequent studies.

In Chapter 3.1 and in extension Chapter 3.2, I found an age bias for both attractiveness and health traits, where young faces were generally perceived as more attractive and healthier than and older faces, however the similarity ends there. Gender bias was only observed in attractiveness ratings, where female faces were rated as more attractive; and motion affected perceptions differently for both genders – moving faces were more attractive, while static faces were perceived to be healthier. In Chapter 3.1, I also demonstrated that while a positive bias boosted ratings of attractiveness, this did not affect health ratings.

The gender bias in attractiveness evaluations highlights the idea that beauty (and indeed youth) is more pertinent to females than males, supporting evolutionary theories on mating strategies (Buss, 1989; Buss & Schmitt, 1993; Buss & Schmitt, 2019). For female faces, we found a linear decline as a function of age, where young female faces received the highest ratings, followed by middle aged, and the old faces receiving the lowest attractiveness ratings. For male faces, there was still the benefit of youth as young male faces received the highest ratings, but I saw that middle and old faces received similar attractiveness ratings. The findings therefore provide support to evolutionary theory of youth being more important for females than for males (Buss, 1989; Buss & Schmitt, 1993; Buss & Schmitt, 2019). Surprisingly, we found no gender bias for health evaluations – both genders were treated the same for each age group.

One explanation for this could be that perceivers use different cues when judging for attractiveness and health, and that the cues used for health perception are consistent regardless of gender. The literature has inconsistent findings as to which factors are more predictive of one trait, but studies have demonstrated that skin colour distribution (Fink et al., 2006; Jones, 2018; Jones et al., 2016; Samson et al., 2011), and averageness (Jones, 2018) independently predict facial health, while averageness, sexual dimorphism and texture contribute to predicting perceived attractiveness (Jones & Jaeger, 2019). From here, it seems that there are facial features that contribute to perception of one social trait more than the other, and this could explain why there is a gender bias in perception of attractiveness, but not in facial health.

Both facial traits have demonstrated a positivity bias, where smiling faces were perceived as more attractive and healthier, supporting previous studies which found that positive affect is more attractive (Ebner et al., 2018) and healthier (Jones et al., 2018). Ebner et al. (2018) argued that the positivity effect on facial attractiveness could be explained as smiling faces generating a positive emotional response to the perceiver, allowing for more positive evaluation of the target.

The difference in motion bias however, highlights the notion that different facial features contribute to perception of attractiveness and health differently. It is important to note that the effect of motion was only observed for smiling faces. Neutral faces received similar evaluations of attractiveness and health, however, dynamic smiling faces were evaluated as more attractive, while static smiling faces were evaluated as healthier. One explanation for this is that dynamic faces allows for some facial features to be more salient than others (Rubenstein, 2005), e.g. facial texture and wrinkles

It is important to note that there is a notable lack of interactions between target sex, emotion, and motion in these two traits despite their strong association with each other. This provides support that observers use different cues and signals when making social judgments. This also highlights perhaps the most important facial feature – age. Note that for attractiveness, there was an interaction of facial age and target sex – this was not observed for facial health where sexes in the different age groups received similar health evaluations.

From here, we could infer that salient signs of facial aging is more pertinent to perception of attractiveness than facial health, particularly as a function of sex.

These findings are the first to demonstrate facial age estimates remained stable across facial sex, motion, and emotion, as well as the first study to investigate the effects of these factors on perceptions of facial health.

Additionally, these findings are also the first to demonstrate that different facial factors such as age, sex, emotion, and motion play a different role to perceptions of facial attractiveness and health. Future research should therefore consider this to ensure that the contribution of these facial factors to specific social traits are examined and accounted for. In particular, this study underlined the importance of incorporating both static and dynamic stimuli when investigating social traits as only using one motion may result in some effects being overlooked and making overgeneralisations of results based on one stimuli type.

Although this thesis did not directly manipulate the facial features commonly related to perceptions of these traits, e.g., symmetry, averageness, and sexual dimorphism, the findings demonstrated that despite the high correlations generally found between evaluations of facial attractiveness and health, and in turn, facial age, these facial traits are perceived independently of each other. In sum, researchers should not assume that facial cues and features contribute similarly to these traits.

5.3 Magnitude of OEE is similar across own ethnicity face recognition abilities

Another salient facial characteristic is the categorisation of ethnicity. The literature has reliably shown an OEE on facial recognition, where individuals are better at recognising faces from their own ethnic group compared to those from other ethnic group. OEE has been reliably shown using recognition and perceptual tasks, however, the degree of this difference as a function of whether an individual is good or bad at recognising faces from their own group has not been investigated. In chapter 4.1, I provided support to the OEE, where both Asian and White participants performed better at recognising faces from their own ethnic groups. However, regardless of how well they performed on own ethnicity faces, their ability to recognise faces from other ethnicity group was not disproportionally better or worse. That is, the good-recognisers in the study did not have a boosted performance, nor did the bad-recognisers show a floor effect. The findings show that the performance on other ethnicity faces was generally the same as the own ethnicity faces, just that individuals were scoring lower on the other ethnicity test. This implies that OEE is a persistent feature of facial recognition ability and the quality of this effect is not boosted or diminished disproportionally based on own ethnicity recognition.

This work has important applications – when classifying face recognition ability a cut-off of scores (2SDs) is usually applied, where individuals who scored lower than 2SDs were categorised as prosopagnosics (Wan et al., 2017) or those who perform incredibly well are categorised as super-recognisers. Previous studies compared these groups' performance on own- and other- ethnicity faces to the performance of the general population, limiting the scope of their findings. Our study did not apply any such classification, which allowed us to observe the full range of performance across the population. With the finding that the magnitude of OEE is similar across the population regardless of performance, we should be able to step away from grouping individuals as 'poor' or 'good' recognisers and use a wider range of abilities.

One interesting finding from this study is that individuals who performed very poorly in the own ethnicity test manifest an OEE similar to the rest of the population. One argument for the poor own ethnicity face recognition performance is that individuals are not able to learn the target face and hence not able to distinguish between faces. Our findings therefore show has an important theoretical implication as this suggests that there is visual learning happening despite of the inherited susceptibility to generally poor face processing ability, otherwise the performance should be at chance-level. On a similar note, those with inherited susceptibility to generally good face processing were also largely superior in recognising other ethnicity faces compared to the rest of the population. This is in line with findings that facial recognition is highly heritable and correlated with general cognitive abilities. (Shakeshaft & Plomin, 2015). This has practical uses, for example, employing super-recognisers for roles that require good face recognition ability such as ID verifications.

5.4 Facial areas used for perceptions of social traits.

Facial recognition studies have also demonstrated that individuals pay attention to different facial areas as a function of target ethnic group when studying target faces. In chapter 4.2, I contributed to this topic by using eye-tracking techniques to investigate which facial areas individuals pay attention to when judging for social traits – facial attractiveness, health and age, as well as employing two ethnic groups (Asian and White) to look at three target ethnicity face groups (Asian, Black, and White).

Here I found that in general, the nose, eye regions, and ‘other’ regions, that is, regions outside the nose, eyes, glabella, and lips, received the most attention across tasks and ethnic groups, while Glabella and Lips regions received the least attention. This generally supports previous studies where the eye regions receive the most fixations and durations. The finding that the nose area received the most attention across the ethnic group was unexpected. This could be an artefact of the study design, where the fixation cross was placed in the middle of the screen prior to the target image appearing. However, as part of our data cleaning, we removed the first fixation index for each trial and the fixations to the nose persists. Future studies could change the fixation cross for every trial to reduce this bias. Apart from this, there were no other main differences in how the target faces were looked at by both participant groups.

In general, for duration variable, we only found an interaction between areas of interest and participant ethnicity on age and attractiveness tasks, and for relative fixation counts, we only found an interaction between areas of interest and target face ethnicity across the three tasks.

5.4.1 No support for OEE when estimating facial attractiveness, health and age.

Overall, we found no support of OEE on facial age, where Asians and Whites are judging for age, they look at facial regions similarly. This is an interesting finding, as I expected to see different patterns emerge. The literature on skin aging has suggested that different ethnic groups show different patterns of signs of aging, where Caucasians tend to develop wrinkles earlier compared to their Asian and Black counterparts, and Asian and Black ethnic groups tend to have uneven skin tone and hyperpigmentation as their main concerns (Rawlings, 2006). Perhaps perceivers are not consciously aware of these differences and therefore are not able to utilise them? It is also important to note that our sample was made up of undergraduate students; therefore, signs of aging at this age group is not as salient as it would be for older age groups. On the other hand, there is a general pattern of aging (Michaud et al., 2015), and perhaps as this is similar across ethnicities, we therefore use the same pattern of eye movements to look at a face when estimating for one's age.

Additionally, Asians and Black individuals are said to be difficult to estimate ages for, therefore I expected that participants would be looking at these two target faces differently than they would a Caucasian face. It could well be that as cues for age vary across ethnicities individuals simply look at the faces the same way and reach an estimation. However, as we did not investigate whether their age estimates were accurate, we cannot establish whether there is an OEE in age estimation. Future studies could explore the accuracy of age estimates as a function of participant and target face ethnicity alongside tracking eye movements.

This finding could have applications for the cosmetic industry, as this implies that observers used similar facial cues to estimate age regardless of the ethnicity of the target face. This could mean that one procedure that works to reduce the signs of clinical aging for one ethnicity would also work for another. This could help shape the development of anti-ageing procedures in the future.

We also did not find support for OEE for attractiveness and health tasks overall, as we did not find a three-way interaction between areas of interest, target face ethnicity, and participant ethnicity. Overall, the data suggest that Asian and Caucasian participants looked at all target faces similarly. There was a subtle difference in how the areas of interest were looked at in terms of duration for attractiveness, and also for relative fixations across the three tasks, but the general pattern remained the same – the facial areas that were paid the most attention were the Nose, both eye regions, and 'other' areas. I did find subtle differences in how much fixations the Lips and Glabella were looked at by Asian and Caucasian participants across the three tasks and these could be probed further in future research.

Overall, I did not find that observers looked at their own-ethnic group facial features more than the others, contradicting findings from Arizpe et al's recognition task (2016). This provides support that individual's patterns of looking at a face differs as a function of task (Kannan et al., 2015).

5.5. Conclusions

In sum, the current thesis investigated different facets relating to perceptions of facial traits. This work provided evidence that perceptions towards use of anti-ageing treatments have not changed despite the increasing popularity and accessibility of such procedures, and demonstrated that skincare behaviour, unlike make-up and cosmetic procedures, was not predictive of female intrasexual competitiveness.

Additionally, I contributed to the face recognition literature by demonstrating that the magnitude of other-ethnicity affect remains stable across own-ethnicity face recognition ability.

Finally, this work contributed to the face processing literature by demonstrating that despite close associations between facial age, attractiveness and health, perceptions of facial age appears to be more stable across various factors such as ethnicity, motion, emotion, and gender, compared to facial attractiveness and health. To the best of my knowledge, this work is also the first to demonstrate that no OEE is observed using eye-tracking technique when judging on facial attractiveness, health, and age.

Chapter 6.

References

American Society of Plastic Surgeons. (2020). *Plastic surgery statistics*. https://www.plasticsurgery.org/news/plastic-surgery-statistics

Anderson, S. L., Adams, G., & Plaut, V. C. (2008). The cultural grounding of personal relationship: The importance of attractiveness in everyday life. *Journal of Personality and Social Psychology*, *95*(2), 352–368. https://doi.org/10.1037/0022-3514.95.2.352

Andersson, S. (1989). Sexual selection and cues for female choice in leks of Jackson's widowbird Euplectes jacksoni. *Behavioral Ecology and Sociobiology*, *25*(6), 403–410. https://doi.org/10.1007/BF00300186

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental Origins of the Other-Race Effect. *Current Directions in Psychological Science*, *22*(3), 173–178. https://doi.org/10.1177/0963721412474459

Archer, J., & Coyne, S. M. (2005). An Integrated Review of Indirect, Relational, and Social Aggression. *Personality and Social Psychology Review*, *9*(3), 212–230. https://doi.org/10.1207/s15327957pspr0903_2

Arizpe, J., Kravitz, D. J., Walsh, V., Yovel, G., & Baker, C. I. (2016). Differences in Looking at Own- and Other-Race Faces Are Subtle and Analysis-Dependent: An Account of Discrepant Reports. *PLOS ONE*, *11*(2), e0148253. https://doi.org/10.1371/journal.pone.0148253

Arnocky, S., & Piché, T. (2014). Cosmetic Surgery as Intrasexual Competition: The Mediating Role of Social Comparison. *Psychology*, *05*(10), 1197–1205. https://doi.org/10.4236/psych.2014.510132

Arnocky, S., Proietti, V., Ruddick, E. L., Côté, T.-R., Ortiz, T. L., Hodson, G., & Carré, J. M. (2019). Aggression Toward Sexualized Women Is Mediated by Decreased Perceptions of Humanness. *Psychological Science*, *30*(5), 748–756. https://doi.org/10.1177/0956797619836106

Barrera, M. E., & Maurer, D. (1981). Recognition of Mother's Photographed Face by the Three-Month-Old Infant. *Child Development*, *52*(2), 714. https://doi.org/10.2307/1129196

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2019). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 363–377. https://doi.org/10.1037/xhp0000607

Bate, S., Bennetts, R. J., Tree, J. J., Adams, A., & Murray, E. (2019). The domain-specificity of face matching impairments in 40 cases of developmental prosopagnosia. *Cognition*, *192*, 104031. https://doi.org/10.1016/j.cognition.2019.104031

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4*.

Bate, S., & Tree, J. J. (2017). The Definition and Diagnosis of Developmental Prosopagnosia. *Quarterly Journal of Experimental Psychology*, *70*(2), 193–200. https://doi.org/10.1080/17470218.2016.1195414

Baudouin, J.-Y., & Tiberghien, G. (2004). Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychologica*, *117*(3), 313–332. https://doi.org/10.1016/j.actpsy.2004.07.002

Beilin, G. (2011). Home-use TriPollar RF device for facial skin tightening: Clinical study results. *Journal of Cosmetic and Laser Therapy*, *13*(2), 69–76. https://doi.org/10.3109/14764172.2011.552607

Bennett, E. V, Clarke, L. H., Kowalski, K. C., & Crocker, P. R. E. (2017). "I'll do anything to maintain my health": How women aged 65–94 perceive, experience, and cope with their aging bodies. *Body Image*, *21*, 71–80. https://doi.org/10.1016/j.bodyim.2017.03.002

Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of Personality and Social Psychology*, *113*(4), 530–546. https://doi.org/10.1037/pspa0000091

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture Shapes How We Look at Faces. *PLoS ONE*, *3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-Movement Strategies in Developmental Prosopagnosia and "Super" Face Recognition. *Quarterly Journal of Experimental Psychology*, *70*(2), 201–217. https://doi.org/10.1080/17470218.2016.1161059

Bowdring, M. A., Sayette, M. A., Girard, J. M., & Woods, W. C. (2021a). In the Eye of the Beholder: A Comprehensive Analysis of Stimulus Type, Perceiver, and Target in Physical Attractiveness Perceptions. *Journal of Nonverbal Behavior*, *45*(2), 241–259. https://doi.org/10.1007/s10919-020-00350-2

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

Bradshaw, H. K., Leyva, R. P., Nicolas, S. C. A., & Hill, S. E. (2019). Costly female appearance-enhancement provides cues of short-term mating effort: The case of cosmetic surgery. *Personality and Individual Differences*, *138*, 48–55. https://doi.org/10.1016/j.paid.2018.09.019

Brooks, A. T. (2010). Aesthetic anti-ageing surgery and technology: women's friend or foe? *Sociology of Health & Illness*, *32*(2), 238–257. https://doi.org/10.1111/j.1467-9566.2009.01224.x

Burns, E. J., Bennetts, R. J., Bate, S., Wright, V. C., Weidemann, C. T., & Tree, J. J. (2017). Intact word processing in developmental prosopagnosia. *Scientific Reports*, *7*(1), 1683. https://doi.org/10.1038/s41598-017-01917-8

Burns, E. J., Martin, J., Chan, A. H. D., & Xu, H. (2017). Impaired processing of facial happiness, with or without awareness, in developmental prosopagnosia. *Neuropsychologia*, *102*, 217–228. https://doi.org/10.1016/j.neuropsychologia.2017.06.020

Burns, E. J., Tree, J. J., & Weidemann, C. T. (2014). Recognition memory in developmental prosopagnosia: electrophysiological evidence for abnormal routes to face recognition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00622

Burt, D. M., & Perrett, D. I. (1997). Perceptual asymmetries in judgements of facial attractiveness, age, gender, speech and expression. *Neuropsychologia*, *35*(5), 685–693. https://doi.org/10.1016/S0028-3932(96)00111-X

Buss, D. M. (1988). The evolution of human intrasexual competition: Tactics of mate attraction. *Journal of Personality and Social Psychology*, *54*(4), 616–628. https://doi.org/10.1037/0022-3514.54.4.616

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*(1), 1–14. https://doi.org/10.1017/S0140525X00023992

Buss, D. M., & Schmitt, D. P. (1993). Sexual Strategies Theory: An evolutionary perspective on human mating. *Psychological Review*, *100*(2), 204–232. https://doi.org/10.1037/0033-295X.100.2.204

Buss, D. M., & Schmitt, D. P. (2019). Mate Preferences and Their Behavioral Manifestations. *Annual Review of Psychology*, *70*(1), 77–110. https://doi.org/10.1146/annurev-psych-010418-103408

Buunk, A. P., & Fisher, M. (2009). Individual differences in intrasexual competition. *Journal of Evolutionary Psychology*, *7*(1), 37–48. https://doi.org/10.1556/JEP.7.2009.1.5

Cai, Z., Hahn, A. C., Zhang, W., Holzleitner, I. J., Lee, A. J., DeBruine, L. M., & Jones, B. C. (2019). No evidence that facial attractiveness, femininity, averageness, or coloration are cues to susceptibility to infectious illnesses in a university sample of young adult women. *Evolution and Human Behavior*, *40*(2), 156–159. https://doi.org/10.1016/j.evolhumbehav.2018.10.002

Campbell, A. (2004). Female competition: Causes, constraints, content, and contexts. *Journal of Sex Research*, *41*(1), 16–26. https://doi.org/10.1080/00224490409552210

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2022). *Bambi: A simple interface for fitting Bayesian linear models in Python*.

Cavallo, P., Proto, M. C., Patruno, C., Sorbo, A. Del, & Bifulco, M. (2008). The first cosmetic treatise of history. A female point of view. *International Journal of Cosmetic Science*, *30*(2), 79–86. https://doi.org/10.1111/j.1468-2494.2007.00414.x

Cenac, Z., Biotti, F., Gray, K. L. H., & Cook, R. (2019a). Does developmental prosopagnosia impair identification of other-ethnicity faces? *Cortex*, *119*, 12–19. https://doi.org/10.1016/j.cortex.2019.04.007

Chao, J. (2021). *Skincare for your soul: Achieving outer beauty and inner peace with korean skincare*. Mango Media.

Chasteen, A. L., Bashir, N. Y., Gallucci, C., & Visekruna, A. (2011). Age and antiaging technique influence reactions to age concealment. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *66 B*(6), 719–724. https://doi.org/10.1093/geronb/gbr063

Chaterjee, A. (2007). Cosmetic Neurology and Cosmetic Surgery: Parallels, Predictions, and Challenges. *Cambridge Quarterly of Healthcare Ethics*, *16*(02). https://doi.org/10.1017/S0963180107070156

Childs, M. J., & Jones, A. (2022). Perceptions of Individuals Who Engage in Age Concealment. *Evolutionary Behavioral Sciences*. https://doi.org/10.1037/ebs0000305

Childs, M. J., Jones, A., Thwaites, P., Zdravković, S., Thorley, C., Suzuki, A., Shen, R., Ding, Q., Burns, E., Xu, H., & Tree, J. J. (2021). Do individual differences in face recognition ability moderate the other ethnicity effect? *Journal of Experimental Psychology. Human Perception and Performance*, *47*(7). https://doi.org/10.1037/xhp0000762

Chiroro, P., & Valentine, T. (1995). An Investigation of the Contact Hypothesis of the Own-race Bias in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *48*(4), 879–894. https://doi.org/10.1080/14640749508401421

Chiu, W. C. K., Chan, A. W., Snape, E., & Redman, T. (2001). Age Stereotypes and Discriminatory Attitudes towards Older Workers: An East-West Comparison. *Human Relations*, *54*(5), 629–661. https://doi.org/10.1177/0018726701545004

Chopik, W. J., Bremner, R. H., Johnson, D. J., & Giasson, H. L. (2018). Age Differences in Age Perceptions and Developmental Transitions. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00067

Clarke, L. H., & Griffin, M. (2008). Visible and invisible ageing: beauty work as a response to ageism. *Ageing and Society*, *28*(5), 653–674. https://doi.org/10.1017/S0144686X07007003

Clarke, L. H., & Griffin, M. (2007). The body natural and the body unnatural: Beauty work and aging. *Journal of Aging Studies*, *21*(3), 187–201. https://doi.org/10.1016/j.jaging.2006.11.001

Clarke, L. H., Repta, R., & Griffin, M. (2007). Non-Surgical Cosmetic Procedures: Older Women's Perceptions and Experiences. *Journal of Women & Aging*, *19*(3–4), 69–87. https://doi.org/10.1300/J074v19n03_06

Coetzee, V., Greeff, J. M., Stephen, I. D., & Perrett, D. I. (2014). Cross-Cultural Agreement in Facial Attractiveness Preferences: The Role of Ethnicity and Gender. *PLoS ONE*, *9*(7), e99629. https://doi.org/10.1371/journal.pone.0099629

Correa-de-Araujo, R., & Yoon, S. S. (Sarah). (2021). Clinical Outcomes in High-Risk Pregnancies Due to Advanced Maternal Age. *Journal of Women's Health*, *30*(2), 160–167. https://doi.org/10.1089/jwh.2020.8860

Cox, A., & Fisher, M. (2008). A framework for exploring intrasexual competition. *Journal of Social, Evolutionary, and Cultural Psychology*, *2*(4), 144–155. https://doi.org/10.1037/h0099348

Cuddy, A. J. C., Norton, M. I., & Fiske, S. T. (2005). This Old Stereotype: The Pervasiveness and Persistence of the Elderly Stereotype. *Journal of Social Issues*, *61*(2), 267–285. https://doi.org/10.1111/j.1540-4560.2005.00405.x

Darrach, H., Ishii, L. E., Liao, D., Nellis, J. C., Bater, K., Cobo, R., Byrne, P. J., Boahene, K. D. O., Papel, I. D., Kontis, T. C., & Ishii, M. (2019a). Assessment of the Influence of "Other-Race Effect" on Visual Attention and Perception of Attractiveness Before and After Rhinoplasty. *JAMA Facial Plastic Surgery*, *21*(2), 96–102. https://doi.org/10.1001/jamafacial.2018.1697

Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2021). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods*, *54*(4), 1889–1901. https://doi.org/10.3758/s13428-021-01705-3

DeBruine, L. M., Jones, B. C., Unger, L., Little, A. C., & Feinberg, D. R. (2007). Dissociating averageness and attractiveness: Attractive faces are not always average. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1420–1430. https://doi.org/10.1037/0096-1523.33.6.1420

de Gelder, B., & Rouw, R. (2000). Paradoxical configuration effects for faces and objects in prosopagnosia. *Neuropsychologia*, *38*(9), 1271–1279. https://doi.org/10.1016/S0028-3932(00)00039-7

DeGutis, J., Mercado, R. J., Wilmer, J., & Rosenblatt, A. (2013). Individual differences in holistic processing predict the own-race advantage in recognition memory. *PLoS ONE*, *8*(4), e58253. https://doi.org/10.1371/journal.pone.0058253

De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science*, *13*(1), 181–187. https://doi.org/10.1111/j.1467-7687.2009.00876.x

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

Department for Work and Pensions. (2023, March 30). *State Pension Age Review published*. https://www.gov.uk/government/news/state-pension-age-review-published

Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception Studies—a review of behavioral findings and methodology. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01355

Dolsky, R. L. (1999). Cosmetic Surgery in the United States: Its past and present. *Dermatologic Surgery*, *25*(11), 886–892. https://doi.org/10.1046/j.1524-4725.1999.99008.x

Dove. (2023). *Dove Self-Esteem Project*. https://www.dove.com/uk/dove-self-esteem-project.html

Draelos, Z. D. (2015). Cosmetics: The medicine of beauty. *Journal of Cosmetic Dermatology*, *14*(2), 91–91. https://doi.org/10.1111/jocd.12146

Drevenstedt, J. (1976). Perceptions of onsets of young adulthood, middle age, and old age. *Journal of Gerontology*, *31*(1), 53–57. https://doi.org/10.1093/geronj/31.1.53

Dubbs, S. L., Kelly, A. J., & Barlow, F. K. (2015). *Ravishing Rivals* (M. L. Fisher, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199376377.013.35

Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, *23*(5), 714–747. https://doi.org/10.1080/02643290500441296

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Ebner, N. C. (2008). Age of face matters: Age-group differences in ratings of young and old faces. *Behavior Research Methods*, *40*(1), 130–136. https://doi.org/10.3758/BRM.40.1.130

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. https://doi.org/10.3758/BRM.42.1.351

Estudillo, A. J., Lee, J. K. W., Mennie, N., & Burns, E. (2020). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, *34*(1), 270–276. https://doi.org/10.1002/acp.3609

Fink, B., Bunse, L., Matts, P. J., & D'Emiliano, D. (2012). Visible skin colouration predicts perception of male facial age, health and attractiveness. *International Journal of Cosmetic Science*, *34*(4), 307–310. https://doi.org/10.1111/j.1468-2494.2012.00724.x

Fink, B., Grammer, K., & Matts, P. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces☆. *Evolution and Human Behavior*, *27*(6), 433–442. https://doi.org/10.1016/j.evolhumbehav.2006.08.007

Fink, B., Klappauf, D., Brewer, G., & Shackelford, T. K. (2014). Female physical characteristics and intra-sexual competition in women. *Personality and Individual Differences*, *58*, 138–141. https://doi.org/10.1016/j.paid.2013.10.015

Fink, B., & Matts, P. (2008). 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*(4), 493–498. https://doi.org/10.1111/j.1468-3083.2007.02512.x

Fink, B., Matts, P. J., Klingenberg, H., Kuntze, S., Weege, B., & Grammer, K. (2008). Visual attention to variation in female facial skin color distribution. *Journal of Cosmetic Dermatology*, *7*(2), 155–161. https://doi.org/10.1111/j.1473-2165.2008.00382.x

Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2005). Facial symmetry and the 'big-five' personality factors. *Personality and Individual Differences*, *39*(3), 523–529. https://doi.org/10.1016/j.paid.2005.02.002

Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2006). Facial symmetry and judgements of attractiveness, health and personality. *Personality and Individual Differences*, *41*(3), 491–499. https://doi.org/10.1016/j.paid.2006.01.017

Fink, B., & Penton-voak, I. (2002). Evolutionary psychology of facial attractiveness. . *Current Directions in Psychological Science*, *11*(5), 154–158.

Fisher, M., & Cox, A. (2011). Four strategies used during intrasexual competition for mates. *Personal Relationships*, *18*(1), 20–38. https://doi.org/10.1111/j.1475-6811.2010.01307.x

Fisher, M. L. (2004). Female intrasexual competition decreases female facial attractiveness. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl_5). https://doi.org/10.1098/rsbl.2004.0160

Flament, F., Bazin, R., Qiu, H., Ye, C., Laquieze, S., Rubert, V., Decroux, A., Simonpietri, E., & Piot, B. (2015). Solar exposure(s) and facial clinical signs of aging in Chinese women: Impacts upon age perception. *Clinical, Cosmetic and Investigational Dermatology*, *8*, 75–84. https://doi.org/10.2147/CCID.S72244

Foos, P. W., & Clark, M. C. (2011). Adult Age and Gender Differences in Perceptions of Facial Attractiveness: Beauty is in the Eye of the Older Beholder. *The Journal of Genetic Psychology*, *172*(2), 162–175. https://doi.org/10.1080/00221325.2010.526154

Foos, P. W., Clark, M. C., & Terrell, D. F. (2006). Adult age, gender, and race group differences in images of aging. *Journal of Genetic Psychology*, *167*(3), 309–325. https://doi.org/10.3200/GNTP.167.3.309-325

Foo, Y. Z., Simmons, L. W., & Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Scientific Reports*, *7*(1), 39731. https://doi.org/10.1038/srep39731

Freckleton, R. P. (2002). On the misuse of residuals in ecology: regression of residuals vs. multiple regression. *Journal of Animal Ecology*, *71*(3), 542–545. https://doi.org/10.1046/j.1365-2656.2002.00618.x

Furnham, A., & Levitas, J. (2012). Factors that motivate people to undergo cosmetic surgery. *Canadian Journal of Plastic Surgery*, *20*(4), 47–50. https://doi.org/10.1177/229255031202000406

Ganel, T. (2015). Smiling makes you look older. *Psychonomic Bulletin & Review*, *22*(6), 1671–1677. https://doi.org/10.3758/s13423-015-0822-7

George, P. A., & Hole, G. J. (1995). Factors Influencing the accuracy of age estimates of unfamiliar faces. *Perception*, *24*(9), 1059–1073. https://doi.org/10.1068/p241059

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. https://doi.org/10.1068/p270295

Goldstein, A. G., & Chance, J. E. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, *23*(3), 211–214. https://doi.org/10.3758/BF03329829

Golle, J., Mast, F. W., & Lobmaier, J. S. (2014). Something to smile about: The interrelationship between attractiveness and emotional expression. *Cognition and Emotion*, *28*(2), 298–310. https://doi.org/10.1080/02699931.2013.817383

Grand Review Research. (n.d.). *Personalized Skin Care Products Market Size, Share & Trends Analysis Report By Product (Face Care, Body Care), By Gender, By Type (Mass, Premium), By Distribution Channel, By Testing, By Region, And Segment Forecasts, 2023 - 2030*. Market Analysis Report. Retrieved May 27, 2023, from https://www.grandviewresearch.com/industry-analysis/personalized-skin-care-products-market-report

Gunn, D. A., de Craen, A. J. M., Dick, J. L., Tomlin, C. C., van Heemst, D., Catt, S. D., Griffiths, T., Ogden, S., Maier, A. B., Murray, P. G., Griffiths, C. E. M., Slagboom, P. E., Westendorp, R. G. J., & Kritchevsky, S. (2013). Facial appearance reflects human familial longevity and cardiovascular disease risk in healthy individuals. *The Journals of Gerontology: Series A*, *68*(2), 145–152. https://doi.org/10.1093/gerona/gls154

Hamilton, W. D., & Zuk, M. (1982). Heritable True Fitness and Bright Birds: A Role for Parasites? *Science*, *218*(4570), 384–387. https://doi.org/10.1126/science.7123238

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45–56. https://doi.org/10.1348/000712607X199981

Harris, M. B. (1994). Growing old gracefully: Age concealment and gender. *Journals of Gerontology*, *49*(4), 149–158. https://doi.org/10.1093/geronj/49.4.P149

Harvey, A. J. (2014). Some effects of alcohol and eye movements on cross-race face learning. *Memory*, *22*(8), 1126–1138. https://doi.org/10.1080/09658211.2013.872278

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, *106*(2), 1017–1027. https://doi.org/10.1016/j.cognition.2007.04.002

Henderson, A. J., Holzleitner, I. J., Talamas, S. N., & Perrett, D. I. (2016). Perception of health from facial cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150380. https://doi.org/10.1098/rstb.2015.0380

Henss, R. (1991). Perceiving age and attractiveness in facial photographs. *Journal of Applied Social Psychology*, *21*(11), 933–946. https://doi.org/10.1111/j.1559-1816.1991.tb00451.x

Hess, U., Adams, R. B., Simard, A., Stevenson, M. T., & Kleck, R. E. (2012). Smiling and sad wrinkles: Age-related changes in the face and the perception of emotions and intentions. *Journal of Experimental Social Psychology*, *48*(6), 1377–1380. https://doi.org/10.1016/j.jesp.2012.05.018

Holland, C. A. C., Ebner, N. C., Lin, T., & Samanez-Larkin, G. R. (2019). Emotion identification across adulthood using the Dynamic FACES database of emotional expressions in younger, middle aged, and older adults. *Cognition and Emotion*, *33*(2), 245–257. https://doi.org/10.1080/02699931.2018.1445981

Horry, R., Cheong, W., & Brewer, N. (2015). The other-race effect in perception and recognition: Insights from the complete composite task. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 508–524. https://doi.org/10.1037/xhp0000042

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion*, *3*(2), 150–166. https://doi.org/10.1037/1528-3542.3.2.150

Hughes, S. M., & Aung, T. (2018). Symmetry in Motion: Perception of Attractiveness Changes with Facial Movement. *Journal of Nonverbal Behavior*, *42*(3), 267–283. https://doi.org/10.1007/s10919-018-0277-4

Hummert, M. L., Garstka, T. A., & Shaner, J. L. (1997). Stereotyping of older adults: The role of target facial cues and perceiver characteristics. *Psychology and Aging*, *12*(1), 107–114. https://doi.org/10.1037/0882-7974.12.1.107

Hurd Clarke, L. C. (2002). Beauty in later life: Older women's perceptions of physical attractiveness. *Canadian Journal on Aging*, *21*(3), 429–442. https://doi.org/10.1017/s0714980800001744

Ida Gobbini, M., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage*, *22*(4), 1628–1635. https://doi.org/10.1016/j.neuroimage.2004.03.049

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review. *Social Psychology Quarterly*, *58*(2), 108. https://doi.org/10.2307/2787149

Jackson, M. C., Counter, P., & Tree, J. J. (2017). Face working memory deficits in developmental prosopagnosia: Tests of encoding limits and updating processes. *Neuropsychologia*, *106*, 60–70. https://doi.org/10.1016/j.neuropsychologia.2017.09.003

JASP Team. (2024). *JASP* (0.18.3). https://jasp-stats.org/

Jones, A., & Jaeger, B. (2019). Biological bases of beauty revisited: The Effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry*, *11*(2), 279. https://doi.org/10.3390/sym11020279

Jones, A. L. (2018). The influence of shape and colour cue classes on facial health perception. *Evolution and Human Behavior*, *39*(1), 19–29. https://doi.org/10.1016/j.evolhumbehav.2017.09.005

Jones, A. L., Batres, C., Porcheron, A., Sweda, J. R., Morizot, F., & Russell, R. (2018). Positive facial affect looks healthy. *Visual Cognition*, *26*(1), 1–12. https://doi.org/10.1080/13506285.2017.1369202

Jones, A. L., & Kramer, R. S. S. (2016). Facial Cosmetics and Attractiveness: Comparing the Effect Sizes of Professionally-Applied Cosmetics and Identity. *PLOS ONE*, *11*(10), e0164218. https://doi.org/10.1371/journal.pone.0164218

Jones, A. L., Porcheron, A., Sweda, J. R., Morizot, F., & Russell, R. (2016). Coloration in different areas of facial skin is a cue to health: The role of cheek redness and periorbital luminance in health perception. *Body Image*, *17*, 57–66. https://doi.org/10.1016/j.bodyim.2016.02.001

Jones, A., Russell, R., & Ward, R. (2015). Cosmetics Alter Biologically-Based Factors of Beauty: Evidence from Facial Contrast. *Evolutionary Psychology*, *13*(1), 147470491501300. https://doi.org/10.1177/147470491501300113

Jones, B. C., Little, A. C., Penton-Voak, I. S., Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Facial symmetry and judgements of apparent health. *Evolution and Human Behavior*, *22*(6), 417–429. https://doi.org/10.1016/S1090-5138(01)00083-6

Judge, T. A., Hurst, C., & Simon, L. S. (2009). Does it pay to be smart, attractive, or confident (or all three)? Relationships among general mental ability, physical attractiveness, core self-evaluations, and income. *Journal of Applied Psychology*, *94*(3), 742–755. https://doi.org/10.1037/a0015497

Juhász, M. L. W., Levin, M. K., & Marmur, E. S. (2017). A review of available laser and intense light source home devices: A dermatologist's perspective. *Journal of Cosmetic Dermatology*, *16*(4), 438–443. https://doi.org/10.1111/jocd.12371

Kalick, S. M. (1979). Aesthetic surgery: how it affects the way patients are perceived by others. *Annals of Plastic Surgery*, *2*(2), 128–134. https://europepmc.org/article/med/426454

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., & Johnson, R. M. (1998). Does Human Facial Attractiveness Honestly Advertise Health? Longitudinal Data on an Evolutionary Question. *Psychological Science*, *9*(1), 8–13. https://doi.org/10.1111/1467-9280.00002

Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, *108*, 67–76. https://doi.org/10.1016/j.visres.2015.01.013

Kellie, D. J., Blake, K. R., & Brooks, R. C. (2021). Behind the makeup: The effects of cosmetics on women's self-objectification, and their objectification by others. *European Journal of Social Psychology*, *51*(4–5), 703–721. https://doi.org/10.1002/ejsp.2767

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kho, S. K., Leong, B. Q. Z., Keeble, D. R. T., Wong, H. K., & Estudillo, A. J. (2023). A new Asian version of the CFMT: The Cambridge Face Memory Test – Chinese Malaysian (CFMT-MY). *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02085-6

Kite, M. E., Stockdale, G. D., Whitley, B. E., & Johnson, B. T. (2005). Attitudes Toward Younger and Older Adults: An Updated Meta-Analytic Review. *Journal of Social Issues*, *61*(2), 241–266. https://doi.org/10.1111/j.1540-4560.2005.00404.x

Kleisner, K., Kočnar, T., Tureček, P., Stella, D., Akoko, R. M., Třebický, V., & Havlíček, J. (2017). African and European perception of African female attractiveness. *Evolution and Human Behavior*, *38*(6), 744–755. https://doi.org/10.1016/j.evolhumbehav.2017.07.002

Korthase, K. M., & Trenholme, I. (1982). Perceived Age and Perceived Physical Attractiveness. *Perceptual and Motor Skills*, *54*(3_suppl), 1251–1258. https://doi.org/10.2466/pms.1982.54.3c.1251

Kościński, K. (2013). Perception of Facial Attractiveness from Static and Dynamic Stimuli. *Perception*, *42*(2), 163–175. https://doi.org/10.1068/p7378

Kruschke, J. (2015). *Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Kumar, S., Massie, C., & Dumonceaux, M. D. (2006). Comparative innovative business strategies of major players in cosmetic industry. *Industrial Management & Data Systems*, *106*(3), 285–306. https://doi.org/10.1108/02635570610653461

Kwart, D. G., Foulsham, T., & Kingstone, A. (2012). Age and Beauty are in the Eye of the Beholder. *Perception*, *41*(8), 925–938. https://doi.org/10.1068/p7136

Lander, K. (2008). Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour*, *75*(3), 817–822. https://doi.org/10.1016/j.anbehav.2007.07.001

Lander, K., & Butcher, N. (2015). Independence of face identity and expression processing: exploring the role of motion. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00255

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390–423. https://doi.org/10.1037/0033-2909.126.3.390

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., Boothroyd, L. G., Durrani, S. J., Stirrat, M. R., Whiten, S., Pitman, R. M., & Hillier, S. G. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1583), 135–140. https://doi.org/10.1098/rspb.2005.3296

Leary, M. R. (1999). Making sense of self-esteem. *Current Directions in Psychological Science*, *8*(1), 32–35.

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual Other-Race Training Reduces Implicit Racial Bias. *PLoS ONE*, *4*(1), e4215. https://doi.org/10.1371/journal.pone.0004215

Lephart, E. D. (2018). A review of the role of estrogen in dermal aging and facial attractiveness in women. *Journal of Cosmetic Dermatology*, *17*(3), 282–288. https://doi.org/10.1111/jocd.12508

Lephart, E. D., & Naftolin, F. (2021). Menopause and the Skin: Old Favorites and New Innovations in Cosmeceuticals for Estrogen-Deficient Skin. *Dermatology and Therapy*, *11*(1), 53–69. https://doi.org/10.1007/s13555-020-00468-7

Lick, D. J., & Johnson, K. L. (2018). Facial Cues to Race and Gender Interactively Guide Age Judgments. *Social Cognition*, *36*(5), 497–516. https://doi.org/10.1521/soco.2018.36.5.497

Lie, H. C., Simmons, L. W., & Rhodes, G. (2009). Does Genetic Diversity Predict Health in Humans? *PLoS ONE*, *4*(7), e6391. https://doi.org/10.1371/journal.pone.0006391

Lindeberg, S., Craig, B. M., & Lipp, O. V. (2019). You look pretty happy: Attractiveness moderates emotion perception. *Emotion*, *19*(6), 1070–1080. https://doi.org/10.1037/emo0000513

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1638–1659. https://doi.org/10.1098/rstb.2010.0404

Longmore, C. A., & Tree, J. J. (2013). Motion as a cue to face recognition: Evidence from congenital prosopagnosia. *Neuropsychologia*, *51*(5), 864–875. https://doi.org/10.1016/j.neuropsychologia.2013.01.022

Lopaciuk, A., & Loboda, M. (2013). Global beauty trends in the 21st century. *Management, Knowledge and Learning International Conference*, 19–21.

MacLin, O. H., Van Sickler, B. R., MacLin, M. K., & Li, A. (2004). A re-examination of the cross-race effect: The role of race, inversion, and basketball trivia. *North American Journal of Psychology*, *6*(2), 189–204. https://www.researchgate.net/profile/M-Maclin/publication/232551758_A_Re-examination_of_the_Cross-race_Effect_The_Role_of_Race_Inversion_and_Basketball_Trivia/links/00b7d52e2a1afe3e14000000/A-Re-examination-of-the-Cross-race-Effect-The-Role-of-Race-Inversion-and-Basketball-Trivia.pdf?_sg%5B0%5D=started_experiment_milestone&origin=journalDetail&_rtd=e30%3D

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Maestripieri, D., Klimczuk, A. C. E., Traficonte, D. M., & Wilson, M. C. (2014). A greater decline in female facial attractiveness during middle age reflects women’s loss of reproductive value. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00179

Mafra, A. L., Varella, M. A. C., Defelipe, R. P., Anchieta, N. M., de Almeida, C. A. G., & Valentova, J. V. (2020). Makeup usage in women as a tactic to attract mates and compete with rivals. *Personality and Individual Differences*, *163*, 110042. https://doi.org/10.1016/j.paid.2020.110042

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02767

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334. https://doi.org/10.1037/h0028434

Mangal, S., Ray, M., Saha, S., Paul, S., & Maji, H. (2021). Perception regarding selection and use of skincare products and the impact of advertisement: a questionnaire-based randomized survey. *Egyptian Pharmaceutical Journal*, *20*(3), 173. https://doi.org/10.4103/epj.epj_20_21

Martínez-González, M. C., Martínez-González, R.-A., & Guerra-Tapia, A. (2018). Esthetic dermatology and emotional well-being according to gender. *Journal of Cosmetic Dermatology*, *17*(3), 410–416. https://doi.org/10.1111/jocd.12389

Matts, P. J., Fink, B., Grammer, K., & Burquest, M. (2007). Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *Journal of the American Academy of Dermatology*, *57*(6), 977–984. https://doi.org/10.1016/j.jaad.2007.07.040

Mayes, A., Murray, P., Gunn, D., Tomlin, C., Catt, S., Wen, Y., Zhou, L., Wang, H., Catt, M., & Granger, S. (2010). Ageing appearance in China: biophysical profile of facial skin and its relationship to perceived age. *Journal of the European Academy of Dermatology and Venereology*, *24*(3), 341–348. https://doi.org/10.1111/j.1468-3083.2009.03418.x

McElreath, R. (2020). *Statistical Rethinking*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

McFarland, M. B. (1999). *A descriptive study of body dysphoria and body image in mid -life women* [Dissertation]. The University of North Dakota.

McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., Yovel, G., Davis, J. M., & O'Connor, K. B. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test–Australian. *Cognitive Neuropsychology, 28*(2), 109–146. https://doi.org/10.1080/02643294.2011.616880

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., Yovel, G., Broughton, M., & Pelleg, M. (2012). A Robust Method of Measuring Other-Race and Other-Ethnicity Effects: The Cambridge Face Memory Test Format. *PLoS ONE*, *7*(10), e47956. https://doi.org/10.1371/journal.pone.0047956

McLellan, B., & McKelvie, S. J. (1993). Effects of age and gender on perceived facial attractiveness. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *25*(1), 135–142. https://doi.org/10.1037/h0078790

Mendelson, B., & Wong, C.-H. (2012). Changes in the Facial Skeleton With Aging: Implications and Clinical Applications in Facial Rejuvenation. *Aesthetic Plastic Surgery*, *36*(4), 753–760. https://doi.org/10.1007/s00266-012-9904-3

Merinville, E., Grennan, G. Z., Gillbro, J. M., Mathieu, J., & Mavon, A. (2015). Influence of facial skin ageing characteristics on the perceived age in a Russian female population. *International Journal of Cosmetic Science*, *37*, 3–8. https://doi.org/10.1111/ics.12252

Merinville, E., Messaraa, C., O'Connor, C., Grennan, G., & Mavon, A. (2018). What Makes Indian Women Look Older—An Exploratory Study on Facial Skin Features. *Cosmetics*, *5*(1), 3. https://doi.org/10.3390/cosmetics5010003

Messaraa, C., Robertson, N., Walsh, M., Hurley, S., Doyle, L., Mansfield, A., Daly, L., Tansey, C., & Mavon, A. (2020). Clinical evidences of benefits from an advanced skin care routine in comparison with a simple routine. *Journal of Cosmetic Dermatology*, *19*(8), 1993–1999. https://doi.org/10.1111/jocd.13252

Michaud, T., Gassia, V., & Belhaouari, L. (2015). Facial dynamics and emotional expressions in facial aging treatments. *Journal of Cosmetic Dermatology*, *14*(1), 9–21. https://doi.org/10.1111/jocd.12128

Mileva, V. R., Jones, A. L., Russell, R., & Little, A. C. (2016). Sex Differences in the Perceived Dominance and Prestige of Women With and Without Cosmetics. *Perception*, *45*(10), 1166–1183. https://doi.org/10.1177/0301006616652053

Mobius, M. M., & Rosenblat, T. S. (2006). Why Beauty Matters. *American Economic Review*, *96*(1), 222–235. https://doi.org/10.1257/000282806776157515

Morey, R. (2024, January 23). *Using the "BayesFactor" package, version 0.9.2+*. https://cran.r-project.org/web/packages/BayesFactor/vignettes/manual.html#glm

Morrison, E. R., Gralewski, L., Campbell, N., & Penton-Voak, I. S. (2007). Facial movement varies by sex and is related to attractiveness. *Evolution and Human Behavior*, *28*(3), 186–192. https://doi.org/10.1016/j.evolhumbehav.2007.01.001

Morrison, E. R., Morris, P. H., & Bard, K. A. (2013). The Stability of Facial Attractiveness: Is It What You've Got or What You Do with It? *Journal of Nonverbal Behavior*, *37*(2), 59–67. https://doi.org/10.1007/s10919-013-0145-1

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

Muise, A., & Desmarais, S. (2010a). Women's Perceptions and Use of "Anti-Aging" Products. *Sex Roles*, *63*(1–2), 126–137. https://doi.org/10.1007/s11199-010-9791-5

Nash, R., Fieldman, G., Hussey, T., Lévêque, J.-L., & Pineau, P. (2006). Cosmetics: They Influence More Than Caucasian Female Facial Attractiveness. *Journal of Applied Social Psychology*, *36*(2), 493–504. https://doi.org/10.1111/j.0021-9029.2006.00016.x

Nellis, J. C., Ishii, M., Papel, I. D., Kontis, T. C., Byrne, P. J., Boahene, K. D. O., Bater, K. L., & Ishii, L. E. (2017). Association of Face-lift Surgery With Social Perception, Age, Attractiveness, Health, and Success. *JAMA Facial Plastic Surgery*, *19*(4), 311–317. https://doi.org/10.1001/jamafacial.2016.2206

Ng, T. W. H., & Feldman, D. C. (2012). Employee voice behavior: A meta-analytic test of the conservation of resources framework. *Journal of Organizational Behavior*, *33*(2), 216–234. https://doi.org/10.1002/job.754

Ng, W.-J., & Lindsay, R. C. L. (1994). Cross-Race Facial Recognition. *Journal of Cross-Cultural Psychology*, *25*(2), 217–232. https://doi.org/10.1177/0022022194252004

Nkengne, A., Bertin, C., Stamatas, G., Giron, A., Rossi, A., Issachar, N., & Fertil, B. (2008). Influence of facial skin attributes on the perceived age of Caucasian women. *Journal of the European Academy of Dermatology and Venereology*, *22*(8), 982–991. https://doi.org/10.1111/j.1468-3083.2008.02698.x

North, M. S., & Fiske, S. T. (2013). Act Your (Old) Age. *Personality and Social Psychology Bulletin*, *39*(6), 720–734. https://doi.org/10.1177/0146167213480043

Office for National Statistics. (2020). *Conceptions in England and Wales: 2018*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/conceptionstatistics/2018

Office for National Statistics. (2023, January 19). *Birth characteristics in England and Wales: 2020*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthcharacteristicsinenglandandwales/2021

Parsa, K. M., Charipova, K., Coerdt, K., Clark, C. M., Wang, H., Chu, E., & Reilly, M. J. (2021). The Role of Age and Gender on Perception of Women After Cosmetic Rhinoplasty. *Aesthetic Plastic Surgery*, *45*(3), 1184–1190. https://doi.org/10.1007/s00266-020-02030-4

Penton-Voak, I. S., & Chang, H. Y. (2008). Attractiveness judgements of individuals vary across emotional expression and movement conditions. *Journal of Evolutionary Psychology*, *6*(2), 89–100. https://doi.org/10.1556/JEP.2008.1011

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and Human Facial Attractiveness. *Evolution and Human Behavior*, *20*(5), 295–307. https://doi.org/10.1016/S1090-5138(99)00014-8

Perry, E. L., & Finkelstein, L. M. (1999). Toward a broader view of age discrimination in employment-related decisions: A joint consideration of organizational factors and cognitive processes. *Human Resource Management Review*, *9*(1), 21–49. https://doi.org/https://doi.org/10.1016/S1053-4822(99)00010-8

Porcheron, A., Mauger, E., & Russell, R. (2013). Aspects of Facial Contrast Decrease with Age and Are Cues for Age Perception. *PLoS ONE*, *8*(3), e57985. https://doi.org/10.1371/journal.pone.0057985

Potter, T., & Corneille, O. (2008). Locating attractiveness in the face space: Faces are more attractive when closer to their group prototype. *Psychonomic Bulletin & Review*, *15*(3), 615–622. https://doi.org/10.3758/PBR.15.3.615

Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology*, *110*(3), 461–479. https://doi.org/10.1111/bjop.12368

Ravina, E. (2008). *Love & Loans: The effect of beauty and personal characteristics in credit markets*. http://users.nber.org/~confer/2008/bfs08/ravina.pdf

Rawlings, A. V. (2006). Ethnic skin types: are there differences in skin structure and function?1. *International Journal of Cosmetic Science*, *28*(2), 79–93. https://doi.org/10.1111/j.1467-2494.2006.00302.x

R Core Team. (2013). The R Project for Statistical Computing. In *Vienna, Austria*. https://www.R-project.org

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* http://www.R-project.org/

Rhodes, G. (2006). The Evolutionary Psychology of Facial Beauty. *Annual Review of Psychology*, *57*(1), 199–226. https://doi.org/10.1146/annurev.psych.57.102904.190208

Rhodes, G., Ewing, L., Hayward, W. G., Maurer, D., Mondloch, C. J., & Tanaka, J. W. (2009). Contact and other-race effects in configural and component processing of faces. *British Journal of Psychology*, *100*(4), 717–728. https://doi.org/10.1348/000712608X396503

Rhodes, G., Lee, K., Palermo, R., Weiss, M., Yoshikawa, S., Clissa, P., Williams, T., Peters, M., Winkler, C., & Jeffery, L. (2005). Attractiveness of Own-Race, Other-Race, and Mixed-Race Faces. *Perception*, *34*(3), 319–340. https://doi.org/10.1068/p5191

Rhodes, G., Lie, H. C., Thevaraja, N., Taylor, L., Iredell, N., Curran, C., Tan, S. Q. C., Carnemolla, P., & Simmons, L. W. (2011). Facial Attractiveness Ratings from Video-Clips and Static Images Tell the Same Story. *PLoS ONE*, *6*(11), e26653. https://doi.org/10.1371/journal.pone.0026653

Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., & Akamatsu, S. (2001). Attractiveness of Facial Averageness and Symmetry in Non-Western Cultures: In Search of Biologically Based Standards of Beauty. *Perception*, *30*(5), 611–625. https://doi.org/10.1068/p3123

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146–174. https://doi.org/10.1037/a0025750

Robertson, D. J., Black, J., Chamberlain, B., Megreya, A. M., & Davis, J. P. (2020). Super-Recognisers show an advantage for other race face identification. *Applied Cognitive Psychology*, *34*(1), 205–216. https://doi.org/10.1002/acp.3608

Roberts, S. C., Saxton, T. K., Murray, A. K., Burriss, R. P., Rowland, H. M., & Little, A. C. (2009). Static and Dynamic Facial Images Cue Similar Attractiveness Judgements. *Ethology*, *115*(6), 588–595. https://doi.org/10.1111/j.1439-0310.2009.01640.x

Rooth, D.-O. (2009). Obesity, Attractiveness, and Differential Treatment in Hiring. *Journal of Human Resources*, *44*(3), 710–735. https://doi.org/10.3368/jhr.44.3.710

Ross, D. A., Richler, J. J., & Gauthier, I. (2015). Reliability of composite-task measurements of holistic face processing. *Behavior Research Methods*, *47*(3), 736–743. https://doi.org/10.3758/s13428-014-0497-4

Rouw, R., & Gelder, B. De. (2002). Impaired face recognition does not preclude intact whole face perception. *Visual Cognition*, *9*(6), 689–718. https://doi.org/10.1080/13506280143000223

Rubenstein, A. J. (2005). Variation in Perceived Attractiveness. *Psychological Science*, *16*(10), 759–762. https://doi.org/10.1111/j.1467-9280.2005.01610.x

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Russel, V. L., Buerkner, P. , Herve, M., Love, J., Riebl, H., & Singman, H. (2017). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/web/packages/emmeans/index.html

Salihu, H. (2003). Childbearing beyond maternal age 50 and fetal outcomes in the United States. *Obstetrics & Gynecology*, *102*(5), 1006–1014. https://doi.org/10.1016/S0029-7844(03)00739-7

Samson, N., Fink, B., & Matts, P. (2011). Interaction of skin color distribution and skin surface topography cues in the perception of female facial age and health. *Journal of Cosmetic Dermatology*, *10*(1), 78–84. https://doi.org/10.1111/j.1473-2165.2010.00538.x

Samson, N., Fink, B., Matts, P. J., Dawes, N. C., & Weitz, S. (2010). Visible changes of female facial skin surface topography in relation to age and attractiveness perception. *Journal of Cosmetic Dermatology*, *9*(2), 79–88. https://doi.org/10.1111/j.1473-2165.2010.00489.x

Sangrigoli, S., & de Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry*, *45*(7), 1219–1227. https://doi.org/10.1111/j.1469-7610.2004.00319.x

Schmitt, D. P., & Buss, D. M. (1996). Strategic self-promotion and competitor derogation: Sex and context effects on the perceived effectiveness of mate attraction tactics. *Journal of Personality and Social Psychology*, *70*(6), 1185–1204. https://doi.org/10.1037/0022-3514.70.6.1185

Schoemann, A. M., & Branscombe, N. R. (2011). Looking young for your age: Perceptions of anti-aging actions. *European Journal of Social Psychology*, *41*(1), 86–95. https://doi.org/10.1002/ejsp.738

Sefcek, J. A., Brumbach, B. H., Vasquez, G., & Miller, G. F. (2007). The Evolutionary Psychology of Human Mate Choice. *Journal of Psychology & Human Sexuality*, *18*(2–3), 125–182. https://doi.org/10.1300/J056v18n02_05

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences*, *112*(41), 12887–12892. https://doi.org/10.1073/pnas.1421881112

Simion, F., & Giorgio, E. Di. (2015). Face perception and processing in early infancy: inborn predispositions and developmental changes. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00969

Slevec, J., & Tiggemann, M. (2010). Attitudes Toward Cosmetic Surgery in Middle-Aged Women: Body Image, Aging Anxiety, and the Media. *Psychology of Women Quarterly*, *34*(1), 65–74. https://doi.org/10.1111/j.1471-6402.2009.01542.x

Slevin, K. F. (2010). "If I had lots of money... I'd have a body makeover:" Managing the Aging Body. *Social Forces*, *88*(3), 1003–1020. https://doi.org/10.1353/sof.0.0302

Sng, O., Williams, K. E. G., & Neuberg, S. L. (2020). Sex-age stereotyping: Social perceivers as lay adaptationists. *Evolution and Human Behavior*, *41*(2), 136–149. https://doi.org/10.1016/j.evolhumbehav.2019.12.001

Sontag, S. (1979). The double standard of aging. In J. Williams (Ed.), *Psychology of women* (pp. 462–478). Academic Press.

Stephen, I. D., & McKeegan, A. M. (2010). Lip Colour Affects Perceived Sex Typicality and Attractiveness of Human Faces. *Perception*, *39*(8), 1104–1110. https://doi.org/10.1068/p6730

Stoehr, A. M., & Kokko, H. (2006). Sexual dimorphism in immunocompetence: what does life-history theory predict? *Behavioral Ecology*, *17*(5), 751–756. https://doi.org/10.1093/beheco/ark018

Sulikowski, D., Ensor, M., & Wagstaff, D. (2022). Mate-value moderates the function of make-up as a signal of intrasexual aggression. *Personality and Individual Differences*, *185*, 111275. https://doi.org/10.1016/j.paid.2021.111275

Sunday, M. A., Richler, J. J., & Gauthier, I. (2017). Limited evidence of individual differences in holistic processing in different versions of the part-whole paradigm. *Attention, Perception, & Psychophysics*, *79*(5), 1453–1465. https://doi.org/10.3758/s13414-017-1311-z

Swami, V., Pietschnig, J., Stewart, N., Nader, I. W., Stieger, S., Shannon, S., & Voracek, M. (2013). Blame it on patriarchy: More sexist attitudes are associated with stronger consideration of cosmetic surgery for oneself and one's partner. *International Journal of Psychology*, *48*(6), 1221–1229. https://doi.org/10.1080/00207594.2012.740566

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study. *Cognition*, *93*(1), B1–B9. https://doi.org/10.1016/j.cognition.2003.09.011

Teuscher, U., & Teuscher, C. (2007). Reconsidering the double standard of aging: Effects of gender and sexual orientation on facial attractiveness ratings. *Personality and Individual Differences*, *42*(4), 631–639. https://doi.org/10.1016/j.paid.2006.08.020

The Aesthetic Society. (2021). *Aesthetic Plastic Surgery National Databank Statistics 2020-2021*. https://cdn.theaestheticsociety.org/media/statistics/2021-TheAestheticSocietyStatistics.pdf

The American Society for Aesthetic Plastic Surgery. (2018). *Cosmetic (Aesthetic) Surgery National Data Bank*. https://cdn.theaestheticsociety.org/media/statistics/2018-TheAestheticSocietyStatistics.pdf

The Body Shop. (2023). *Rise Up With Self Love*. https://www.thebodyshop.com/en-gb/about-us/activism/self-love/a/a00038

Thornhill, R., & Gangestad, S. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*(12), 452–460.

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior*, *27*(2), 131–144. https://doi.org/10.1016/j.evolhumbehav.2005.06.001

Thornton, M. J. (2013). Estrogens and aging skin. *Dermato-Endocrinology*, *5*(2), 264–270. https://doi.org/10.4161/derm.23872

Tian, L., Bashir, N. Y., Chasteen, A. L., & Rule, N. O. (2020). The Effect of Age-Stigma Concealment on Social Evaluations. *Basic and Applied Social Psychology*, *42*(4), 219–234. https://doi.org/10.1080/01973533.2020.1741359

Tinlin, R. M., Watkins, C. D., Welling, L. L. M., DeBruine, L. M., Al-Dujaili, E. A. S., & Jones, B. C. (2013). Perceived facial adiposity conveys information about women's health. *British Journal of Psychology*, *104*(2), 235–248. https://doi.org/10.1111/j.2044-8295.2012.02117.x

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of Competence from Faces Predict Election Outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Trujillo, L. T., Jankowitsch, J. M., & Langlois, J. H. (2014). Beauty is in the ease of the beholding: A neurophysiological test of the averageness theory of facial attractiveness. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(3), 1061–1076. https://doi.org/10.3758/s13415-013-0230-2

Tybur, J. M., Fan, L., Jones, B. C., Holzleitner, I. J., Lee, A. J., & DeBruine, L. M. (2022). Re-evaluating the relationship between pathogen avoidance and preferences for facial symmetry and sexual dimorphism: A registered report. *Evolution and Human Behavior*, *43*(3), 212–223. https://doi.org/10.1016/j.evolhumbehav.2022.01.003

Ueda, R., Kuraguchi, K., & Ashida, H. (2016). Asymmetric Effect of Expression Intensity on Evaluations of Facial Attractiveness. *SAGE Open*, *6*(4), 215824401667756. https://doi.org/10.1177/2158244016677569

Unicef. (n.d.). *Dove Working with Unicef to improve body confidence and self-esteem*. Retrieved July 31, 2023, from https://www.unicef.org.uk/corporate-partners/dove/

Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

Verosky, S. C., Porter, J., Martinez, J. E., & Todorov, A. (2018). Robust effects of affective person learning on evaluation of faces. *Journal of Personality and Social Psychology*, *114*(4), 516–528. https://doi.org/10.1037/pspa0000109

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

Voegeli, R., Schoop, R., Prestat-Marquis, E., Rawlings, A. V, Shackelford, T. K., & Fink, B. (2021). Cross-cultural perception of female facial appearance: A multi-ethnic and multi-centre study. *PLOS ONE*, *16*(1), e0245998. https://doi.org/10.1371/journal.pone.0245998

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*, *27*(2), 265–277. https://doi.org/10.1037/a0025065

Wagenmakers, E.-J., & Ly, A. (2023). History and nature of the Jeffreys–Lindley paradox. *Archive for History of Exact Sciences*, *77*(1), 25–72. https://doi.org/10.1007/s00407-022-00298-3

Wagstaff, D. L. (2018). Comparing mating motivations, social processes, and personality as predictors of women's cosmetics use. *Evolutionary Behavioral Sciences*, *12*(4), 367–380. https://doi.org/10.1037/ebs0000119

Wagstaff, D. L., & Sulikowski, D. (2022). The impact of sexual strategies, social comparison, and Instagram use on makeup purchasing intentions. *Evolutionary Behavioral Sciences*. https://doi.org/10.1037/ebs0000285

Walker, P. M., & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology*, *20*(4), 461–475. https://doi.org/10.1002/acp.1191

Wang, X., Chen, H., Chen, Z., & Yang, Y. (2021). Women's Intrasexual Competition Results in Beautification. *Social Psychological and Personality Science*, *12*(5), 648–657. https://doi.org/10.1177/1948550620933403

Wan, L., Crookes, K., Dawel, A., Pidcock, M., Hall, A., & McKone, E. (2017). Face-blind for other-race faces: Individual differences in other-race recognition impairments. *Journal of Experimental Psychology: General*, *146*(1), 102–122. https://doi.org/10.1037/xge0000249

Williams, J. R. (2015). Ethics and Medical Research. In *Medical Ethics Manual*.

Willis, J., & Todorov, A. (2006). First Impressions. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilmer, J. B. (2017). Individual Differences in Face Recognition: A Decade of Discovery. *Current Directions in Psychological Science*, *26*(3), 225–230. https://doi.org/10.1177/0963721417710693

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology*, *17*(3), 365–373. https://doi.org/10.1002/acp.898

Wyckoff, J. P., Asao, K., & Buss, D. M. (2019). Gossip as an intrasexual competition strategy: Predicting information sharing from potential mate versus competitor mating strategies. *Evolution and Human Behavior*, *40*(1), 96–104. https://doi.org/10.1016/j.evolhumbehav.2018.08.006

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00633

Zebrowitz, L. A. (2011). *Ecological and Social Approaches to Face Perception*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.013.0003

Zhang, L., Adique, A., Sarkar, P., Shenai, V., Sampath, M., Lai, R., Qi, J., Wang, M., & Farage, M. A. (2020). The Impact of Routine Skin Care on the Quality of Life. *Cosmetics*, *7*(3), 59. https://doi.org/10.3390/cosmetics7030059

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research*, *105*, 61–69. https://doi.org/10.1016/j.visres.2014.09.006

Zhou, X., Elshiekh, A., & Moulson, M. C. (2019). Lifetime perceptual experience shapes face memory for own- and other-race faces. *Visual Cognition*, *27*(9–10), 687–700. https://doi.org/10.1080/13506285.2019.1638478

Chapter 7.

Appendix

Appendix A. Motivations behind age concealment (Harris, 1994)

How acceptable would it be to appear younger than you are for the following reasons?

1. Self-esteem
   1 2 3 4 5 6 7
   Not at all acceptable Extremely acceptable

2. Employment
   1 2 3 4 5 6 7
   Not at all acceptable Extremely acceptable

3. Attracting/pleasing a partner
   1 2 3 4 5 6 7
   Not at all acceptable Extremely acceptable

4. Pleasing others
   1 2 3 4 5 6 7
   Not at all acceptable Extremely acceptable

5. Vanity
   1 2 3 4 5 6 7
   Not at all acceptable Extremely acceptable

Appendix B. Signs of aging (Harris, 1994)

Please evaluate the following signs of aging

1. Gray hair
   1 2 3 4 5 6 7
   Very unattractive Very attractive

2. White hair
   1 2 3 4 5 6 7
   Very unattractive Very attractive

3. Balding
   1 2 3 4 5 6 7
   Very unattractive Very attractive

4. Facial wrinkles
   1 2 3 4 5 6 7
   Very unattractive Very attractive

5. Facial sagging*
   1 2 3 4 5 6 7
   Very unattractive Very attractive

6. Uneven facial skin tone*
   1 2 3 4 5 6 7
   Very unattractive Very attractive

7. Wrinkled neck
   1 2 3 4 5 6 7
   Very unattractive Very attractive

8. Change in body shape
   1 2 3 4 5 6 7
   Very unattractive Very attractive

*added items

Appendix C. Desired procedures – Harris, 1994

If money was not an issue, would you do the following age concealment techniques:

1. Colour or highlight hair to cover gray
   0 1 2 3
   Would never do Definitely would do

2. Use a wig to cover thinning or balding hair
   0 1 2 3
   Would never do Definitely would do

3. Use hair transplant or chemical to increase hair
   0 1 2 3
   Would never do Definitely would do

4. Use cosmetics to cover wrinkles
   0 1 2 3
   Would never do Definitely would do

5. Use hand-held products at home to reduce wrinkles, e.g. radiofrequency and light therapy
   0 1 2 3
   Would never do Definitely would do

6. Use chemical peels, microdermabrasion to look younger
   0 1 2 3
   Would never do Definitely would do

7. Use Botox or other injectables to look younger
   0 1 2 3
   Would never do Definitely would do

8. Get face-lift or other cosmetic surgeries to look younger
   0 1 2 3
   Would never do Definitely would do

Appendix D. Anxiety towards Aging (Lasher & Faulkander, 1993)

With regards to your physical appearance, please indicate whether you agree to the statements below:

	Definitely Disagree	Somewhat Disagree	Neither agree nor disagree	Somewhat Agree	Definitely Agree
1. I have never lied about my age in order to appear younger.	1	2	3	4	5
2. It doesn't bother me at all to imagine myself as being old.	1	2	3	4	5
3. I have never dreaded the day I would look in the mirror and see gray hairs.	1	2	3	4	5
4. I have never dreaded looking old.	1	2	3	4	5
5. When I look in the mirror, it bothers me to see how my looks have changed with age.	1	2	3	4	5

Appendix E. Aesthetic Dermatology scales – Martinez-Gonzales et al, 2017

Please indicate whether you agree or disagree with the following statements:

	Total disagreement	Disagreement	Agreement	Total Agreement
The results of aesthetic dermatology can help me:				
1. . . .do my work and tasks in better spirits	1	2	3	4
2. . . .have desire to live	1	2	3	4
3. . . . have a better mood	1	2	3	4
4. . . .enjoy going out and see people	1	2	3	4
5. . . .feel more hopeful	1	2	3	4
6. . . .see the positive side of things	1	2	3	4
7. . . .feel more able to overcome mistakes and weaknesses	1	2	3	4
8. . . .improve my relationship with my partner	1	2	3	4
9. . . .open many doors in my life	1	2	3	4
10. . . .improve my social relationships (friends and family)	1	2	3	4
11. . . .enhance my success at work	1	2	3	4
12. . . .feel in better health	1	2	3	4
13. . . .feel better with myself and more self-confident.	1	2	3	4
14. I know what the injectable wrinkle fillers consist on	1	2	3	4
15. I know what “botox” or “botulinum toxin” treatment consists on	1	2	3	4
16. I know what laser rejuvenation for the skin consists on	1	2	3	4
17. Sometimes I have thought of undergoing some aesthetic treatment	1	2	3	4
18. Aesthetic Dermatology treatments not only corrects aesthetic defects, but they also help prevent	1	2	3	4

ageing and preserve skin functions				
19. I would undergo some aesthetic treatment if advised by my dermatologist.	1	2	3	4

Appendix F. Intrasexual Competition Scale – Buunk and Fisher, 2009

I can’t stand it when I meet another man/woman who is more attractive than I am.1,2
1 2 3 4 5 6 7
Not at all applicable Completely applicable

When I go out, I can’t stand it when women/men pay more attention to a same-sex friend of mine than to me. 3,4
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I tend to look for negative characteristics in attractive men/women.2
1 2 3 4 5 6 7
Not at all applicable Completely applicable

When I’m at a party, I enjoy it when women/men pay more attention to me than other men/women.6
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I wouldn’t hire a very attractive man/woman as a colleague.
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I just don’t like very ambitious men/women.5
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I tend to look for negative characteristics in men/women who are very successful.
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I wouldn’t hire a highly competent man/woman as a colleague.
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I like to be funnier and more quick witted than other men/women.5,6
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I want to be just a little better than other men/women.
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I always want to beat other men/women.5,6
1 2 3 4 5 6 7
Not at all applicable Completely applicable

I don’t like seeing other men/women with a nicer house or a nicer car than mine.

1 2 3 4 5 6 7
Not at all applicable Completely applicable

Answers vary from "not at all applicable" (1) to "completely applicable" (7)
1 Women score significantly higher than men in Dutch sample
2 Women score significantly higher than men in Canadian sample
3 Women score marginally significantly (p = .058) higher than men in Dutch sample
4 Women score marginally significantly (p = .055) higher than men in Canadian sample
5 Men score significantly higher than women in Dutch sample
6 Men score significantly higher than women in Canadian sample

Appendix G. Vignettes and evaluations of targets following Chasteen et al. (2011)

- Motivation – romantic partner; concealment type – moderate
  - Angela is a middle-aged woman who wants to maintain a more youthful appearance to look for a *romantic partner*. She regularly uses *non-invasive, hand-held device* that she could use at home as part of her anti-aging routine.
- Motivation – romantic partner; concealment type – major
  - Barbara is a middle-aged woman who wants to maintain a more youthful appearance to look for a *romantic partner*. She regularly uses *professional treatments such as Botox and dermal fillers* as part of her routine.
- Motivation – employment; concealment type – moderate
  - Carol is a middle-aged woman who wants to maintain a more youthful appearance to find a *better job*. She regularly uses *non-invasive, hand-held device* that she could use at home as part of her anti-aging routine.
- Motivation – employment; concealment type – major
  - Dorothy is a middle-aged woman who wants to maintain a more youthful appearance to find a *better job*. She regularly uses *professional treatments such as Botox and dermal fillers* as part of her routine.
- Motivation – self-esteem; concealment type – moderate
  - Elaine is a middle-aged woman who wants to maintain a more youthful appearance to increase her *self-esteem*. She regularly uses *non-invasive, hand-held device* that she could use at home as part of her anti-aging routine.
- Motivation – self-esteem; concealment type – major
  - Fatima is a middle-aged woman who wants to maintain a more youthful appearance to increase her self-esteem. She regularly uses *professional treatments such as Botox and dermal fillers* as part of her routine.

Appendix H. Coefficients of fixed effects and their interactions.

Parameters	*b*	SE	df	*t*	*p*
(Intercept)	51.572	1.026	643.93	50.28	**0.001**
Age	1.974	1.023	643.93	1.93	0.054
Gender	-4.47	2.1	643.93	-2.128	**0.034**
Motivation (Romantic)	-3.266	0.925	1480	-3.529	**0.001**
Motivation (Self-esteem)	3.417	0.925	1480	3.693	**0.001**
Concealment	8.923	0.925	1480	9.642	**0.001**
ICS	-4.327	1.044	643.93	-4.146	**0.001**
Age * Gender	-1.111	2.25	643.93	-0.494	0.621
Age * Motivation (Romantic)	-1.999	0.923	1480	-2.166	**0.03**
Age * Motivation (Self-esteem)	-2.394	0.923	1480	-2.594	**0.01**
Gender * Motivation (Romantic)	2.841	1.895	1480	1.499	0.134
Gender * Motivation (Self-esteem)	-1.853	1.895	1480	-0.978	0.328
Age * Concealment	-1.199	0.923	1480	-1.299	0.194
Gender * Concealment	2.432	1.895	1480	1.283	0.2
Motivation (Romantic) * Concealment	1.449	1.309	1480	1.107	0.269
Motivation (Self-esteem) * Concealment	1.625	1.309	1480	1.241	0.215
Age * ICS	-0.509	0.845	643.93	-0.602	0.547
Gender * ICS	4.901	2.098	643.93	2.336	**0.02**
Motivation (Romantic) * ICS	0.276	0.942	1480	0.293	0.77
Motivation (Self-esteem) * ICS	0.148	0.942	1480	0.157	0.875
Concealment * ICS	-0.305	0.942	1480	-0.324	0.746
Age * Gender * Motivation (Romantic)	1.537	2.03	1480	0.757	0.449
Age * Gender * Motivation (Self-esteem)	2.89	2.03	1480	1.424	0.155
Age * Gender * Concealment	2.136	2.03	1480	1.053	0.293
Age * Motivation (Romantic) * Concealment	1.392	1.305	1480	1.066	0.286
Age * Motivation (Self-esteem) * Concealment	0.935	1.305	1480	0.717	0.474
Gender * Motivation (Romantic) * Concealment	-1.219	2.68	1480	-0.455	0.649
Gender * Motivation (Self-esteem) * Concealment	-3.131	2.68	1480	-1.168	0.243
Age * Gender * ICS	-0.722	2.07	643.93	-0.349	0.727

Age * Motivation (Romantic) * ICS	0.416	0.763	1480	0.546	0.585
Age * Motivation (Self-esteem) * ICS	0.444	0.763	1480	0.582	0.561
Gender * Motivation (Romantic) * ICS	-0.999	1.893	1480	-0.528	0.598
Gender * Motivation (Self-esteem) * ICS	-1.201	1.893	1480	-0.634	0.526
Age * Concealment * ICS	-0.217	0.763	1480	-0.285	0.776
Gender * Concealment * ICS	0.202	1.893	1480	0.107	0.915
Motivation (Romantic) * Concealment * ICS	1.489	1.332	1480	1.118	0.264
Motivation (Self-esteem) * Concealment * ICS	1.584	1.332	1480	1.189	0.234
Age * Gender * Motivation (Romantic) * Concealment	-1.359	2.87	1480	-0.474	0.636
Age * Gender * Motivation (Self-esteem) * Concealment	-2.203	2.87	1480	-0.767	0.443
Age * Gender * Motivation (Romantic) * Concealment * ICS	-1.252	1.868	1480	-0.67	0.503
Age * Gender * Motivation (Self-esteem) * Concealment * ICS	-0.562	1.868	1480	-0.301	0.764
Age * Gender * Concealment * ICS	0.383	1.868	1480	0.205	0.837
Age * Motivation (Romantic) * Concealment * ICS	-0.159	1.079	1480	-0.147	0.883
Age * Motivation (Self-esteem) * Concealment * ICS	-0.047	1.079	1480	-0.043	0.965
Gender * Motivation (Romantic) * Concealment * ICS	-0.961	2.677	1480	-0.359	0.72
Gender * Motivation (Self-esteem) * Concealment * ICS	-1.878	2.677	1480	-0.701	0.483
Age * Gender * Motivation (Romantic) * Concealment * ICS	0.712	2.642	1480	0.27	0.788
Age * Gender * Motivation (Self-esteem) * Concealment * ICS	0.317	2.642	1480	0.12	0.905

Note: Parameters in bold are significant. '*' denotes interaction between variables.

Appendix I. Power calculations for CFMT.