City Research Online

City, University of London Institutional Repository

Citation: Cenac, Z. (2021). The face and race: facial recognition, categorisation, and variability. (Unpublished Doctoral thesis, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/25894/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

THE FACE AND RACE:

FACIAL RECOGNITION, CATEGORISATION, AND VARIABILITY

Zarus Julian Cenac

This thesis is submitted for the degree of Doctor of Philosophy (PhD)

City, University of London

Department of Psychology

October 2019

TABLE OF CONTENTS

Chapter 1 : INTRODUCTION 12
1.1 The face and race 12
1.1.1 Race terminology 15
1.2 The other-race effect 15
1.2.1 Measurement and theory 16
1.2.1.1 Paradigms and quantification 16
1.2.1.2 Face stimuli 18
1.2.1.3 Overview of theories 19
1.2.2 Objective variability 22
1.2.3 Experience 26
1.2.3.1 Individuation and dimensions 26
1.2.3.2 Levels of experience 30
1.2.3.2.1 Experience to recognition 31
1.2.3.2.2 Experience and expertise 32
1.2.3.3 Early experience 33
1.2.4 Experience and motivation 36
1.2.4.1 The valence of experience 39
1.2.4.1.1 Emotional states and observed emotion 40
1.2.4.1.2 Prejudice 42
1.2.4.1.3 Valenced contact and other-race recognition 43
1.2.5 Perceptual expertise 45
1.2.5.1 Featural 47
1.2.5.2 Configural 50
1.2.5.3 Holistic 51
1.2.6 Perceptually-driven: Subjective variability 55
1.2.6.1 Other-race subjective homogeneity 56
1.2.6.2 The other-race effect 57
1.3 Gender categorisation 59
1.3.1 Gender/sexual dimorphism 59
1.3.2 Experience and bias 61
1.3.3 Expertise 62
1.4 Race categorisation 63
1.4.1 Boundary 64
1.4.2 Proficiency 68
1.5 Outline 70
1.5.1 Experience and identity recognition 70
1.5.2 Facial variability 71
1.5.3 Gender categorisation 72
1.5.4 Race categorisation and gender 73
Chapter 2 : EXPERIENCE AND IDENTITY RECOGNITION 75
2.1 Introduction 75
2.2 General method 76
2.2.1 Participants 76
2.2.2 Materials 76
2.2.2.1 Questionnaires 76
2.2.2.2 Face-matching task 81
2.2.3 Procedure 82
2.2.4 Data preparation 83

2.3 Experiment 1 84
2.3.1 Results and discussion 86
2.4 Experiment 2 89
2.4.1 Results and discussion 90
2.5 General discussion 94
2.5.1 Experience 94
2.5.2 Variability 94
2.5.3 Fixation point placement 96
2.5.4 Computer-generated faces 97
2.5.5 Conclusion 99
Chapter 3 : OBJECTIVE AND SUBJECTIVE VARIABILITY 100
3.1 Objective variability 101
3.1.1 Introduction 101
3.1.2 Method 104
3.1.2.1 Skeletal face 104
3.1.2.2 Full face 108
3.1.3 Results 112
3.1.3.1 Skeletal face 112
3.1.3.1.1 Variance 112
3.1.3.1.2 Pattern variability 115
3.1.3.1.3 Interdimensional correlations 116
3.1.3.1.4 Standardised generalised variance 116
3.1.3.2 Full face 119
3.1.4 Discussion 121
3.1.4.1 Skeletal face 121
3.1.4.2 Full face 123
3.2 Subjective variability 124
3.2.1 Introduction 124
3.2.2 Results 128
3.2.3 Discussion 130
3.3 General discussion 130
3.3.1 Routes to variability 131
3.3.2 Magnitudes 131
3.3.3 Conclusion 133
Chapter 4 : GENDER CATEGORISATION 134
4.1 Introduction 134
4.2 Method 136
4.2.1 Participants 136
4.2.2 Materials 137
4.2.2.1 Questionnaires 137
4.2.2.2 Face stimuli 138
4.2.3 Procedure 142
4.2.4 Data preparation 143
4.3 Results 145
4.3.1 Other-race gender effect 145
4.3.2 Other-race expertise disadvantage for gender 147
4.3.3 Other-race gender categorisation bias 150
4.3.4 Correlations 152
4.4 Discussion 152
4.4.1 Expertise 153

4.4.2 Bias 153
4.4.3 Experience 154
4.4.4 Gender/sexual dimorphism 156
4.4.5 Conclusion 157
Chapter 5 : RACE CATEGORISATION AND GENDER 158
5.1 Experiment 4 158
5.1.1 Introduction 158
5.1.2 Method 158
5.1.2.1 Participants 158
5.1.2.2 Stimuli 159
5.1.2.3 Procedure 160
5.1.2.4 Data preparation 161
5.1.3 Results and discussion 161
5.1.3.1 Race category boundary 162
5.1.3.2 Race categorisation proficiency 163
5.1.3.3 Gender categorisation ability 164
5.2 Experiment 5 165
5.2.1 Introduction 165
5.2.1.1 Facial orientation 166
5.2.1.1.1 Morphology 166
5.2.1.1.2 Skin tone 167
5.2.1.2 Gender 167
5.2.2 Method 169
5.2.2.1 Participants 169
5.2.2.2 Stimuli 170
5.2.2.3 Procedure 171
5.2.2.4 Data preparation 172
5.2.3 Results and discussion 172
5.2.3.1 Race category boundary 172
5.2.3.2 Race categorisation proficiency 173
5.3 Experiment 6 175
5.3.1 Introduction 175
5.3.2 Method 176
5.3.2.1 Participants 176
5.3.2.2 Stimuli 176
5.3.2.3 Procedure 176
5.3.2.4 Data preparation 177
5.3.3 Results and discussion 177
5.3.3.1 Race category boundary 177
5.3.3.2 Race categorisation proficiency 178
5.4 General discussion 179
5.4.1 Upright faces 180
5.4.1.1 Boundary 180
5.4.1.2 Proficiency 182
5.4.1.3 Gender categorisation 185
5.4.2 Orientation 186
5.4.2.1 Boundary 186
5.4.2.2 Proficiency 186
5.4.3 Luminance 187
5.4.4 Conclusion 188

Chapter 6 : GENERAL DISCUSSION 190
6.1 Experience 190
6.1.1 Recap and implications 190
6.1.2 Limitations and future research 191
6.1.2.1 Experience and eye-movements 191
6.2 Objective and subjective facial variability 193
6.2.1 Recap and implications 193
6.2.2 Limitations and future directions 195
6.2.2.1 The number of faces and dimensions 195
6.2.2.2 Dimension comparisons 196
6.2.2.3 Objective variability: Behavioural 197
6.2.2.4 Expertise 198
6.3 Gender categorisation and race 199
6.3.1 Recap and implications 199
6.3.2 Limitations and future directions 201
6.3.2.1 Non-engagement 201
6.3.2.2 Expertise 202
6.3.2.3 Bias 203
6.4 Race categorisation 204
6.4.1 Recap and implications 204
6.4.2 Limitations and future directions 205
6.4.2.1 Participant race and experience 205
6.4.2.2 Racial prototypicality 206
6.4.2.3 Bias 206
6.4.2.4 Morphology and skin tone 208
6.4.2.5 Multiracial categorisation 209
6.4.2.6: Oval 209
6.5 Conclusion 210
REFERENCES 211
APPENDICES 251

LIST OF TABLES

Table 2.1 Non-valenced Inter-Racial Contact Questionnaire items..................... 79
Table 2.2 Inter-Racial Contact Questionnaire postively- and negatively-valenced items.. 79
Table 2.3 For Experiment 1, ordinal alphas regarding items of the Inter-Racial Contact Questionnaire (non-valenced) 88
Table 2.4 For Experiment 2, ordinal alphas for valenced items in the Inter-Racial Contact Questionnaire.. 92
Table 2.5 Predictors in a multiple regression concerning the recognition of East Asian faces by non-East-Asians.. 93
Table 2.6 Regressing predictors on recognition accuracy for African faces amongst non-Blacks... 93
Table 3.1 A list of dimensions in the skeletal face... 106
Table 3.2 Facial diversity analysis.. 120
Table 3.3 Relationships between other-race subjective homogeneities and other-race effects 129
Table 4.1 Ordinal alphas regarding non-valenced items of the Inter-Racial Contact Questionnaire... 138
Table 5.1 Luminance of some stimuli used in Experiment 4............................. 171

LIST OF FIGURES

Figure 1.1 The pathway to the other-race effect. 22
Figure 1.2 Exemplar-based face-space. 29
Figure 1.3 The race category boundary. 65
Figure 2.1 An example trial from the face recognition task. 83
Figure 2.2 The other-race experience disadvantage. 86
Figure 2.3 Face recognition performance of Caucasian participants for own- and other-race faces. 87
Figure 2.4 Hypothetical links between experience and face recognition for various extents of facial diversity. 95
Figure 3.1 Defining unequal variability from migratory distances and confidence intervals. 108
Figure 3.2 Mean variance of the face declines as migratory distance increases. 114
Figure 3.3 The relationship between pattern variability and migratory distance. 115
Figure 3.4 Standardised generalised variance diminished as migratory distance expanded. 118
Figure 3.5 Magnitudes of other-race subjective homogeneity and the other-race effect. 129
Figure 4.1 Example stimuli from a male-female morph continuum. 141
Figure 4.2 A representation of a presentation in the aperture task (gender categorisation). 143
Figure 4.3 Gender categorisation for own- and other-race faces. 147
Figure 4.4 Performance on the aperture task for gender discrimination. 149
Figure 4.5 Point of subjective equality regarding gender categorisation. 151
Figure 5.1 Monoracial race category boundaries for upright faces. 163
Figure 5.2 Race categorisation proficiency for upright faces. 164
Figure 5.3 Race category boundaries regarding faces presented upright and inverted. 173
Figure 5.4 Proficiency for monoracially categorising faces at different facial orientations. 174
Figure 5.5 The boundary between race categories concerning luminance-typical and luminance-equalised faces. 178
Figure 5.6 Racial categorisation aptitude with typical and equalised facial luminance. 179
Figure 5.7 Prototypicality of genders for race categories, and biases. 182

ACKNOWLEDGEMENTS

I thank my supervisors: Dr Richard Cook and Dr Kielan Yarrow. I also thank research collaborator Dr Katie Gray. I additionally thank Dr Hans Christian Petersen for the nonparametric bootstrap, Dr Claire Gordon and colleagues, and Dr William W. Howells for the creation of anthropometric datasets, Dr Sarah Bate and colleagues for the availability of their face recognition dataset, and Dr Debbie S. Ma, Dr Joshua Correll, and Dr Bernd Wittenbrink for the Chicago Face Database. Thanks are given to City, University of London for the MPhil/PhD studentship programme.

DECLARATION

I allow the University Librarian to make single copies of this thesis (whether in full or partial) solely for the reason of study.

Parts of the thesis have been represented in:

Conference presentation

Objective facial variability (Chapter 3): Cenac, Z., & Cook, R. (2017, September). *Facial variability in the context of the other-race effect*. Paper presented at the 19th Annual Conference of the British Association for Biological Anthropology and Osteoarchaeology, Liverpool, United Kingdom.

ABSTRACT

Observers make use of facial details to recognise the identity of an individual person, and categorise them in various ways such as in terms of gender and race. However, race can have an influence on the accuracy with which faces are recognised and categorised by gender, whilst gender may affect facial race categorisation. This thesis explored the *face and race* under two topics: the *recognition of identity*, and the *categorisation of gender and race*. Regarding identity recognition, clarification is required concerning the mechanisms which underlie the phenomenon of face recognition memory being lesser for other-race than own-race faces (the other-race effect); the role of childhood experience is uncertain, as is whether facial variability is constant between races/ethnicities, and if the other-race effect is considerably perceptually-driven. As for gender and race categorisation, it is not clear what leads to gender being categorised less proficiently for other- than own-race faces (the other-race gender effect), and in what ways there is an influence of gender on race categorisation. On the topic of identity recognition, this thesis found that i) childhood experience did not relate to the other-race effect or the recognition of other-race faces (Chapter 2), ii) the morphological variability of the face lessened with increasing migratory distance (from inside of Africa), which raises the possibility of racial/ethnic variability differences moderating the other-race effect (Chapter 3), and iii) results favoured a perceptual basis to the other-race effect for Caucasian, yet not East Asian, observers (Chapter 3). On gender categorisation, it was demonstrated that the other-race gender effect related to other- vs. own-race differences in the local facial processing of gender and gender categorisation bias (Chapter 4). As for race categorisation, facial gender affected race category boundaries and race categorisation precision; results suggested that observers account for the lighter skin tone of females (rather than the darker male skin tone) when categorising race, and that morphology (rather than luminance) drives the effect of gender (Chapter 5). Implications are discussed in the context of the other-race effect, the other-race gender effect, and race categorisation.

LIST OF ABBREVIATIONS

ANSUR: 1988 Anthropometric Survey of U.S. Army Personnel
ANSUR II: 2012 Anthropometric Survey of U.S. Army Personnel
B: Black
C: Caucasian
CI: Confidence interval
CFMT: Cambridge Face Memory Test
CFPT: Cambridge Face Perception Test
Dim.: dimorphism
EA: East Asian
F: Female
Geo.: geometric
IRCQ: Inter-Racial Contact Questionnaire
M: Male
Mig. dist.: migratory distance
MIC: mean interdimensional correlation
MPV: mean pattern variability
MV: mean variance
PI20: The 20-item prosopagnosia index
PSE: Point of subjective equality
SGV: standardised generalised variance

Chapter 1: INTRODUCTION

1.1 The face and race

With the migration of anatomically modern humans across the Earth from an origin in Africa (Manica, Amos, Balloux, & Hanihara, 2007), *Homo sapiens* came to consist of multiple race groups, and a biological element to race is indicated by groupings which have emerged from genetics studies (e.g., Risch, Burchard, Ziv, & Tang, 2002). Yet there is also a social contribution in categorising race, as indicated by the placement of category boundaries between races being affected by a priming manipulation (Krosch & Amodio, 2014) and temporal changes in the racial categorisation of individuals (Saperstein & Penner, 2012). There are racial differences within a number of facial dimensions (Farkas, Katic, & Forrest, 2005; Porter, 2004; Porter & Olson, 2001), and the face can indeed signal race, as race groups can be distinguished from each other on the basis of facial information (Ge et al., 2009; Valentine & Endo, 1992) and amounts of racial genetic ancestry can be deduced from the face (Klimentidis & Shriver, 2009).

Whilst being useful for categorising race (e.g., Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011; Krosch & Amodio, 2014; Pilucik & Madsen, 2017; Valentine & Endo, 1992), the human face is also helpful for observers in recognising the facial identity of an individual (Duchaine & Nakayama, 2006; Murphy & Cook, 2017), and categorising gender (Chatterjee & Nakayama, 2012; Yamaguchi, Hirukawa, & Kanazawa, 1995; Zhao & Hayward, 2010). However, race can seem to affect how well faces are recognised in terms of individual identity (Hancock & Rhodes, 2008) and categorised by gender

(Zhao & Bentin, 2008).

Compared to the body, the face is morphologically heterogeneous regarding within-dimension variability, and this diversity indicates the relative usefulness of the face for recognising the identity of individuals (Sheehan & Nachman, 2014).[1] People are worse at recognising the individual identities of unfamiliar faces than they are at recognising familiar faces (Burton, Wilson, Cowan, & Bruce, 1999). Amongst unfamiliar faces, identity recognition can be worse for an observer when faces are from a race category which is different from their own (i.e., *other*-race) as opposed to the same (*own*-race) (Chiroro & Valentine, 1995). Relatively poorer recognition memory for those other-race faces has been termed the other-race effect (e.g., O'Toole, Deffenbacher, Abdi, & Bartlett, 1991; Hancock & Rhodes, 2008). The other-race effect is called by a number of different names, such as the cross-race effect, and the own-race bias (Brigham, Bennett, Meissner, & Mitchell, 2007; Chiroro, Tredoux, Radaelli, & Meissner, 2008). Similarly, an other-ethnicity effect (also known as the own-ethnicity effect) (e.g., Horry, Wright, & Tredoux, 2010) can occur (i.e., lesser recognition memory for other- than own-ethnicity faces) even concerning different ethnic groups of the same race (McKone et al., 2012).

Interestingly, it seems that the relative difficulty with other-race faces

[1] There has been research which favours the body being heterogeneous relative to the face/head (Lucas & Henneberg, 2016), but this depends on the definition of variability. Lucas and Henneberg (2016) used eight morphological measurement dimensions from the face/head, and another eight from the body. Starting with one dimension, they gradually increased the number of dimensions, one at a time, to find the number where no two individuals would be the same as each other across all of the selected dimensions. The number of dimensions required, such that an individual was unique, was (numerically) lower for the body than the face/head; Lucas and Henneberg (2016) stated that the human "body is more variable than the face" (p. 533). Still, their search for uniqueness is of a different nature to considering the variability within dimensions.

(Malpass & Kravitz, 1969) can also occur outside of identity recognition; whilst sexual dimorphism between males and females is present in the face (e.g., Ferrario, Sforza, Pizzini, Vogel, & Miani, 1993; Mydlová, Dupej, Koudelová, & Velemínská, 2015; Samal, Subramani, & Marx, 2007; Tanikawa, Zere, & Takada, 2016), male/female sex categorisation can be better amongst own-race faces than other-race faces (O'Toole, Peterson, & Deffenbacher, 1996) (i.e., the other-race sex effect), as can gender categorisation using male and female categories (Zhao & Bentin, 2008) which is an occurrence that is called the other-race gender effect.

This thesis covers two themes on the overall topic of *the face and race*: i) (individual) identity recognition, and ii) the categorisation of gender and race. The remainder of Chapter 1 will review existing research on each of these two themes as a springboard for the research presented in Chapters 2 through 5. Identity recognition will be covered in Section 1.2, whilst gender and race categorisation will respectively be considered in Sections 1.3 and 1.4. Concerning identity recognition, Section 1.2 details theories and evidence as to why the other-race effect occurs. As for gender and race categorisation, Section 1.3 reflects on whether the other-race effect is a subset of a more general difficulty with other-race faces rather than merely concerning facial identity recognition; it considers the possibility of a broader problem being at work by reviewing research on the other-race gender effect (e.g., Zhao & Hayward, 2010). Also, on gender and race categorisation, Section 1.4 considers the ability of gender to affect race categorisation. Lastly, in Section 1.5, the goals and objectives of this thesis are outlined.

1.1.1 Race terminology

In this thesis, an effort is made to avoid *colour* terms (e.g., Black, White) whenever possible, and to use land-based terms. Black refers to "[a] person with African ancestral origins, who self identifies, or is identified, as Black, African or Afro-Caribbean" but "[i]n some circumstances … signifies all non-white minority populations" (Bhopal, 2004, p. 443); Black may not necessarily always mean African. However, in the current thesis, it is generally assumed that the previous research referred to which used *Black* participants/stimuli (e.g., Pauker et al., 2009) did so regarding Africans alone, given the locations of those studies and the stimuli that they used. Nonetheless, a search of the literature on the topics of the other-race effect and race categorisation did show that *Black* has been used for non-Africans (see Section 1.2.2). Where previous research has used the term Black, this thesis will not substitute with African unless it is clear that solely Africans were used in the Black category.

1.2 The other-race effect

The importance of the other-race effect can be judged from what it may lead to. Embarrassment has been considered as a possible outcome for the person observing the face (Hugenberg, Young, Bernstein, & Sacco, 2010), and has, anecdotally, been stated to occur not only for the observer, but also for the misrecognised person (Wan et al., 2017). Still, it has been the relevance of the other-race effect for eyewitness misidentifications which has been a focal point for theories and experiments (Brigham et al., 2007; Knuycky, Kleider, & Cavrak, 2014).

In the United States of America, of the initial 190 cases where incorrect eyewitness identifications and convictions occurred but DNA evidence subsequently lead to exonerations, 159 were rape convictions (Garrett, 2011). These “cases might not have gone forward had the victim not been able to identify the defendant; the victim identifications were often crucial to closing the case” (furthermore, the accused was unfamiliar to the victim in the majority of cases) (Garrett, 2011, p. 51). This indicates the importance of eyewitness misidentifications in the convictions which were overturned via DNA. Figures from the Innocence Project show that, in the United States, eyewitness misidentifications featured in 246 cases where DNA evidence later quashed convictions, and indicate that a substantial percentage of the 246 were other-race (Innocence Project, 2017).[2] Furthermore, problems of other-race identification need not merely stem from eyewitnesses and suspects being of different races, but also from the process of creating police lineups, as suggested by less stringency when forming other-race lineups in the experimental setting (Brigham & Ready, 1985).

1.2.1 Measurement and theory

1.2.1.1 Paradigms and quantification

Different paradigms have been used to explore the other-race effect (e.g., Gwinn, Barden, & Judd, 2015; McKone et al., 2012). The core of an other-race

[2] The Innocence Project states that 42% of the 246 were other-race (Innocence Project, 2017). They use Latino as a race category (Innocence Project, n.d.); fifty-three percent of Latino/Hispanic persons selected White monoracial in the 2010 US census (Humes, Jones, & Ramirez, 2011) (on average, Latinos in the United States of America have 65% European Caucasian ancestry, Bryc, Durand, Macpherson, Reich, & Mountain, 2015), yet it should be noted that there is no Latino race category featured in the U.S. Census (United States Census Bureau, 2017). Nevertheless, assuming that the percentages of exonerees by race (whether eyewitness misidentifications were involved or not) (Innocence Project, n.d.) are representative of instances of eye misidentifications, this would still be suggestive of a sizeable proportion of the 246 being other-race whether Latino is considered as a race or not.

effect experiment usually consists of i) sequentially presenting participants with unfamiliar own- and other-race faces (learning phase), followed by ii) a disparate activity which lasts a number of minutes, and then iii) testing the ability of each participant to distinguish between faces from the learning phase and novel faces, with *old*/*new* decisions made upon viewing each face (test phase) (e.g., Hills & Lewis, 2011; Young & Hugenberg, 2012). As another option, Caucasian and East Asian versions of the Cambridge Face Memory Test (CFMT) measure the ability of an observer to recognise target faces amongst distracters of the same race under various viewing conditions, with any one of the targets being present on each experimental trial (Duchaine & Nakayama, 2006; McKone et al., 2012). The CFMT is a computer-based face recognition task which is useful in differentiating between typical and atypically poor facial recognition ability (Duchaine & Nakayama, 2006) (also see Section 3.2.1). In an alternative paradigm (a sequential matching task), learning and test phases can occur per trial; a singular face is displayed, then it vanishes before the same face is presented once more or a different face is shown, and a participant responds whether the *same* face was shown twice, or if they were *different* faces to each other (e.g., Lindsay, Jack, & Christian, 1991; Michel, Rossion, Bülthoff, Hayward, & Vuong, 2013).[3]

Face recognition performance may be tallied in a number of ways (e.g., Nguyen & Pezdek, 2017). For instance, it can be measured via sensitivity scores, such as *d′*, which is calculated from hits (e.g., correct identifications of faces at

[3] Despite the brief retention interval in the sequential face-matching task, in the paradigm "comparison is always between a target image and a memory trace" (Megreya, White, & Burton, 2011, p. 1476), i.e., memory is still involved.

test as having been present at learning) given false alarms (e.g., incorrectly identifying *new* faces as being *old*, or *different* as *same*), with the other-race effect present when *d'* is larger for own-race faces than it is for other-race faces (Bothwell, Brigham, & Malpass, 1989; Michel et al., 2013; Shriver & Hugenberg, 2010). Another metric for investigating the other-race effect is response bias (e.g., *C*) (Rhodes, Locke, Ewing, & Evangelista, 2009); being more lax when making *old* or *same* decisions for other-race faces (Cassidy, 2011) could be due to other-race faces subjectively appearing more similar, thereby encouraging the old/same response for those faces (Wells & Olson, 2001). For the Cambridge Face Memory Test, the percentage of correct identifications of target faces is the accuracy measure (Duchaine & Nakayama, 2006), therefore the other-race effect has arisen when the percentage is greater for own-race faces than it is for other-race faces (McKone et al., 2012).

1.2.1.2 Face stimuli

Experiments on the other-race effect differ in terms of the non-facial traits included in stimuli; the face alone has been used (e.g., Michel, Rossion, Han, Chung, & Caldara, 2006) and, additionally, features beyond it but still part of the head nonetheless, such as the ears and head hair from the scalp (e.g., Baldwin, Keefer, Gravelin, & Biernat, 2013). Hence, in this thesis, the word *face* is not always being applied in a strict fashion.

Whilst some experiments have presented a different image of the unfamiliar face at test than at learning (Crookes et al., 2015), others have shown the same unfamiliar face image at both phases (e.g., Young & Hugenberg, 2012; Johnson & Fredrickson, 2005). Under the face processing model of Bruce and Young (1986), different types of information can be extracted from the image of a face. One type, structural, refers to the actual facial traits whilst another type, pictorial, are properties within the image that may change across different images of that face (e.g., lighting) (Bruce & Young, 1986). Indeed, in their other-race effect study, Gwinn et al. (2015) used different pictures at test than learning such that "[u]nlike some prior research, this provided a pure test of memory for the target individuals, rather than memory for particular photographs" (p. 4).

Yet it has been suggested that, unlike familiar faces, the processing of unfamiliar faces is greatly dependent on pictorial information, and hence their processing is affected by picture changes (Hancock, Bruce, & Burton, 2000; Burton, 2013); given that studies on the other-race effect use unfamiliar faces, it could be argued that processing would substantially be picture-based. The question then becomes whether performance in other-race effect experiments where learning and test images match (e.g., Hancock & Rhodes, 2008) is reflective of participants using facial information. In Hancock and Rhodes (2008), who used the same unfamiliar image at learning and test, experience with other-race persons (and hence their faces, as measured by a questionnaire) related to the magnitude of the other-race effect. This relationship with experience indicates that results in other-race effect experiments considerably involve face processing, and this would be true whether such experiments do or do not employ the same image of a face at learning and test.

1.2.1.3 Overview of theories

The other-race effect is not short on theories which attempt to explain it (e.g., Valentine, 1991; Levin, 2000; Sporer, 2001). One line of thought was that races can be of unequal facial variability, and that the other-race effect occurs due to own-race faces being more objectively variable than faces of an other-race, however, previous research has found no racial differences in facial structural variability (Goldstein, 1979a). Unequal variability has generally been rejected as a theory (e.g., Ng & Lindsay, 1994) although not entirely (as a contributor to the other-race effect) (Rossion & Michel, 2011). Ways of grouping the currently prominent theories vary, as do the labels which those clusters attract (Ng & Lindsay, 1994; Meissner & Brigham, 2001). The most popular types of theories which have emerged from the literature are experiential, socio-cognitive, and convergences of the two alongside motivation (Bernstein, Young, & Hugenberg, 2007; Wan, Crookes, Reynolds, Irons, & McKone, 2015; Hugenberg, Wilson, See, & Young, 2013).

Experiential perspectives construe the other-race effect as having arisen from greater experience individuating own-race faces (than other-race faces), thereby resulting in the possession of better expertise for own-race faces, and causing superior recognition memory for them (Valentine, 1991; Walker & Hewstone, 2006). Socio-cognitive approaches consider the other-race effect as a consequence of social categorisation, with individuation happening more for ingroup than outgroup members (Levin, 2000); such an approach can essentially be collapsed under the motivational stream, with motivation being greater for individuating (versus categorising) ingroup persons than members of an outgroup (Hugenberg et al., 2010; Wan et al., 2015). As for motivation more generally, it exerts control over the *use* of expertise (Baldwin et al., 2013; Hugenberg et al., 2010).

Different types of expertise have been explored regarding the other-race effect: *featural*, *configural*, and *holistic* (e.g., Mondloch et al., 2010; Wang et al., 2015). Featural is in regard to components (e.g., the nose), configural (i.e., second-order relations) concerns the spaces between components, and holistic refers to processing the object (in this case the face) as whole rather than simply as separate elements (Diamond & Carey, 1986; Hayward, Rhodes, & Schwaninger, 2008; Maurer, Le Grand, & Mondloch, 2002). For each of these three expertise types, relatively greater own-race expertise has been theorised to have a function in the other-race effect (Hayward et al., 2008; Rhodes, Ewing, Hayward, Maurer, Mondloch, & Tanaka, 2009). Therefore, other-race faces may subjectively *seem* homogeneous, not due to being objectively less variable than own-race faces (Goldstein, 1979a), but because of lower expertise (Hugenberg et al., 2010). Therefore, other-race subjective homogeneity would be a consequence of experience and motivation (Figure 1.1); other-race subjective homogeneity may then lead to the other-race effect (Goldstein, 1979a).

In the remainder of Section 1.2, theories of the other-race effect are discussed, beginning with objective facial variability (Section 1.2.2), followed by experiential (Section 1.2.3), and then motivational (including socio-cognitive) alongside experiential (Section 1.2.4). Afterwards, each of the three different types of expertise are covered (Section 1.2.5) along with other-race subjective homogeneity (Section 1.2.6).

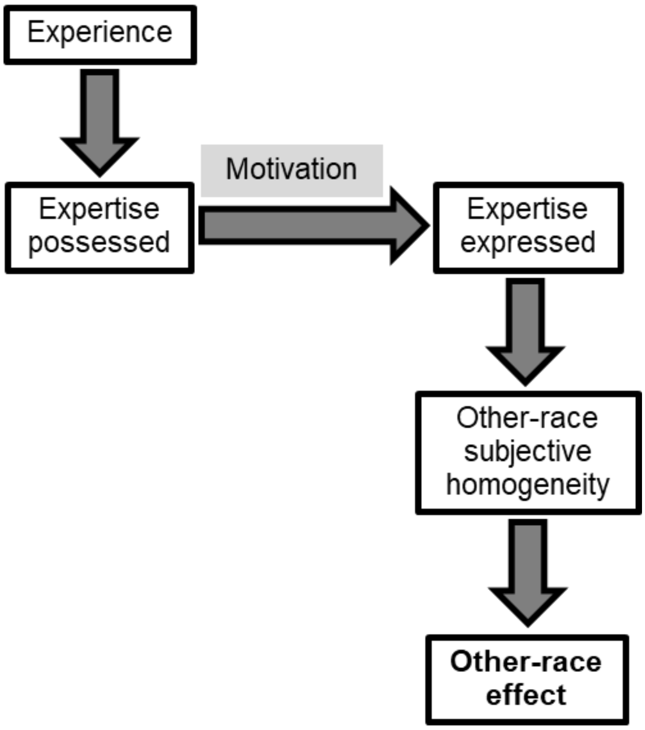


Figure 1.1. This represents a summary of theorisations of the other-race effect (e.g., Goldstein 1979a; Hugenberg et al., 2010; Rossion & Michel, 2011; Valentine, 1991), i.e., being more experienced with own-race faces than other-race faces confers a relatively higher possessed level of expertise (e.g., holistic) for own-race faces, and motivation directs the use of expertise (i.e., greater expressed expertise for own-race), culminating in other-race subjective homogeneity from which the other-race effect manifests.

1.2.2 Objective variability

An early proposed reason for the other-race effect was that it occurs when faces within an other-race category are less variable than own-race faces (hence causing a relative difficulty distinguishing between other-race faces), nonetheless, a study in the 1970s found races to be equally diverse (Goldstein, 1979a; Meissner & Brigham, 2001). That study, Goldstein (1979a), is essentially used as the go-to article concerning races being of matching objective facial variability (e.g., Byatt & Rhodes, 2004; Papesh & Goldinger, 2010). Even so, Section 1.2.2 argues that there are still uncertainties concerning whether objective morphological variability is the same from one race to another.

Goldstein (1979a) used measurement data that were largely of the face proper (sourced from previous studies) and undertook racial comparisons within and between genders. Regarding specific comparisons, for brevity, only within-gender ones are summarised next. It was found that, for males, morphological dimensions of Caucasian groups (in Ireland and Russia) were as variable as Japanese East Asians. Compared to Caucasian males in Hawaii, Japanese East Asian males were noted as more variable on approximately 80% of dimensions, however “only 25 white and 33 Japanese East Asian men were measured … so conclusions involving these data should be drawn with caution” (Goldstein, 1979a, p. 189). Pertaining to females, Goldstein (1979a) found that Caucasians (Ireland) matched the variability of Japanese East Asians.[4] Amongst males, Caucasians had the same variability as Blacks. No comparison was undertaken between females regarding Blacks and Caucasians as there were very few measurement dimensions available which were common to both groups (Goldstein, 1979a). Blacks were not compared to Japanese East Asians (Goldstein, 1979a).

Overall, it was stated that “[c]omparisons among three racial groups yielded no evidence for racial differences in facial heterogeneity, but features of Japanese women's faces may display more variation than the other faces studied” (Goldstein, 1979a, p. 187). However, noting that Africans hold a higher genetic diversity than non-Africans, Rossion and Michel (2011) suggested that within-

[4] Although, on this comparison, Goldstein (1979a) stated it there was “no evidence in these data for any race-related difference in facial variability” (p. 189), Goldstein (1979b), interpreted the comparison in Goldstein (1979a) between Japanese East Asian females and Caucasian females as showing the former as being more heterogeneous.

race facial morphology is likely not equally variable across race groups, such as proposing that Africans are heterogeneous in comparison to Caucasians.

It may be worth considering why dissimilar variabilities were not apparent in Goldstein (1979a). For a measurement (e.g., the width of the nose), Goldstein made decisions on differences in variability by seeing whether a coefficient of variation, (*SD*/*M*)*100, was numerically larger, smaller, or had the same value as one from another race category. To assess relative variability, the number and portion of comparisons falling into each of those three decision outcomes was looked at. It could be that this methodology would not have been sufficiently sensitive for finding unequal variability (when variability actually differed). Reh and Scheffler (1996) stated that, on coefficients of variation, "significance tests and calculations of confidence intervals are not generally performed" (p. 449); that was in the 1990s, and would have been the situation in 1979. Nevertheless, Phipps, German, and Smith (1988) used an ANOVA to compare coefficients of variations for facial measurements (the underlying bones and teeth) of Navajo Native Americans and Caucasians between 10 and 12 years of age; overall, no racial variability difference was found.

Measurements for the Black category in Goldstein (1979a) were derived from two studies. As described by Goldstein (1979a), one was represented mainly by men and women born in the United States of America, with some being of the Caribbean. Generally, the genetic ancestry of both African Americans and African Caribbeans is mostly African (Benn-Torres et al., 2008; Bryc, Durand, Macpherson, Reich, & Mountain, 2015; Shriver et al., 2003; Yaeger et al., 2008). African Americans are genetically diverse (Tishkoff et al., 2009), which would presumably apply to other African Diasproran groups such as those in the Caribbean. Following the Rossion and Michel (2011) assertion, if the extent of genetic heterogeneity is reflected facially, there could be an expectation that African Diasporans would be relatively variable.

The other data used in the Black category of Goldstein (1979a) was of men in Bougainville Island, the indigenous population of which is Melanesian (Friedlaender et al., 1971; Friedlaender et al., 2008; Kariks, Kooptzoff, & Walsh, 1957). Regarding that population data, Friedlaender (1975) (where Goldstein obtained the data) removed “off-islanders” (p. 117) from the overall analysis, hence the data from Bougainville Island which was used in Goldstein (1979a) was very certainly of Melanesians. Melanesians have been described as a group who “physically resemble” Africans (Friedlaender et al., 2008, p. 0174), however, in terms of genetic ancestry, Melanesians are separate from Africans (Friedlaender et al., 2008; Risch et al., 2002; Tishkoff et al., 2009) as shown with Melanesians in Bougainville specifically (McEvoy et al., 2010). Indeed, “[o]utside the Pacific, East Asian populations are apparently the closest (but still very distant) relatives of Melanesians. Africans and Europeans are the most distant” (Friedlaender et al., 2008, p. 0187). In Goldstein (1979a), due to a lack of measurements, some coefficient of variation values for Black men were only from the primarily American-born study, others from Bougainville, and the remainder an average of the two. Hence, for males, comparisons were between Caucasians and an assemblage of potentially heterogeneous (African) and not-so-variable (non-African) groups.

With the use of purely descriptive comparisons in Goldstein (1979a), and the possible higher variability of Africans not directly tested with respect to general facial morphology, it was therefore questionable whether races are of equal structural diversity. Yet the idea of the other-race effect being caused by other-race objective homogeneity is said to be opposed by examples of the effect occurring for different groups (Cassidy, 2011), and, indeed, designs which cross two races in participants and stimuli can show other-race effects within both participant groups (e.g., Michel et al., 2013). Additionally, given the influences of experience (Hancock & Rhodes, 2008) and motivation (Baldwin et al, 2013), it is highly unlikely that differences in variability (if they exist) would be *the* key force behind the other-race effect. However, they could moderate it (Rossion & Michel, 2011).

1.2.3 Experience

A research area on the other-race effect which has garnered far more attention than objective variability is that of racial experience. The core idea underlying the experiential component of the other-race effect is that an observer has a relative dearth in experience for faces of an other-race category (compared to own-race faces), which translates into an observer possessing less expertise for other-race faces than own-race faces, and the other-race effect ultimately transpires as a result (e.g., Rossion & Michel, 2011). Section 1.2.3 presents a review of theorisations and studies pertaining to the experiential element. It is asserted that clarity is lacking on i) whether experience within certain childhood life stages relates to the other-race effect in adulthood, and ii) what type of experiences within childhood are key for the adulthood other-race effect.

1.2.3.1 Individuation and dimensions

Some aspects of the face seem better for individuation than others (e.g., Hills, Cooper, & Pake, 2013; Tanaka & Farah, 1993). The subset of dimensions which make a considerable contribution to encoding identity are the dimensions which are the most useful for distinguishing between the faces which an observer has experienced (oftentimes largely own-race faces) (Valentine, 1991). Races may differ in how relatively useful dimensions are for individuation, i.e., the dimensions that are best to use for one race may not apply so well to another race (Hills & Lewis, 2011); Caucasian faces are recognised better when a fixation cross (which guides the first fixation) appears before the upper face (nose bridge) rather than the lower face (nose tip), whilst the opposite holds true for Black faces, which indicates that the upper face contains relatively individuating details for Caucasian faces, whilst the (relatively) central/lower face has that function for the faces of Blacks (Hills et al., 2013; Hills & Pake, 2013). Furthermore, fixating sooner to the mouth region of African faces is beneficial for recognition (unlike for Caucasian faces) (McDonnell, Bornstein, Laub, Mills, & Dodd, 2014). The situation outside of Caucasian faces and Black faces is undetermined. Nevertheless, attending to dimensions that are useful for differentiating between own- but not other-race faces could produce the other-race effect (Hills et al., 2013; Hills & Lewis, 2006, 2011).

The importance of dimension subsets is captured by the face-space models of Valentine (1991). Face-space is a memorial representation of faces which an observer has encountered, with those faces plotted in various dimensions (Valentine, 1991). The dimensions themselves “are not specified” as of yet, and they could be featural, second-order relational, or holistic (Valentine, Lewis, & Hills, 2016, p. 1998). Assuming that a person has relatively great own-race experience, the dimensions which dominate the face-space of that person are the ones which are relevant for encoding the identity of own-race faces, therefore faces from a different race are more clustered in face-space than own-race faces

(Valentine, 1991). Due to facial differences between races (Farkas, et al., 2005; Porter, 2004; Porter & Olson, 2001), faces of an other-race congregate away from own-race faces (Valentine, 1991). This thesis takes the position that the housing of face-space in memory does not negate the idea that other-race subjective homogeneity results in the other-race effect; the relative weight (attention) given to dimensions would lead to faces being perceived in line with dimensional weightings, hence other-race subjective homogeneity would manifest, and encodings into face-space would arise from those perceptions.

There are two face-space models: exemplar-based and norm-based (Valentine, 1991). In the exemplar-based variant, faces are encoded along dimensions without respect to any norm (Figure 1.2); because own-race exemplars are more dispersed than other-race exemplars, there is less confusability between own-race exemplars, and the other-race effect is a consequence (Valentine, 1991). In the norm-based version, each face is encoded into dimensions with regards to their dissimilarity from a face norm which is housed at the centre of face-space (and there is only one face norm). A face is represented on a multidimensional vector which possesses an angle and distance from the norm. The norm would be extrapolated from faces encountered, i.e., generally weighted towards own-race faces, therefore, faces which are own-race would have a higher diversity in their vectors than other-race faces do. Because of this diversity, own-race faces are more distinguishable than other-race faces concerning their individual identities (Valentine, 1991).

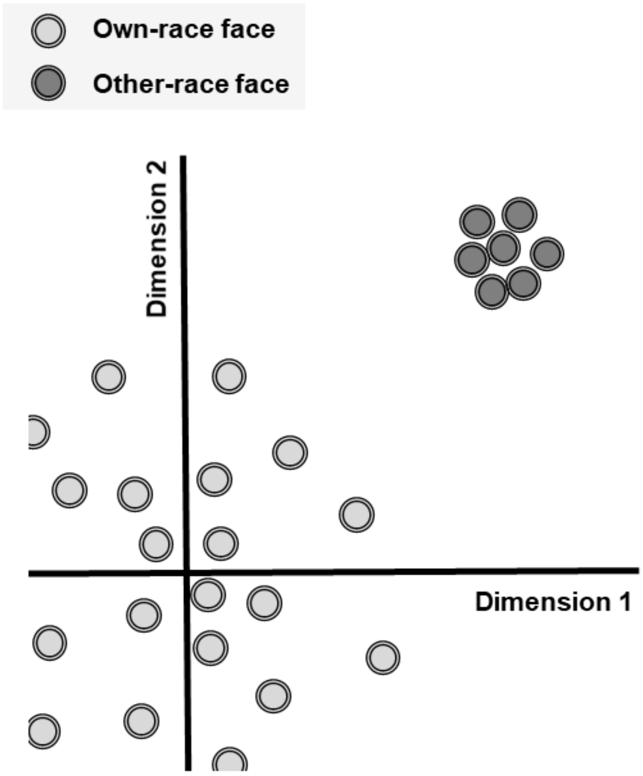


Figure 1.2. Exemplar-based face-space, featuring faces from two race categories (own-race and one other-race) which are plotted in two hypothetical dimensions for simplicity (based on Valentine & Endo, 1992).

If dimensions are equally appropriate for own- and other-race faces, an other-race effect (i.e., for unfamiliar faces, and given that experience is greater for own-race faces than other-race faces) would not occur under the exemplar-based account, but it would happen under the norm-based version due to angles (relative to the one norm) being less diverse for other-race faces (Valentine, 1991).

The existence of one overall face norm is supported by face categorisations (intact face vs. scrambled) being slower for other-race faces than own-race, as vector lengths would be longer for other-race faces (Valentine, 1991). That said, there is evidence favouring the presence of multiple norms (Baudouin & Gallay, 2006; Jaquet, Rhodes, & Hayward, 2008). Gender norms are suggested by the distinctiveness of faces being judged in terms of gender (Baudouin & Gallay, 2006). The existence of race norms is indicated by outcomes of perceptual adaptation, with presentations of two races during adaptation in opposing directions (for example, inner facial features are contracted for own-race but expanded for other-race faces) causing racially-dominant aftereffects (to continue the example, own-race faces seem expanded post-adaptation relative to how they looked at baseline, whilst other-race faces appear contracted) (Jaquet et al., 2008), yet the actuality of multiple norms would not quash there still being an overall face norm (Rhodes, Watson, Jeffery, & Clifford, 2010).

1.2.3.2 Levels of experience

The link between experience and the other-race effect has been explored in a number of ways. One way involves training observers to differentiate between other-race faces (Heron-Delaney et al., 2011; Hills & Lewis, 2006), such as by associating a face with a specific number (Goldstein & Chance, 1985) or letter (Lebrecht, Pierce, Tarr, & Tanaka, 2009; Tanaka & Pierce, 2009). Training Caucasian observers on individuating the lower face region diminishes the Caucasian-Black other-race effect, unlike training on the upper face, which suggests that effective individuation training redefines which dimensions are primarily used to encode identity (Hills & Lewis, 2006).

Alternatively, questionnaires have been used to look into the experiential element of the other-race effect (e.g., Young & Hugenberg, 2012). Questionnaires tally other-race experience (Zhao, Hayward, & Bülthoff, 2014a, 2014b) which has been quantified as percentages of the faces encountered (Cloutier, Li, & Correll, 2014) or in absolute terms (Hancock & Rhodes, 2008). Whether there is a relationship between other-race experience and the other-race effect has been explored by way of correlational (Young & Hugenberg, 2012) and regression analyses (Hancock & Rhodes, 2008; Zhao et al., 2014b), with the idea being that an experiential component is indicated by other-race experience increasing whilst the magnitude of the other-race effect decreases (e.g., Young & Hugenberg, 2012).

Of the experiential questionnaires used in research regarding the other-race effect, three commonly-used and relatively contemporary ones are by Hancock and Rhodes (2008) and Walker and Hewstone (2006). Item-wording can be adjusted so that the group for which experience is quantified is a particular group which is present in face stimuli (e.g., Young & Hugenberg, 2012). Regarding the Hancock and Rhodes (2008) measure, there are seven items, the majority of which inquire into "social interactions" (Hancock & Rhodes, 2008, p. 49), for example, "I socialize a lot with Caucasian people", and responses are given on a six-point agreement scale (Hancock & Rhodes, 2008, p. 56). Walker and Hewstone (2006) presented two questionnaires, with each consisting of items which had five-point scales. One of the Walker and Hewstone (2006) questionnaires focussed on the quantity of contact. Their other questionnaire was designed to capture experience individuating, and featured items such as, a person of a group "has comforted me when I have been feeling sad" (Walker & Hewstone, 2006, p. 468).

1.2.3.2.1 Experience to recognition

Relationships between questionnaire-measured other-race experience and the other-race effect have been clear (e.g., Zhao et al., 2014a). Other-race experience, as measured by the Hancock and Rhodes (2008) questionnaire, is predictive of the magnitude of the other-race effect (Hancock & Rhodes, 2008;

Zhao et al., 2014b). Experience, with items derived from the Walker and Hewstone (2006) contact questionnaire, also predicts the size of the other-race effect (Zhao et al., 2014a, 2014b). Interestingly, in Experiment 1 of Young and Hugenberg (2012), the correlation between the other-race contact (Hancock & Rhodes, 2008, items) and the other-race effect became evident when motivation to individuate other-race faces was raised by way of instructions to individuate, whilst there was no correlation in the absence of those instructions (i.e., a control condition); an experiential contribution was apparent when experiential differences (rather than motivational differences) particularly drove the other-race effect (Young & Hugenberg, 2012).

1.2.3.2.2 Experience and expertise

As for expertise, whether there is generally a relationship between other-race experience (questionnaire-based) and types of other-race disadvantages in expertise tasks would seem ambiguous (e.g., Bukach Cottle, Ubiwa, & Miller, 2012; Rhodes, Ewing, et al., 2009). This would be concerning for the pathway from experience to the other-race effect via expertise (although, see Section 1.2.5 for a brief discussion on processing and memory). Other-race experience was not a significant predictor of other-race featural and configural disadvantages in Rhodes, Ewing, et al. (2009) (using the Hancock & Rhodes, 2008, questionnaire), nor the other-race holistic and featural disadvantages in Zhao et al. (2014b) (Hancock & Rhodes, 2008, items were used with respect to one holistic task, whilst the Walker & Hewstone, 2006, contact questionnaire was employed regarding another holistic activity and featural performance). Yet experience (using the Walker & Hewstone. 2006, contact questionnaire) predicted the other-race configural disadvantage in Zhao et al. (2014b), and the other-race holistic disadvantage was found to correlate with experience (Walker & Hewstone, 2006, individuation questionnaire) in Bukach et al. (2012).

1.2.3.3 Early experience

Perceptual experience from childhood may be particularly important for the presence of the other-race effect during adulthood (Rossion & Michel, 2011; Wan et al., n.d.). This line of reasoning implies that the other-race effect is apparent in childhood; indeed it is, such as amongst 3-year-old Africans and Caucasians (Suhrke et al., 2014). Yet there is evidence supporting the other-race effect not being present throughout *all* of childhood, but only after a certain period of time during early infancy (e.g., Kelly et al., 2009). The development of the other-race effect in infancy can be understood in the context of *perceptual narrowing* (e.g., Kelly et al., 2007), which has been defined as "a decline in the ability, during the 1st year of life, to discriminate unfamiliar types of perceptual stimuli, such as two faces of another race or species" (Scott & Monesson, 2009, p. 676).

Experiments on perceptual narrowing with respect to face recognition during infancy have used a paradigm where the infant is habituated to a face, followed by presentation of the same face and a different face, with the idea being that face recognition ability is indicated by the proportional amount of time the infant looks at the *different* face rather than the *same* face, i.e., longer proportional looking times for the different (i.e., novel) face indicates better face recognition ability (e.g., Heron-Delaney et al., 2011; Kelly et al., 2007, 2009; Scott & Monesson, 2009). Such experiments have shown that experience individuating a face type can spare narrowing with respect to that face type (Scott & Monesson, 2009). Specifically, whilst 6-month-olds proportionally look at novel macaque faces in excess of chance, three months later those who received individuation training still proportionally viewed novel faces of macaques greater than chance, unlike infants who either partook in categorisation training or merely looked at the macaque faces (Scott & Monesson, 2009).

Perceptual narrowing applies amongst the human races, such that, regarding African, Caucasian and East Asian faces, three-month-old East Asian and Caucasian observers proportionally view novel facial identities within each group above chance levels, however, at 6- and 9-months-old respectively, East Asians and Caucasians solely look at novel own-race faces greater than chance, which suggests that the other-race effect arises gradually over the first year (Kelly et al., 2007; Kelly et al., 2009).

However, the plasticity of the face processing system in childhood allows for some adjustments to be made after the first year, and these adjustments are in line with faces experienced in the perceptual environment of the observer (e.g., Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005). For instance, in Sangrigoli et al. (2005), Caucasian faces were recognised better than own-race faces by adult East Asians who resided in Korea prior to being adopted by Caucasians between 3- and 9-years-of-age (6 on average) and raised in Europe; this suggests that considerable plasticity remains at 3-years-old, and experience before that age need not drive the other-race effect (Sangrigoli et al., 2005).

The use of experiential questionnaires has given some insight into which periods of life are important for the appearance of the other-race effect in adulthood (Davis, Hudson, Ma, & Correll, 2016; Wan et al., 2013; Wan et al., n.d.). Wan et al. (n.d.) conducted correlation analyses between the magnitude of the other-race effect and various measures of contact at different times of life. Looking at the proportion of significant correlations vs. non-correlations, the apparently greater prevalence of correlations between measures of other-race (and own-race) experience and the other-race effect (in adulthood) pertaining to experience between 5 and 12 years of age (five significant correlations out of 10), rather than between 12 and 18 years (no correlations) or current contact (1/14 correlations), is indicative of relatively early childhood experience being important for the other-race effect (Wan et al., n.d.).

Interestingly, Wan et al. (2013) examined whether other-race experience disadvantages (own-race experience minus other-race experience) correlated with the other-race effect. Studies on contact and recognition differences usually employ *other-race contact* as the experiential index (e.g., Davis et al., 2016), and not this *racial difference* in contact. Experiential measures were percentage contact between the times of 5-12 years, 12-18 years, and adulthood, separately for classmates, friends, and neighbours; the only contact difference which correlated with the size of the other-race effect was the one for classmates at 5-12 years (Wan et al., 2013).

Results in Davis et al. (2016) did not demonstrate that childhood contact relates to the adulthood other-race effect. Davis et al. (2016) applied a questionnaire by Cloutier et al. (2014) which enquires into the percentage composition of contact regarding persons whom the respondent knew when the respondent was 0-6, 6-12, or 12-18 years old. Davis et al. (2016) took an average of timespans to provide a number for childhood contact (i.e., 0-18 years) of Caucasians with Blacks. Whilst the experiment was planned with response latencies in mind (Davis et al., 2016), the supplementary portion of their article did show that other-race effects occurred regarding accuracy measures (error rates and sensitivity). Magnitudes of the other-race effect did not change with other-race contact during childhood (Davis et al., 2016). Childhood contact did, however, relate to a proposed indicator of the other-race holistic disadvantage measured in response latencies (Davis et al., 2016).

Overall, experience serves a dual function of acquiring (Hills & Lewis, 2011) and retaining the ability to recognise faces (Heron-Delaney et al., 2011; Scott & Monesson, 2009). It has been suggested that childhood experience is a determiner of the other-race effect (Rossion & Michel, 2011), yet, for two accuracy metrics, magnitudes of the other-race effect itself were not affected by childhood contact (0-18 years broadly) in Davis et al. (2016). However, there is evidence which is supportive of experience from some childhood timespans being more important than others for the prevalence of the other-race effect in adulthood (Wan et al., n.d.). Given previous studies (Davis et al., 2016; Wan et al., 2013; Wan et al., n.d.), it seems unclear which timespans of childhood contact are valuable for the adulthood other-race effect, if any.

1.2.4 Experience and motivation

Through experience, an observer possesses levels of expertise for different race categories, however, those magnitudes of expertise may not be used to their fullest whether for own-race faces (Bernstein et al., 2007) or other-race faces (Baldwin et al., 2013). Section 1.2.4 covers the motivational element of the other-race effect, and its partnership with experience.

A determiner of expertise engagement is social categorisation (Bernstein et al., 2007). Regarding social categorisation and the other-race effect, the *feature-selection hypothesis* of Levin (2000) holds a socio-cognitive factor as not only influential, but (rather than experiential differences) wholly responsible for other-race faces being processed less expertly and consequently being recognised relatively poorly; this opposes an experience-expertise approach. Levin (2000) reasoned that own-race persons are thought of more as individuals than other-race persons are, which leads to observers looking for (and encoding) individuating aspects of the face more so for own-race faces than other-race faces. Other-race faces are considered more at the level their race category, and therefore they are encoded more in regard to facial details which align with race category membership (i.e., homogeneously) rather than facial information which differentiates between faces of that other-race (Levin, 2000).

Moreover, own- and other-race designations were interpreted by Levin (2000) as being ingroup and outgroup categorisations respectively. Effects of non-racial social groupings would be compatible with the feature-selection hypothesis and incompatible with a pure experience-expertise theory, and such effects do occur as shown by Caucasians exhibiting a recognition advantage for ingroup faces even when ingroup and outgroup faces are all Caucasian (Bernstein et al., 2007). Nevertheless, there are occurrences which a strict socio-cognitive theory cannot provide a reason for, and these occurrences are i) correlations between experience and the other-race effect (e.g., Hancock & Rhodes, 2008; Young & Hugenberg, 2012), and ii) effects of motivation outside of social categorisation (Baldwin et al., 2013).

Regarding the second occurrence, informing participants of the other-race effect and providing instructions which encourage them to individuate faces (particularly other-race faces) has been sufficient to cause the other-race effect to not be evident (Hugenberg, Miller, & Claypool, 2007; Rhodes, Locke, et al., 2009; Young & Hugenberg, 2012). Furthermore, Baldwin et al. (2013) reduced the other-race effect via evoking outcome dependency for an other-race target, and, in another experiment, Baldwin et al. (2013) found that the recognition of own- and other-race faces was equalised by having other-race faces appear larger than own-race ones via the Ponzo illusion. Baldwin et al. (2013) proposed that these manipulations increased how vital an other-race person seemed to be for achieving goals, which enhanced the motivation to individuate those faces. A staunch socio-cognitive perspective cannot explain such influences on the other-race effect (which are separate of social groupings) and yet an experience-expertise viewpoint is also unable to provide a reason (Baldwin et al., 2013).

The inability of experience-expertise and socio-cognitive strands to separately account for the other-race effect has led to the ascent of models that combine approaches (e.g., Sporer, 2001; Hugenberg et al., 2010). Sporer (2001) suggested an *in-group/out-group model.* This theory allows for socio-cognitive effects and contains perceptual expertise as a factor, but no role is allocated for motivation at encoding distinct from grouping. The *categorisation-individuation model* of Hugenberg et al. (2010) not only includes this more general motivation, but assigns motivation to a trio of key factors alongside perceptual expertise (via experience) and socio-cognitive (motivational) aspects. Within that model, motivation affects recognition by directing expertise use. The theory has socio-cognitive elements which do influence the employing of expertise, but only through motivation, with the idea being that observers can have a greater motivation to individuate ingroup faces than outgroup faces. Therefore, within the categorisation-individuation model, there are three key own- vs. other-races *differences* through which an other-race effect can present: i) a relative lacking in possessed other-race expertise which the observer holds from experience, ii) motivational (including the socio-cognitive factor), with motivation to individuate other-race faces being comparatively low thereby leading to a lower level of expertise being applied, and iii) a combination of both low experience-expertise and low motivation.

1.2.4.1 The valence of experience

Motivation may promote the gaining of expertise from visual exposure to faces (Hugenberg et al., 2010), i.e., something which affects face recognition performance by influencing the motivation to individuate may also guide the acquisition (and maintenance) of expertise, which would influence face recognition. Interestingly, the items (five in total) in the Walker and Hewstone (2006) individuation experience questionnaire would appear to largely have been of positive valence. For instance, Walker and Hewstone (2006) (who used race categories of South Asian and White [cf. Risch et al., 2002; however, see Basu, Sarkar-Roy, & Majumder, 2016; Reich, Thangaraj, Patterson, Price, & Singh, 2009]) included items such as “I have looked after or helped a South Asian … friend when someone was causing them trouble or being mean to them” (p. 468). In their study, amongst White participants, experience individuating South Asians predicted the recognition of South Asian faces, after controlling for recognising White faces and the quantity of contact with South Asians (Walker & Hewstone, 2006).[5] This raises the possibility that positive (rather than negative) experiences may deliver particularly individuating experience.

[5] A corresponding analysis did not happen with South Asian participants, as they recognised South Asian and White faces equally well (Walker & Hewstone, 2006). Analysis has occurred in an experiment testing whether the other-race effect correlates with the size of the other-race disadvantage in featural processing despite there being no difference in own- compared to other-race featural performance (i.e., this type of difference was not present on average) in that study (DeGutis et al., 2013), however, the researchers did note that “a somewhat restricted range in” that difference “may decrease this correlation” (DeGutis et al., 2013, p. 8).

The manner in which the valence of contact potentially affects other-race recognition could be supposed from the literature of two areas: emotion (emotional states and observing emotion) and prejudice. The perspective of Section 1.2.4.1 assumes that positively-valenced experiences and positive emotions are intertwined, as are negatively-valenced experiences and negative emotions. This section reviews literature regarding emotion, prejudice, valenced contact, and face recognition, and it determines that uncertainty remains regarding whether valenced contact relates to other-race recognition.

1.2.4.1.1 Emotional states and observed emotion

Evidence favouring a positive emotional state improving other-race recognition was found by Johnson and Fredrickson (2005), yet not subsequently (Curby, Johnson, & Tyson, 2012; Gates, 2008). In Johnson and Fredrickson (2005), other-race faces were recognised (in terms of *d*') better after positive emotion induction (by way of a comedic video) than in negative emotion (horror video) or neutral conditions, with the other-race effect occurring in neutral and negative conditions, but not the positive condition.[6] The positive boost for other-race faces has been explained by positive emotion enhancing holistic processing, or encouraging the inclusion of other-race faces into an ingroup (Johnson & Fredrickson, 2005). Amongst own-race faces, holistic processing is greater for ingroup faces (own-university) than outgroup (other-university) (Hugenberg & Corneille, 2009), which is believed to stem from the motivation to individuate differing between ingroup and outgroup faces (Hugenberg et al., 2010), therefore

[6] Interestingly, this occurred not only when videos were presented prior to the learning phase (in the first experiment), but also when videos were displayed solely after learning (second experiment) (Johnson & Fredrickson, 2005).

placing other-race faces within an ingroup may have increased holistic processing.

However, research elsewhere using positive and negative emotion conditions has found recognition (across faces of the own-race and another) to be unaffected by emotional state (Gates, 2008). Additionally, after accuracy with own- and other-race faces were combined due to there being no influence of race and there being no interactions regarding race, a positive emotional state has been found to leave holistic face processing unaffected (d'), whilst a negative emotional state decreased holistic processing, perhaps due to negative states causing a focus on local processing (Curby et al., 2012).

As for observed emotions, happy own- and other-race faces have been recognised better than angry faces (there was a main effect of expression regardless of race, and no interaction between expression and race), which can be explained by angry expressions diverting attentional resources away from encoding identity (Kikutani, 2018).

On the other hand, it could be argued that anger would be expected to bolster other-race recognition (e.g., Ackerman et al., 2006); faces are recognised better when paired with mean behaviours than nice behaviours, which can be explained by observers particularly remembering threatening individuals in order to protect themselves (Kinzler & Shutts, 2008), and it has been noted that anger "implies threatening intent", hence angry individuals can be individuated, remembered, and avoided for protection (Ackerman et al., 2006, p. 837), with the presence of anger reasoned to increase the motivation to individuate other-race faces (Hugenberg et al., 2010). Indeed, angry other-race faces have been recognised better than neutral other-race faces in some experiments (Ackerman et al., 2006; Young & Hugenberg, 2012) yet not in others (Corneille, Hugenberg, & Potter, 2007; Gwinn et al., 2015). On the whole, it seems ambiguous as to whether positive or negative emotions (whether in terms of states or observed emotion) are of aid to other-race recognition, and, by extension, the gaining and maintenance of other-race experience.

1.2.4.1.2 Prejudice

Prejudice has been "generally defined as a negative evaluation or antipathy toward a social group or its members" (Miller, Smith, & Mackie, 2004, p. 221). Positive and negative types of other-race contact have respectively been linked to increased and decreased prejudice for the other-race (Barlow et al., 2012). This conjures the question of whether the link between individuation experience and other-race recognition in Walker and Hewstone (2006) may have operated via positive other-race contact decreasing other-race prejudice. Indeed, it has been noted that an early theory of the other-race effect "was that individuals with less prejudiced racial attitudes would be more *motivated* [emphasis added] to differentiate other-race members, when compared with more prejudiced persons" (Meissner & Brigham, 2001, p. 7).

The evidence appears varied regarding whether prejudice (for other-race persons) is a factor in the other-race effect and the recognition of other-race faces (e.g., Ferguson, Rhodes, Lee, & Sriram, 2001; Lavrakas, Burl, & Mayzner, 1976; Walker & Hewstone, 2008). Concerning implicit prejudice, in Ferguson et al. (2001) persons of high and low implicit racial prejudice had equal magnitudes of the other-race effect, whilst individuals of high implicit prejudice recognised faces better than persons of low implicit prejudice (irrespective of faces being own- or other-race). A positive association occurred between implicit prejudice and the other-race effect in Walker and Hewstone (2008), yet no such relationship happened in Jerovich (2017). As for explicit prejudice, Ferguson et al. (2001) found that individuals of high prejudice and persons of low prejudice exhibited equal magnitudes of the other-race effect. Regarding correlations between the other-race effect and explicit prejudice, an association was not evident in Slone et al. (2000). No relationship between explicit prejudice and other-race recognition was demonstrated in Lavrakas et al. (1976) and Slone et al. (2000), whilst Ferguson et al. (2001) found that people of high explicit prejudice had worse recognition of faces than individuals of low explicit prejudice (regardless of whether faces were own- or other-race).

If prejudice is a determiner of the other-race effect and other-race recognition, it may not solely be the current level of prejudice that matters, but also prejudice when observing faces (and acquiring or keeping expertise) in the past. Indeed, prejudice against other-race persons can be changed situationally, for instance a competitive mentality yields a greater level of explicit prejudice against other-race persons relative to a cooperative mentality (Sassenberg, Moskowitz, Jacoby, & Hansen, 2007), and explicit attitudes towards a minimal outgroup improve when participants believe that they have cooperated with a member of that outgroup on a task, whilst a competition condition does not lead to attitude changes with respect to positivity/negativity (Stiff & Bowen, 2016).

1.2.4.1.3 Valenced contact and other-race recognition

A previous study, Jerovich (2017), had examined whether positive or negative contact with other-race persons is correlated with other-race recognition. One questionnaire item was used to quantify positive contact, and another singular item was employed for measuring negative contact (Jerovich, 2017).

However, the use of single-item (over multi-item) measures can obscure relationships, which may be, in part, due to single-item measures not accounting for within-person variability, yet this variability can be reduced across multiple items (Credé, Harms, Niehorster, & Gaye-Valentine, 2012). Indeed, no correlations were apparent in Jerovich (2017) between positive or negative other-race contact and other-race recognition.

Interestingly, Walker and Hewstone (2006) did not find a correlation between individuating (seemingly positive) experience and face recognition. However, other-race individuating contact predicted other-race recognition in a regression analysis when other-race quantity of contact and own-race recognition were also used as predictors (Walker & Hewstone, 2006). Own-race recognition was included "in order to control for overall face discrimination abilities" (Walker & Hewstone, 2006, p. 470); it would be speculative to suggest that the lack of relationships between other-race contact valences and other-race recognition in Jerovich (2017) occurred due to own-race (or general) recognition prowess being unaccounted for.

However, given the diversity of face recognition aptitude and correlations in face recognition ability from one stimulus race to another (e.g., Wan et al., 2017), an inclusion of own-race (or overall) face recognition in the analysis when exploring the potential relationship between contact valence and other-race recognition may have rendered more detectable the relationships between types of valenced contact and other-race recognition. Additionally, given concerns regarding the number of questionnaire items (Credé et al., 2012), it may be desirable to use a multi-item measure rather than a single-item instrument to count positive or negative contact.[7] Overall, it appears that the valence of contact would seem to be quite an uncharted area in the context of the other-race effect, although results in Walker and Hewstone (2006) may suggest the importance of positive contact.

1.2.5 Perceptual expertise

Experience has its impact on the other-race effect through expertise (Hugenberg et al., 2010) (Figure 1.1). In regard to the other-race effect, previous research has considered three types of expertise (featural, configural, and holistic), for instance, studies have tested whether magnitudes of other-race expertise disadvantages relate to the other-race effect (e.g., DeGutis, Mercado, Wilmer, & Rosenblatt, 2013; Zhao et al., 2014b). It has been suggested that configural sensitivity drives face recognition, (Rhodes, Brake, Taylor, & Tan, 1989), which implies that an other-race configural disadvantage would cause the other-race effect. Given that holistic processing correlates with face recognition accuracy (Caucasian CFMT) (Richler, Cheung, & Gauthier, 2011a), the idea that the other-race effect arises chiefly from superior own-race holistic processing (Rossion & Michel, 2011) seems reasonable. Furthermore, the holistic advantage may bolster featural and configural processing (Rossion & Michel, 2011), and holistic processing does indeed affect featural and configural perception (Hayward, Crookes, Chu, & Favelle, 2016), therefore one might expect each type of expertise to relate to the other-race effect. Section 1.2.5 addresses whether previous research has found other-race expertise disadvantages, and if each

[7] Slone et al. (2001) used a questionnaire to measure other-race experience, which included a subset of several items that inquired into the pleasantness of other-race contact, however, the relationship between the pleasantness subset and other-race recognition (or the other-race effect) was not specifically examined.

disadvantage associates with the other-race effect. This section argues that the impression from previous research is that other-race expertise disadvantages do not have a relationship with the other-race effect, and that this casts doubt over the other-race effect being perceptually-driven.

However, expert *processing* tasks can involve a delay, i.e., between the presentation of a face at *learning* and a presentation at *test* (rather than using simultaneous displays of the face[s]), hence tasks can be thought of as testing memory (e.g., Rhodes, Ewing et al., 2009), although they are taken to show the use of certain types of expert processing (e.g., Zhao et al., 2014b). A prime consideration of the other-race effect as being encoding-driven (from the *learning* phase) (Hugenberg et al., 2010) does suggest that other-race disadvantages in processing memory tasks would largely be conceived as being due to own- vs. other-race differences in expert processing rather than differences arising in memory (storage or retrieval). Nevertheless, whether own- vs other-race disparities in processing tasks arise from processing or memory may be of concern for featural and configural memory tasks (but not for holistic tasks, as they are to do with disruptions and enhancements, see Section 1.2.5.3). Arguably, greater own- than other-race performance in a featural or configural memory task could arise from better own-race expert processing at learning or test phases[8], or superior storage or retrieval of own-race featural/configural information from memory. One could expect that the use of memory in featural and configural tasks would only enhance any correlations between measured other-race disadvantages in featural/configural memory tests and the other-race effect (vs.

[8] Expert processing need not be limited to taking place solely at learning, as results in previous research indicate an occurrence at test (Ho & Pezdek, 2016).

when memory would not be used in a processing task, i.e., in simultaneous presentations); an overall absence of correlations would imply no relationship between other-race featural/configural disadvantages in *processing* and the other-race effect.

1.2.5.1 Featural

There is mixed evidence regarding the other-race featural disadvantage (e.g., Mondloch et al., 2010; Tanaka, Kiefer, & Bukach, 2004), and results in previous research do not favour this disadvantage being a contributing factor to the other-race effect (DeGutis et al., 2013; Zhao et al., 2014b). Featural processing of own- vs. other-race faces has been studied in a number of ways: scrambling the location of features (Hayward et al., 2008), the part-whole task (DeGutis et al., 2013), feature morphing (Rhodes, Hayward, & Winkler, 2006), and the Jane/Ling task (Mondloch et al., 2010).

In the scrambling paradigm, own- and other-race faces are presented with their features in their typical locations at learning, yet the placement of their features is jumbled at test so that memory for features is selectively tested, thereby indicating whether featural processing is greater for own-race faces than it is for other-race faces (e.g., Hayward et al., 2008). This paradigm has been used with Caucasian and East Asian faces alongside Caucasian and East Asian observers (Hayward et al., 2008; Mondloch et al., 2010; Rhodes, Ewing, et al., 2009; Zhao et al., 2014b); other-race featural disadvantages have been apparent in some studies, and of the same size regardless of observer race (Hayward et al., 2008; Rhodes, Ewing, et al., 2009), or been found amongst East Asian observers but not Caucasians (Mondloch et al., 2010), or not being apparent in either race of observer (Zhao et al., 2014b).

Per trial of the part-whole task, a target whole face (Caucasian, or East Asian) is initially presented before an interval, after which participants decide which one of two (test) stimuli are of the same identity as the target (with one of the stimuli matching the identity of the target, and the other being a distractor) (Tanaka, et al., 2004); in part trials, the two stimuli are isolated facial features (just the eyes, the nose, or the mouth), and in whole trials they are faces that differ in terms of singularly the eyes, the nose, or the mouth (Tanaka, et al., 2004). Performance in part trials (i.e., recognition of features in isolation) has been found to be equal for own-race and other-race faces regarding Caucasian observers in one study (DeGutis et al., 2013), but an other-race *advantage* occurred for Caucasian observers in two other studies (Michel, Caldara, & Rossion, 2006; Tanaka et al., 2004), whilst East Asian observers have exhibited an other-race disadvantage in an experiment (Michel, Caldara, & Rossion, 2006) but had the same performance for own- and other-race faces in another (Tanaka et al., 2004).[9]

A variant of the part-whole task was delivered to Caucasian and East Asian observers using own- and other-race faces, wherein the two test stimuli (one on the left, the other on right) were eyes, nose, and mouths but not in their usual

[9] It is assumed that the *Asian* stimuli used in the part-whole task of Tanaka et al. (2004) and some subsequent studies (e.g., DeGutis et al., 2013) are specifically East Asian. This is based on i) the Asian facial identities which were used as backdrops (i.e., to encapsulate the manipulated features) being derived from Koreans, and target features being placed into those identities having the same "race" as them (DeGutis, DeNicola, Zink, McGlinchey, & Milberg, 2011, p. 2507), ii) the appearances of example stimuli (e.g., Crookes, Favelle, & Hayward, 2013; Tanaka et al., 2004), and iii) Hayward, Crookes, and Rhodes (2013) describing the Tanaka et al. (2004) stimuli as Asian, with Hayward et al. (2013) using *Asian* to refer to East Asians in particular; they use the same term (and therefore definition) for Asian participants in Tanaka et al. (2004) and Michel, Caldara, and Rossion (2006).

configural relationship; nonetheless, feature recognition was the same for own- and other-race faces (Zhao et al., 2014b).[10]

As for feature morphing, changes to nose shape and the lightness of the eyebrows and lips have been used, with Caucasian and East Asian faces and observers, and the other-race featural disadvantage (in upright faces) has been present (Rhodes et al., 2006). In the Jane/Ling task, a target face is presented (Caucasian or East Asian), and a same/different decision is made for a subsequent face which is the same as the target or differs in terms of features or configurations; the other-race featural disadvantage occurred for observers who were Caucasian, although not for East Asian observers (Mondloch et al., 2010).

Importantly, a relationship between the other-race effect and the other-race featural disadvantage has not been demonstrated in either of the two studies which have inquired into it (DeGutis et al., 2013; Zhao et al., 2014b). Specifically, the magnitude of the other-race effect (as indexed via Caucasian and East Asian CFMTs) did not relate to the size of other-race featural disadvantages as measured by feature-scrambling at test (Zhao et al., 2014b) and part trials of the part-whole task (DeGutis et al., 2013).

All in all, the presence of the other-race disadvantage regarding features seems mixed across and within paradigms (e.g., Mondloch et al., 2010; Tanaka et al., 2004; Mondloch et al., 2010). Given that previous research has found no relationship between magnitudes of the other-race featural disadvantage and the

---

[10] Zhao et al. (2014b) used faces from the Max Planck Institute Face Database; Asian faces from this database have been described as having been "acquired from Chinese, Taiwanese, Vietnamese, Japanese, and Korean individuals" (Michel et al., 2013, p. 1206). Furthermore, given that Zhao et al. (2014b) used Chinese participants, and used a questionnaire to measure contact with *East Asians*, the stimuli they used would have been East Asian.

other-race effect (DeGutis et al., 2013; Zhao et al., 2014b), differences in featural expertise would not seem to dictate the other-race effect.

1.2.5.2 Configural

Generally, the presence of the own-race configural disadvantage seems more reliable than the featural disadvantage (e.g., Mondloch et al., 2010). The negative effect of inverting faces (i.e., an upside-down facial orientation rather than upright, which impairs recognition) (Yin, 1969), has been construed as reflecting a disruption of configural processing (Diamond & Carey, 1986), and previous research has explored the face inversion effect for own- and other-race faces (e.g., Hancock & Rhodes, 2008; Herzmann, Minor, & Curran, 2018; MacLin, Van Sickler, MacLin, & Li, 2004; Rhodes et al., 1989; Valentine & Bruce, 1986a). The other-race disadvantage in the face inversion effect (i.e., the face inversion effect for own-race faces minus the face inversion effect for other-race faces) has been found to predict the other-race effect (Hancock & Rhodes, 2008). However, whilst face inversion does disrupt configural processing (Goffaux & Rossion, 2007; Le Grand, Mondloch, Maurer, & Brent, 2001; Yovel & Kanwisher, 2004), face inversion has been noted as affecting other types of expertise too (e.g., Maurer et al., 2002), and there are indeed instances of detrimental effects of inversion on featural and holistic processing (e.g., Rossion & Boremanse, 2008; Yovel & Kanwisher, 2004); it has been stated that "the demonstration of an inversion effect by itself does not constitute evidence for a particular type of face processing" (Maurer et al., 2002, p. 258). Therefore, any test of whether own vs. other face inversion effect differences relate to the other-race effect (i.e., Hancock & Rhodes, 2008) would not specifically test whether the other-race configural disadvantage relates to the other-race effect.

The other-race configural disadvantage has been explored more directly by presenting blurred faces at test (non-blurred at learning) (e.g., Hayward et al., 2008), the Jane/Ling task (Mondloch et al., 2010), and morph continua of configural changes (Rhodes et al., 2006). The blurring of faces has been used to remove featural details and consequently tests memory for configurations in particular, thereby indicating the extent of configural processing (Hayward et al., 2008). Such a manipulation has consistently demonstrated an other-race configural disadvantage, as evident in experiments using Caucasian and East Asian observers (Hayward et al., 2008; Mondloch et al., 2010; Zhao et al., 2014b) or solely East Asians (Rhodes, Ewing et al., 2009). Such consistency has not been found regarding the Jane/Ling task (trials where configural alterations were made between learning and test); an other-race configural disadvantage has presented amongst Caucasian observers, yet not East Asian observers (Mondloch et al., 2010). Regarding configural morphing, the other-race configural disadvantage has been evident for Caucasian and East Asian observers with stimuli from those races (Rhodes et al., 2006). One study has tested whether magnitudes of the other-race configural disadvantage correlate with the other-race effect, and this was with the blurring paradigm; no correlation occurred (Zhao et al., 2014b). Therefore, as it stands, it seems that neither other-race disadvantages in featural nor configural expertise determine the other-race effect (DeGutis et al., 2013; Zhao et al., 2014b).

1.2.5.3 Holistic

The paradigms which have been used to investigate the other-race holistic disadvantage are the part-whole task (DeGutis, DeNicola, Zink, McGlinchey, & Milberg, 2011; DeGutis et al., 2013; Michel, Caldara, & Rossion, 2006;

Mondloch et al., 2010; Tanaka et al., 2004), its modified variant (Zhao et al., 2014b), and the composite task (e.g., Bukach et al., 2012; Hayward et al., 2016; Horry, Cheong, & Brewer, 2015; Michel, Rossion, et al. 2006; Zhao et al., 2014b). For the part-whole task, the metric of holistic processing is greater accuracy on whole trials than part trials (i.e., the part-whole effect) (Tanaka et al., 2004). In the composite task (for identity), *same*/*different* decisions are made to face pairs, with participants instructed to make their decisions solely based on the upper or lower half of the face, whilst ignoring the distractor half (e.g., Richler, Cheung, & Gauthier, 2011b).

Broadly, there are two types of composite task, one being the partial design and the other being the complete design, and these designs would seem to diverge in what they measure, as their composite effects (indexes of holistic processing) are not correlated (Richler & Gauthier, 2014). In the partial design, holistic processing is measured as the misalignment of face halves reducing the negative influence from the distractor half on performance (as the face is not presented as a whole); subtracting performance on misaligned trials from aligned trials provides the magnitude of the composite effect (e.g., Michel, Rossion, et al., 2006). For the partial design, within a face pair, distractor halves never match each other, however, in the complete design those matches are displayed (along with mismatches from the partial design), therefore, as illustrated in previous research (Richler & Gauthier, 2014), the complete design features a full crossing of alignment conditions with congruency conditions (congruency, as in between decision and distractor face halves, e.g., *same* decision with two *same* distractor halves). The idea is that holistic processing is reflected in the magnitude to which congruency leads to better performance than incongruency in aligned conditions (due to distractor influence) than misaligned, hence in the complete design the composite effect is the magnitude of the interaction between congruency and alignment (e.g., Horry et al., 2015).

Examinations of the other-race holistic disadvantage using the part-whole task or its modified variant have used Caucasian and East Asian stimulus faces (e.g., Mondloch et al., 2010; Zhao et al., 2014b). Concerning the part-whole task, previous research generally supports the other-race holistic disadvantage occurring for Caucasian observers, with the disadvantage being present in more instances (Crookes et al., 2013; DeGutis et al., 2011, 2013; Michel, Caldara, & Rossion, 2006; Tanaka et al., 2004) than not (Mondloch et al., 2010). For East Asian participants, results appear more constant, as the other-race holistic advantage has not been evident in previous research (Crookes et al., 2013; Mondloch et al., 2010; Tanaka et al., 2004). In the modified part-whole task, the other-race holistic disadvantage (Caucasian and East Asian participants) has not presented (Zhao et al., 2014b).

Regarding the composite face task, in the partial design, the other-race holistic disadvantage occurred amongst Caucasian observers in Michel, Rossion, et al. (2006), yet, in that same study, it did not manifest for East Asian observers or in Mondloch et al. (2010) for their participants (Caucasian and East Asian). However, in the partial design, evidence favours the composite effect being vulnerable to response bias, unlike the complete design (Richler et al., 2011b), and this advantage of the complete design is supported by the magnitude of the composite effect correlating with response bias in the partial, but not the complete, design (Richler & Gauthier, 2014); studies using the complete design are generally absent of the other-race holistic disadvantage (Bukach et al., 2012;

Harrison, Gauthier, Hayward, & Richler, 2014[11]; Hayward et al., 2016; Horry et al., 2015; Zhao et al., 2014b).[12]

Previous research is contrary to the other-race effect relating to the other-race holistic disadvantage (Horry et al., 2015; Michel, Caldara, & Rossion, 2006; Michel, Rossion, et al., 2006; Zhao et al., 2014b). In experiments using the part-whole tasks alongside the usual manners of calculating part-whole effects (i.e., whole-trial accuracy minus part-trial accuracy) and other-race disadvantages in holistic processing or recognition (other-race face performance subtracted from the performance found with own-race faces), relationships between the other-race holistic advantage and the other-race effect have not been evident in any previous study (DeGutis et al., 2013; Michel, Caldara, & Rossion, 2006; Zhao et al., 2014b).

However, DeGutis et al. (2013) also used regression to calculate metrics of the part-whole effect (whole-trial accuracy after controlling for part-trial accuracy) and own-race advantages (own-race performance controlling for other-race), and found that own-race advantages in holistic processing and recognition correlated. However, when such a manner of analysis was used (on different data) in Zhao et al. (2014b), no such association was demonstrated. As for the composite task, there have been no correlations between magnitudes of the other-race holistic disadvantage and the other-race effect, whether the partial design was used (Michel, Rossion, et al., 2006) or the complete design (Horry et al.,

---

[11] From descriptions in Hayward et al. (2013), it can be ascertained that *Asian* (for participants and stimuli) in Harrison et al., (2014) referred to East Asian specifically.

[12] In Harrison et al. (2014), a sequential composite task was used, with a presentation duration manipulation applied to the probe face. A significant *t*-test outcome indicated a disadvantage at the lengthiest duration (800 ms), yet there was no interaction between duration and stimulus race which Harrison et al. (2014) concluded made "this result difficult to interpret" (p. 850).

2015; Zhao et al., 2014b).

All in all, other-race disadvantages within each of the three types of perceptual expertise would not seem to drive the other-race effect, whether that expertise is featural (DeGutis et al., 2013; Zhao et al., 2014b), configural (Zhao et al., 2014b), or holistic (Michel, Caldara, & Rossion, 2006; Michel, Rossion, et al., 2006; Zhao et al., 2014b). This is problematic for accounts of the other-race effect (e.g., Hugenberg et al., 2010; Rossion & Michel, 2011) which construe the other-race effect as reflecting a perceptual difficulty which arises from an other-race expertise deficit (see Figure 1.1). Consequently, the status of the other-race effect as a perceptual problem would seem questionable.

1.2.6 Perceptually-driven: Subjective variability

Subjective homogeneity (between faces) is the extent to which the differences between different faces are perceived to be small. Whilst the other-race effect is theorised to result from other-race subjective homogeneity (i.e., the idea that the other-race effect is perceptually-driven) (Hugenberg et al., 2010; Goldstein, 1979a), the lack of links between other-race expertise disadvantages and the other-race effect (Section 1.2.5) could indicate that other-race subjective homogeneity does not generally produce the other-race effect. Indeed, the other-race effect would appear to have a non-perceptually-driven (e.g., memory-driven) aspect (Marcon, Meissner, Frueh, Susa, & MacLin, 2010);[13] other-race subjective homogeneity might not be so vital after all. Section 1.2.6 considers whether

[13] In Marcon et al. (2010), as the retention interval between learning and test lengthened (10, 400, 1,400, and 2,400 ms), Hispanics showed a greater magnitude of recognition advantage for Hispanic faces over African American faces (Marcon et al, 2010). This supports a memory-driven path in the other-race effect occurring (Marcon et al., 2010) as, overall, Hispanics are largely of non-African ancestry and primarily a mixture of European Caucasian and Native American ancestries (Klimentidis, Miller, & Shriver, 2009).

other-race subjective homogeneity is a reality, and if it may lead to the other-race effect; if there is a causal relationship, the uncertain status of the other-race effect as a perceptually-driven phenomenon (Section 1.2.5) would be remedied.

1.2.6.1 Other-race subjective homogeneity

Concerning experiments on other-race subjective homogeneity (featuring Caucasian participants alongside Caucasian and East Asian stimuli), earlier experiments do not generally support the presence of other-race subjective homogeneity (Goldstein & Chance, 1976, 1978, 1979), whilst more contemporary experiments do (Byatt & Rhodes, 2004; Lorenzino, Caminati, & Caudek, 2018; Proietti, Laurence, Matthews, Zhou, & Mondloch, 2019). A series of experiments by Goldstein and Chance (1976, 1978, 1979), overall, found evidence favouring other-race perceptual homogeneity in only one experiment out of six. They studied other-race subjective homogeneity by way of various paradigms. For instance, in Goldstein and Chance (1976), participants were presented with pairs of same/different faces (i.e., simultaneous presentations of a face pair), and they decided whether a pair of stimuli were really the same face or different faces.

Yet the multiple paradigms across the Goldstein and Chance experiments each may “not measure the same behaviour” (Goldstein & Chance, 1979, p. 113), i.e., other-race subjective homogeneity could be a reality, but not expressed equally across tasks. Nonetheless, more recent studies have consistently demonstrated other-race subjective homogeneity across different paradigms, whether using similarity ratings (Byatt & Rhodes, 2004), same/different decisions to individual identities (Proietti et al., 2019), or faces on morph continua (Lorenzino et al., 2018).

Other-race subjective homogeneity might not take place across all observer races. Experiment 2 of Crookes et al. (2015) featured both of their stimulus races in their participant groups (East Asian, and Caucasian) and found that the magnitude of other-race subjective homogeneity was lower for East Asian observers than Caucasian observers. The data compared were the *numerical* sizes of other-race subjective homogeneity (i.e., other-race performance subtracted from own-race performance) which allowed magnitudes to be compared, however, inferential tests were not presented as to whether other-race subjective homogeneity occurred in either race. Nevertheless, the greater magnitude for Caucasian participants (in Australia) than East Asian participants (in Hong Kong, China) (Crookes et al., 2015) would indicate that other-race subjective homogeneity did happen for Caucasians, whilst the descriptive statistics presented in Crookes et al. (2015) may suggest that other-race subjective homogeneity might not have been present for East Asian participants, for whom own-race and Caucasian faces may appear to have been of equal perceptually variability.

1.2.6.2 The other-race effect

For theories of the other-race effect, finding the mere presence/absence of other-race subjective homogeneity is less important than discovering whether other-race subjective homogeneity relates to the other-race effect. In Byatt and Rhodes (2004), who used Caucasian participants solely and own-race and East Asian faces, other-race subjective homogeneity occurred, as derived from the application of multidimensional scaling to similarity ratings (i.e., to measure subjective variability) which also predicted identity recognition performance and that own-race faces would be recognised more accurately than other-race faces.

This would *indicate* a link between other-race subjective homogeneity and the other-race effect. Nevertheless, examples of stimuli presented in Figure 2 of Byatt and Rhodes (2004) portrayed East Asian faces as being lit differently than Caucasian faces. The light source seems to have been closer (or brighter) for East Asian faces, such that noses cast noticeable shadows across the central face, and the upper face was otherwise particularly highlighted. This striking lighting for East Asian faces may have contributed to their subjective homogeneity, and influenced predictions. Furthermore, Byatt and Rhodes (2004) did not *directly* test whether other-race subjective homogeneity and the other-race effect are related (although it would seem highly likely for Caucasian observers).

Outcomes in Hancock and Rhodes (2008) may be somewhat suggestive of the other-race effect being perceptually-driven for both Caucasians and East Asians. In Hancock and Rhodes (2008), the other-race disadvantage in the face inversion effect predicted the other-race effect, and the strength of this relationship did not differ between Caucasian and East Asian observers; as face inversion does not uniquely affect any one of the expertise types of interest (e.g., Yovel & Kanwisher, 2004), and other-race disadvantages in either featural, configural, or holistic expertise do not generally relate to the other-race effect (Section 1.2.5), this could suggest that a combination of other-race expertise disadvantages determines the other-race effect for both races, and the results of Hancock and Rhodes (2008) would consequently support the existence of a perceptually-driven route to the other-race effect for Caucasian and East Asian observers. Although Hancock and Rhodes (2008) did not examine other-race subjective homogeneity, their outcomes would be congruent with other-race subjective homogeneity causing the other-race effect for Caucasians and East

Asians alike.

Still, despite a sizeable amount of research on other-race subjective homogeneity (Byatt & Rhodes, 2004; Crookes et al., 2015; Goldstein & Chance, 1976, 1978, 1979; Lorenzino et al., 2018; Proietti et al., 2019), no study had directly examined whether other-race subjective homogeneity has a link to the other-race effect. This leaves uncertainty regarding whether a route does exist from other-race subjective homogeneity to the other-race effect, which is critical for understandings of why the other-race effect occurs (see Figure 1.1).

1.3 Gender categorisation

Is the other-race effect a subset of more general issues with other-race (compared to own-race) faces? This question can be answered by seeing whether difficulties with other- vs. own-race faces occur in arenas outside of individuation, such as in gender categorisation (Zhao & Bentin, 2008; Zhao & Hayward, 2010).[14] Section 1.3 affirms that it is not clear which factors drive the other-race gender effect: races differing in the extent of facial dissimilarity between males and females (Zhao & Bentin, 2008), experience, bias, or expert processing.

1.3.1 Gender/sexual dimorphism

The other-race gender/sex effect has previously been explored in three studies (O'Toole, et al., 1996; Zhao & Bentin, 2008; Zhao & Hayward, 2010). Each presented East Asian and Caucasian faces, with O'Toole et al. (1996) and

---

[14] Nevertheless, the processing of individual identity and gender seem intertwined, such that searches for a target identity within an array are facilitated (in terms of accuracy and response latency) when the target is of a different gender than distractors (Zhao & Hayward, 2013). This overlap between identity and gender need not mean that the experiential or/and motivational elements of the other-race effect (Hancock & Rhodes, 2008; Rossion & Michel, 2011; Young & Hugenberg, 2012) are also present in the other-race gender effect.

Zhao and Bentin (2008) featuring East Asian and Caucasian observers, whilst Zhao and Hayward (2010) included East Asian observers alone. In O'Toole et al. (1996), overall, Caucasian faces were categorised more accurately (male, female sex categorisations). Additionally, they found that observer race and stimulus race interacted, which, in tandem with mean *d*'s (presented in text on p. 672 of O'Toole et al., 1996, rather than the adjusted ones presented in a figure on that page), demonstrated that the other-race sex effect occurred for Caucasian observers, but not for East Asian observers; indeed, regarding East Asian observers, *d*' was numerically greater for other-race faces (O'Toole et al., 1996). As for Zhao and Bentin (2008), gender was categorised more accurately in Caucasian faces than East Asian faces across both observer groups. This would suggest that there is "probably" a greater physical male-female difference in Caucasian faces than in Chinese East Asian faces (Zhao & Bentin, 2008, p. 1098).

Evidence favouring higher (objective) sexual dimorphism in the faces of Caucasians than East Asians has been found outside of accuracy, with multidimensional scaling of similarity ratings (Caucasian and *Asian* participants) suggesting a larger perceptual dissimilarity between males and females amidst Caucasian faces than the faces of East Asians, which may stem from objective similarity (Hopper, Finklea, Winkielman, & Huber, 2014).

Nonetheless, in Zhao and Hayward (2010) (for intact, upright faces), East Asian observers were as accurate in their gender categorisations of own-race faces as they were for Caucasian faces. Interestingly, the other-race gender effect had not been studied outside of the East-Asian/Caucasian dynamic; it could be that, compared to yet another race, facial dimorphism is lower for East Asian faces or/and higher amongst Caucasian faces.[15]

1.3.2 Experience and bias

The role of experience in the other-race gender effect is an open topic. It has been suggested that races may differ concerning which facial traits are the most useful for gender categorisation (Yamaguchi, et al., 1995). Racial differences (comparing European Americans to African Americans) in the magnitude of sexual dimorphism have been observed in four out of 19 morphological traits of the skull (Kittoe, 2013), which implies that there is variation between races in the optimality of dimension weightings in gender categorisation. If there are such differences, just as deploying own-race dimension weightings to other-race faces may cause the other-race effect (Hills et al., 2013; Hills & Lewis, 2011; Valentine, 1991), applying optimal own-race dimensions for discriminating gender on to other-race faces could result in the other-race gender effect (if those weightings are less useful for other-race faces).

Yet there may be another route for experience to have an influence. Given that races have morphological differences (Farkas et al., 2005; Porter, 2004; Porter & Olson, 2001) and dimorphism with respect to males and females (e.g., Samal et al., 2007; Steyn & İşcan, 1998), to an observer, faces of a race may seem generally more male (or female) than faces of another race (Johnson, Freeman, & Pauker, 2012). In the context of own- and other-race faces, for an

[15] In Yao (2014), comparisons of the extent of facial sexual dimorphism were undertaken between population groups, and several of those comparisons were interracial. Bootstrapping resulted in 10,000 data points per population group being used in the analyses of interest (e.g., in an ANOVA regarding differences between males and females) (Yao, 2014). Unequal magnitudes of sexual dimorphism were found between populations, including interracially (Yao, 2014), however, these significant outcomes could have resulted from using a very large amount of data points in analyses.

observer from Race A, Race B faces may be closer to Race A males than Race A females, therefore causing a bias (relative to Race A, an other-race gender categorisation bias) to categorise Race B faces as male (greater experience with Race A would lead to Race A informing gender norms more than Race B faces would).

Another factor which may influence bias and categorisation performance are associations (which observers may hold) between race and gender/sex, for example *Black* and *male* (Goff, Thomas, & Jackson, 2008; Johnson et al., 2012). In Experiment 1 of Goff et al. (2008), there were 169 participants (over 80% of whom were Caucasian); 58 of those participants categorised the gender of Black and Caucasian faces, and rated those faces for masculinity/femininity. Compared to Caucasian faces, Blacks were perceived as more masculine (i.e., a relative bias) and gender was categorised less accurately (Goff et al., 2008).[16] More specifically, categorisation was less accurate for Black women than for Black men, Caucasian women (Goff et al., 2008), and (as can be assumed from other comparisons and Figure 2 of Goff et al., 2008) Caucasian men. Goff et al. (2008) do note that experience may change race-gender associations. Hence one may expect that gaining other-race experience (thereby reducing the other-race experience disadvantage) would lower the other-race gender categorisation bias, and minimise the other-race gender effect as a result.

1.3.3 Expertise

As for expertise, potential other-race disadvantages in holistic and featural processing of gender have been examined in one experiment (Zhao & Hayward,

[16] The sub-sample of 58 would likely have largely been Caucasian, therefore results would tentatively support an occurrence of the other-race gender effect.

2010). Gender is processed holistically (Baudouin & Humphreys, 2006; Murphy & Cook, 2017; Zhao & Hayward, 2010). Interestingly, matching extents of holistic processing of gender occur for own- and other-race faces (Zhao & Hayward, 2010). Arising from experience, holistic processing is posited to influence the other-race effect (e.g., Rossion & Michel, 2011); perhaps a similar experience-expertise route may occur regarding the other-race gender effect even if dimorphism is greater amidst Caucasian faces. As for featural processing, features do contain gender-differentiating information (Brown & Perrett, 1993), and an other-race featural disadvantage for gender has been evident (using scrambled faces) (Zhao & Hayward, 2010).

Given the holistic and featural processing of gender (e.g., Zhao & Hayward, 2010), experience could inform expert processing, hence a lesser other-race experience disadvantage could reduce the size of the other-race expertise disadvantage in gender processing. In turn, a smaller expertise disadvantage may reduce the other-race gender effect. However, prior research had not examined relationships between other- vs. own-race differences in experience (as indexed by a questionnaire for instance), expert processing of gender, and gender discrimination. Additionally, the link between the other-race gender categorisation bias and both experience and the other-race gender effect had not been examined.

1.4 Race categorisation

Whilst race can affect gender categorisation (Section 1.4), gender itself may be a determiner of race categorisation (e.g., Carpinella, Chen, Hamilton, & Johnson, 2015). Experiments on race categorisation often involve presentations of monoracial faces (Ge et al., 2009: Zhao & Bentin, 2011), or facial morph continua populated with phenotypes which range from one race category to another (Freeman et al., 2011; Krosch & Amodio, 2014). For instance, morph continua have been used for studying race category boundaries (e.g., Webster, Kaping, Mizokami, & Duhamel, 2004). This type of boundary can be conceptualised as a point along a racial continuum between two race norms (e.g., Race 1, and Race 2) where a face subjectively seems to equally belong to Race 1 as much as Race 2, i.e., a point of subjective equality (PSE) (e.g., Krosch & Amodio, 2014). Additionally, race categorisation has been explored through response latencies, which have been employed as a measure of categorisation proficiency (e.g., Carpinella et al., 2015). Section 1.5.1 considers whether gender affects race category boundaries, whilst Section 1.5.2 focuses on the potential effect of gender on the precision with which race is categorised.

1.4.1 Boundary

The race category boundary (i.e., race PSE) is construed as being a product of experience (Webster et al., 2004), such that relatively greater own-race experience reduces sensitivity to own-race-specifying traits in comparison to other-race-specifying traits (as in perceptual adaptation), hence shifting the PSE (on the objective scale of the racial morph continuum) towards the own-race category, with other-race traits being more perceptually salient (Benton & Skinner, 2015; Webster & MacLeod, 2011). Therefore, regarding two races (Races A and B for instance), faces which are objectively halfway between the Race A norm and the Race B norm (or midway between unambiguous members of those races) would be perceived more as Race B by Race A persons than by persons of Race B (Halberstadt, Sherman, & Sherman, 2011; Webster et al., 2004) (Figure 1.3). The race category boundary is not only related to experience;

it has been found to be influenced by economic sparseness (Krosch & Amodio, 2014). A further factor which may affect the race category boundary (regarding faces) could be gender.

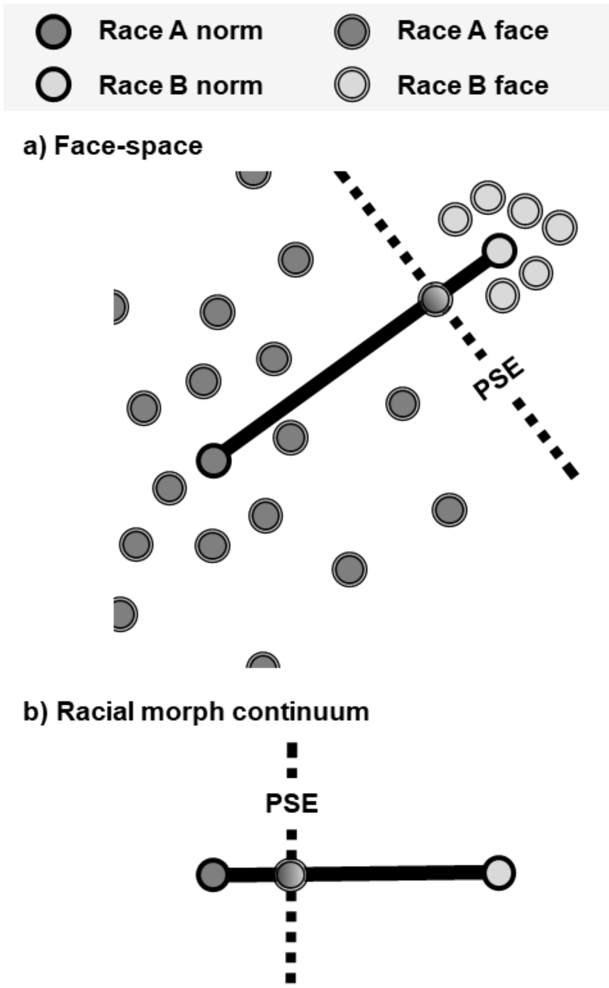


Figure 1.3. For Race A observers: a) due to face-space dimension weightings being more optimal for Race A faces, in terms of face-space the PSE would be closer to the Race B norm than the Race A norm; b) as for the location of the PSE as measured on the objective scale of the race morph continuum, the PSE would be nearer the own-race norm (the Race A norm). Amongst Race B observers, PSE locations would be the opposite of the locations found with Race A observers.

Results in Davenport (2016) may hint at this. A type of race categorisation is self-categorisation, and this can be affected by a variety of factors including gender (Davenport, 2016). This was demonstrated in Davenport (2016) with university students in America who had one parent who was categorised White, and another who was categorised Asian, Black, or Latino; amongst persons of White/Asian, White/Black, and White/Latino ancestry, females were more likely than males to self-categorise as multiracial (compared to Asian, Black, or Latino respectively) (Davenport, 2016). This gender difference may possibly arise from the perception of observers (Davenport, 2016). Indeed, as noted by Davenport (in Wallace, 2016, para. 8), "the different ways that biracial people are viewed by others influences how they see themselves". Therefore, an effect of gender on the race category boundary of faces could be expected.

Previous experiments have explored possible effects of gender on the racial categorisation of others (Carpinella et al., 2015; Ho, Sidanius, Levin, & Banaji, 2011). In Experiment 2A of Ho et al. (2011), family trees (not faces) were presented which consisted of four grandparents and their grandson or granddaughter. Participants (Caucasians who were in the United States) were provided with the *race* of each of the four grandparents. Each grandparent was labelled as Asian, Black, or White. In a presentation, grandparents could all be from one race, or from different ones, i.e., two from a race alongside two of another race, or three from a race and one of another (although there were no Asian and Black combinations). Participants were told whether the grandchild was a grandson (blue text) or a granddaughter (pink text), and symbols were also used to signify gender. Race categorisation decisions pertained to which race the grandchild was; the gender of the grandchild did not affect race categorisation for biracial grandchildren (who were of equal ancestry from two races) (Ho et al., 2011).

Another experiment in Ho et al. (2011) (Experiment 2B, which also featured solely Caucasian participants, with testing occurring once more in America) presented faces for grandparents (Asian, Black, or White) rather than race labels, and the symbols for gender were made larger to make gender more salient. Regarding the categorisation of biracials with half their ancestry being White and the other half being either Asian or Black, results indicated that hypodescent (categorising as Asian or Black more than White) was more prevalent for male faces than female faces (Ho et al., 2011). Therefore, in terms of race norms, the race category boundary on a morph continuum (Figure 1.3b) may be nearer to the Caucasian norm for males than females. Still, across these two experiments of Ho et al. (2011), it would seem particularly unclear if gender does affect race categorisation as an influence of gender was indicated in one experiment yet not another.

As for the racial categorisation of faces themselves, in Studies 1 and 4 of Carpinella et al. (2015) gender affected the categorisation of racially ambiguous FaceGen faces (which were set to be halfway between one race and another, and were on continua from highly masculine to highly feminine) within types of biracial ancestries. For instance, regarding Caucasian/Black faces, increasing femininity heightened the chance of categorising faces as Caucasian and lowered the likelihood of Black categorisations, and, concerning Caucasian/East-Asian faces, femininity also facilitated the probability of Caucasian categorisations whilst reducing the chance of Asian categorisations (Carpinella et al., 2015). These outcomes imply that gender would affect race PSEs, with female faces being perceived as more Caucasian (and less monoracial non-Caucasian) than male faces.

The occurrence of femininity increasing the Caucasian categorisation of biracial faces (Carpinella et al., 2015) would suggest that the greater propensity of biracial women to self-categorise as multiracial than biracial men in the USA (Davenport, 2016) may be (in part) due to biracial women indeed being perceived as more White and less Asian/Black/Latino than the faces of biracial men. Nonetheless, some racial differences *within* FaceGen faces and *amongst* real faces have been described as not being the same, such as in regard to chin length and the flatness of the nasal bones (Holland, 2009). Furthermore, race categorisation can differ between FaceGen and real faces, as the prevalence of Black over White (hypodescent) categorisations of Black/Caucasian biracial faces can be less apparent with FaceGen faces than real faces (Gaither, Chen, Pauker, & Sommers, 2019). Consequently, it should not be definitively stated that the aforementioned gender influences on race categorisation in Carpinella et al. (2015) apply to the racial categorisation of real faces.

Interestingly, it could be that effects of gender on race category boundaries are not the same for different ancestries. In terms of self-categorisation, in Davenport (2016) for White/Asians, males were of a greater likelihood than females to self-categorise as White (than Asian). However, amongst White/Latinos, and White/Blacks, males and females did not differ in their likelihood of White self-categorisations (compared self-categorisations of Latino for White/Latinos, and Black for White/Blacks).

1.4.2 Proficiency

Effects of gender on race categorisation proficiency have been studied in a few experiments, by way of response latencies and accuracy (e.g., Carpinella et al., 2015; Li & Tse, 2016; Li, Tse, & Sun, 2018). Response latencies have been construed as representing the norm-to-exemplar distance, with a longer latency representing a greater distance (Valentine & Bruce, 1986b). Therefore, in the context of face-space, faster race categorisation responses to a gender could suggest that the race-norm is nearer to exemplars of that gender. Regarding Caucasian and East Asian faces, in Ge et al. (2009) there was no effect of facial gender on race categorisation. However, in Experiment 1A of Li et al. (2018), who also used Caucasian and East Asian faces, race categorisations were faster for male faces than female faces. (A stimulus race and stimulus gender interaction was explored in Li et al., 2018 Experiment 1A in the context of the other-race categorisation advantage [race categorisations being faster for other- than own-race faces] for male faces and for female faces rather than concerning whether effects of gender differed *within* races.)

Gender effects within races have been analysed by Carpinella et al. (2015), Li and Tse (2016), and Thomas, Dovidio, and West (2014), who discovered shorter (correct) latencies for males than females amongst Black faces, and shorter latencies for females than males for Caucasian faces (Carpinella et al., 2015; Li & Tse, 2016; Thomas et al., 2014), yet there was no gender disparity in response latencies for East Asian faces (Carpinella et al., 2015; Li & Tse, 2016).[17]

Influences of gender on race categorisation proficiency outside of response

[17] As for why race categorisation proficiency may be affected by gender, measures of the racial prototypicality of faces (via FaceGen metrics) may suggest an objective route, as they have shown that males are more prototypical than females within faces of Blacks, females have greater prototypicality than males amongst Caucasians, and that there is no racial prototypicality difference between the genders for East Asian faces (Carpinella et al., 2015).

latencies have not been examined often. Indeed, race categorisation has almost been at ceiling in some prior research on gender and race categorisation latencies (Ge et al., 2009; Li et al., 2018, Experiment 1A).[18] Regarding accuracy, in Li and Tse (2016), where there was no main effect of gender, and, although an interaction between gender (male, female) and race (Black, Caucasian, East Asian) occurred, further relevant tests (e.g., on gender differences within races) were not presented. Hence, the field seems quite open concerning gender effects on the facial race categorisation of others, and not only with respect to the race category boundary, but also race categorisation proficiency.

1.5 Outline

In Sections 1.2 to 1.4, gaps in the literature were identified; Section 1.5 presents an overview of the objectives underlying the six psychology experiments and the analysis of pre-existing anthropometric and psychological data which constitute Chapters 2-5. Sections 1.5.1 and 1.5.2 pertain to the topic of *identity recognition*, whilst Sections 1.5.3 and 1.5.4 are in regard to *gender and race categorisation*.

1.5.1 Experience and identity recognition

Section 1.2.3 argued that the importance of childhood timespans for the other-race effect during adulthood remained unclear, and Section 1.2.4.1 demonstrated that previous research had rarely made explorations concerning whether the valence of contact relates to the recognition of other-race faces.

[18] In Ge et al. (2009), the focus was on the other-race effect and the other-race categorisation advantage, nonetheless, as "[p]reliminary analyses showed that the effects of participant gender and face gender were not significant" (p. 1202), it can be inferred that there was no difference between genders (of face stimuli) in terms of race categorisation accuracy. As for Li et al. (2018) Experiment 1A, males and females were categorised with equal accuracy.

Chapter 2 intended to address this topic. Participants were of various racial backgrounds. A questionnaire was administered which tallied experience with three race categories (Caucasian, Black, and East Asian) during the timespans of 0-6, 6-12, and 12-18 years-of-age (i.e., the life stages used in Cloutier et al., 2014), and items were featured which were regardless of the valence of contact (which are referred to as *non-valenced* items) or pertained to valence (positive or negative). Face recognition was measured in a delayed sequential face-matching task which featured FaceGen faces of Caucasians, Africans, and East Asians.

In Experiment 1, the relationship between experience and face recognition ability was explored for Caucasian observers by seeing whether racial differences (i.e., own-race minus an other-race) in experience (non-valenced) at each of the childhood timespans correlated with magnitudes of the adulthood other-race effect. Experiment 2 assessed whether other-race contact (positively or negatively valenced) during childhood predicted accuracy for recognising other-race faces during adulthood.

1.5.2 Facial variability

Whether races are of equal structural diversity (pertaining to the face) is equivocal (Section 1.2.2), and it is uncertain if the other-race effect has a perceptually-driven contribution (Section 1.2.6). Chapter 3 attempted to bring some resolution to these topics. Regarding structural variability, the chapter used three datasets (i.e., Gordon et al., 2012, 2014; Howells, 1996), none of which were employed in Goldstein (1979a) or Phipps et al. (1988), and the methods of statistical analysis applied in Chapter 3 were different to the ones used in either paper. Whereas Goldstein (1979a) did not consistently compare the variability of Africans to any non-African group, Chapter 3 did. Previous research had shown that the morphological diversity (mean variance) of the skull diminished as modern humans moved farther away from their origin, with that origin being within Africa (e.g., Manica et al., 2007); Chapter 3 tested whether such a decline occurred in the skeletal structure of the face. This decline was examined in terms of types of variability (including the mean variance), and between-attribute correlations were also explored. Using measurements of the face in its everyday state, Chapter 3 also sought to determine if Blacks, Caucasians, and Native Americans are of equal structural variability. Lastly, data from Bate, Bennetts et al. (2018b) was analysed for the purpose of exploring whether the extent of other-race subjective homogeneity predicts the magnitude of the other-race effect for Caucasian and East Asian observers.

1.5.3 Gender categorisation

Section 1.3 stated that previous research had not tested whether there are associations between the other-race experience disadvantage, the other-race expertise disadvantage for gender, the other-race gender categorisation bias, and the other-race gender effect. Chapter 4 (i.e., Experiment 3) sought to cover these themes by exploring the gender categorisation of East Asian, Caucasian, and Black participants, specifically to determine i) whether the other-race gender effect relates to the other-race experience disadvantage, other-race expertise disadvantages in gender processing, and the other-race gender categorisation bias, and ii) if the other-race experience disadvantage has associations with expertise disadvantages and the other-race gender categorisation bias.

Experience was measured with the same questionnaire that was employed in Chapter 2, with analysis focussing solely on non-valenced contact. The aperture paradigm of Murphy and Cook (2017) was used as the metric of the holistic processing of gender, and the gender processing of local facial areas. In this paradigm, compared to observing all of the face at once, viewing the face through an aperture (which crosses the face) reduces the recognition of the identity, emotion, age, and gender of faces, which suggests that the aperture condition disrupts the holistic processing of each of those types of facial information (Murphy & Cook, 2017). Chapter 4 adapted the aperture task; participants categorised the gender of Caucasian, Black, and East Asians faces which were from continua between a race norm and individual identities (in a gender categorisation task, Murphy & Cook, 2017, solely presented Caucasian faces). The other-race gender effect and other-race gender categorisation bias were measured in a gender categorisation task which featured faces of the three aforementioned groups drawn from male-female morph continua.

1.5.4 Race categorisation and gender

All in all, it is unclear i) if gender affects the perceptual race categorisation of others (race category boundary and race categorisation precision) and ii) whether the gender effect on categorisation is the same for different types of biracial ancestries (Section 1.4). Gender effects could be tied to gender categorisation ability; the less an observer can perceive gender differences, the smaller the difference between genders in race categorisation may be. Moreover, it was untested whether the effect of gender is influenced by the orientation in which a face is presented, and if facial luminance plays a role in the effect of gender. In Chapter 5, these points were explored across three experiments. Experiment 4 consisted of two race categorisation tasks (Caucasian-to-Black and Caucasian-to-East-Asian continua) formed of male and female upright faces, and a facial gender categorisation task (upright Caucasian male-to-female faces), whilst Experiment 5 featured one race categorisation activity (Caucasian-to-Black continua) in which the visual orientation of male faces and female faces was manipulated, so that faces were presented upright and upside-down, and the race categorisation task of Experiment 6 manipulated the luminance of Caucasian-to-Black faces so that the possible contribution of luminance to the gender difference in the Caucasian/Black category boundary could be tested.

Chapter 2: EXPERIENCE AND IDENTITY RECOGNITION

2.1 Introduction

Clarity is required regarding whether other-race experience disadvantages in childhood life stages relate to the other-race effect (Section 1.2.3.3). Furthermore, it is quite uncertain if the valence of other-race contact during childhood is important for other-race recognition; the items in the Walker and Hewstone (2006) individuation experience questionnaire would appear to have generally been positively valenced, which lead to the question of whether the valence of contact matters (Section 1.2.4.1). As stated in Section 1.2.4.1, Jerovich (2017) used single-item scales for other-race positive contact, and for the negative variant, and did not control for face recognition ability, and these factors could have contributed towards correlations between types of contact and other-race recognition not being evident.

Both experiments featured in Chapter 2 (Experiments 1 and 2) employed the same questionnaires and face recognition task. They addressed topics raised in Sections 1.2.3.3 and 1.2.4.1. Experiment 1 explored whether the childhood other-race experience disadvantage relates to the other-race effect amongst adults. The topic of Experiment 2 was the relationship between other-race valenced contact during childhood (using multi-item scales) and recognition accuracy for other-race faces in adulthood, when controlling for a number of variables including a representative of face recognition prowess (see Section 2.2.2.1).

2.2 General method

2.2.1 Participants

There were 113 participants (37 males, 76 females; $M_{\text{age}} = 25.29$, $SD_{\text{age}} = 8.68$). Due to changes in face recognition ability with age (Germine, Duchaine, & Nakayama, 2011), an age limit of 18-to-50-years-old was used. Whilst a subset of participants, being 52 Caucasians (18 males, and 34 females; $M_{\text{age}} = 27.88$, $SD_{\text{age}} = 10.35$), constituted the participants of Experiment 1, all 113 were included in Experiment 2. Each experimental session occurred at City, University of London, wherein ethics approval was given.

2.2.2 Materials

2.2.2.1 Questionnaires

Three questionnaires were employed across the experiments. One was the 20-item prosopagnosia index (PI20), which measures the impressions which observers have of their personal face recognition prowess (Shah, Gaule, Sowden, Bird, & Cook, 2015). PI20 scores correlate with face recognition ability (Shah et al., 2015), and the questionnaire has been used as part of a set of tests for detecting developmental prosopagnosia (Biotti & Cook, 2016; Biotti, Gray, & Cook, 2017). Accordingly, in Experiment 2, PI20 scores were used as a substitute measure of general face recognition ability. Correlations between the PI20 and face recognition indicate that PI20 scores of 65 upward suggest the presence of developmental prosopagnosia (Shah et al., 2015). Therefore, any Caucasians who scored 65 or in excess were removed from analyses, although non-Caucasians with such scores were not removed. The 2011 census shows that the majority of the population of the UK in England is Caucasian (Office for National Statistics, n.d.). Experiments 1 and 2 occurred in England, and PI20 items do not specify race, therefore non-Caucasians might respond to PI20 items with their recognition of Caucasian faces in mind to a considerable extent, whilst Caucasians would largely be focussed on own-race recognition. Given the other-race effect (e.g., Zhao et al., 2014b), in England one would expect non-Caucasians to struggle with face recognition (in general) more than Caucasians, and therefore have higher PI20 scores compared to Caucasians; a score of 65 or more for non-Caucasians may be due to difficulties with Caucasian faces (in the absence of problems with own-race faces) rather than developmental prosopagnosia. PI20 scores were not used as a basis for removal in any instance where it was unknown whether participants were Caucasian.

Another questionnaire was the Inter-Racial Contact Questionnaire (IRCQ), which was used to gather the age, gender, nationality and race/ethnicity of participants, and tally their contact (non-valenced [i.e., without respect to valence] and valenced) with Caucasians, Blacks, and East Asians when the participant was between the age-ranges (years) which were applied in Cloutier et al. (2014): 0-6, 6-12, and 12-18. There was concern over participants not applying the term *African* across Africans regarding faces which they have seen, for instance, under the “Black / African / Caribbean / Black British” heading, the UK Census has separate boxes for “African”, for “Caribbean”, and for “[a]ny other Black / African / Caribbean background” (Office for National Statistics, 2011, p. 8). Because of this, *Black* was used. In the questionnaire, race/ethnicity categories for self-categorisation were selected on the basis of U.S. Census categories (United States Census Bureau, 2017) and recommendations stated by the Office for National Statistics in the UK (Office for National Statistics, n.d.).

For each contact item, participants responded with an integer from 1 to 7, with 1 standing for disagree strongly, and 7 meaning agree strongly. The first block was composed of non-valenced items, and the second block featured the positively and negatively valenced items. Items were not limited to face-to-face encounters as effects on face recognition by experience outside of in-person situations has been evident, such as via faces appearing in books (Heron-Delaney et al., 2011) or on screens during individuation training (Hills & Lewis, 2006), and outcomes in previous research have indicated that exposure to other-race faces in the media (television and film) increases the use of an other-race (vs. a more general) norm when encoding other-race faces (Wang & Zhou, 2016).

The IRCQ featured 12 items, six of which were non-valenced items (Table 2.1), whilst the other six were valenced (three positively, and three negatively) (displayed in Table 2.2). The first, second, and third items of the non-valenced block were derivatives of items from Hancock and Rhodes (2008). In the valenced block, the second item was adapted from an item which was featured in the Walker and Hewstone (2006) individuating experience questionnaire ("[a] South Asian … person has comforted me when I have been feeling sad", Walker & Hewstone, 2006, p. 468), and the fifth item in the IRCQ was partially obtained from another item of the Walker and Hewstone questionnaire ("I have asked a South Asian person to be on my team or in my group during sports or activities", Walker & Hewstone, 2006, p. 468).

Table 2.1

Non-Valenced Items in the IRCQ

Item	Text
1	Most days, I encountered peers with (race) faces in educational or social contexts
2	In my local community, many people were (race)
3	Most days, I had face-to-face interactions with (race) people
4	I saw many (race) individuals in TV shows, films, and online videos
5	I saw many (race) individuals in printed media (e.g., newspapers, magazines, books)
6	Many of the characters depicted in the advertising materials I was exposed were (race)

Table 2.2

Valenced IRCQ Items

Item	Valence	Text
1	Negative	I was teased or bullied by (race) persons
2	Positive	I was comforted by (race) persons when upset
3	Positive	In the popular media or in my local community, I often encountered (race) people who were good role models (e.g., teachers, doctors, sporting stars)
4	Negative	In the popular media or in my local community, I often encountered (race) people who were bad role models (e.g., bullies, criminals, sporting cheats)
5	Positive	I often collaborated with (race) peers in group play or team sports
6	Negative	I often competed against (race) peers in group play or team sports

Items 1, 4, and 6 were designed to reflect negative valence, whilst Items 2, 3, and 5 represented positive valence. Following Section 1.2.4.1, the negatively-valence items were constructed particularly with respect to threat and prejudice (when contact occurred, rather than at present). Item 1 concerned bullying, and Item 2 pertained towards negative role models, i.e., both threatening and harmful. Given that negative other-race experience positively relates to increased explicit prejudice for that race (whilst positive other-race experience has a negative relationship to prejudice) (Barlow et al., 2012), these sorts of experiences may also enhance prejudice. Regarding Item 6 (intergroup competition), compared to an induced cooperative mentality, a competitive state enhances explicit prejudice towards other-race persons (Sassenberg et al., 2007). Positive items were designed to pair with, and oppose, negative items (i.e., Item 2 paired with 1, 3 with 4, and 5 with 6).

The third questionnaire was the Internal and External Motivation to Respond Without Prejudice Scales (IMS and EMS; Plant & Devine, 1998). Plant and Devine (1998) thought of "the internal motivation to respond without prejudice as resulting from internalized and personally important nonprejudiced standards", whilst the external motivation arose "from social pressure to comply with nonprejudiced norms" (p. 813). A desire to seem unprejudiced could cause participants to lower and increase responses on negative and positive IRCQ items respectively; these adjustments may be greater for persons high on IMS/EMS.[19]

---

[19] In the USA, Whites are concerned about being thought to be racially prejudiced, such that they engage their cognitive resources in interracial situations to prevent their behaviour from seeming to be prejudiced, whilst other groups are worried about matching racial stereotypes and being the recipients of prejudice, and they use cognitive resources to avoid those outcomes (Richeson & Shelton, 2003, 2007).

To take a step to account for this, when multiple regression was used to assess whether the valence of contact predicted other-race recognition, IMS and EMS scores (both centred) and their interaction were included as predictors.

IMS and EMS items were originally used on Caucasian respondents, and the items inquired into their prejudice motivations regarding Blacks (Plant & Devine, 1998). Items have previously been adapted concerning prejudice motivations towards Arabs (Fehr & Sassenberg, 2010). In the current experiments, participants were recruited from many races, including ones not reflected in the face stimuli. IMS and EMS items were not tailored towards any particular race of respondent. It was reasoned that participants may find it particularly odd to answer questions on their motivations to respond in a racially non-prejudiced fashion towards their own race, and accordingly, their own-race responses could be considerably noisy, and disrupt responses to the two other categories of interest as participants might attempt to give largely uniform responses from one race to another in order to not seem *unequally* prejudiced/non-prejudiced. As a compromise, IMS and EMS responses were asked for *other-race* categories in general rather than any specific race.

2.2.2.2 Face-matching task

FaceGen stimuli have previously been utilised when studying the other-race effect (Chang, Murray, & Yassa, 2015; Matheson & McMullen, 2011; Pauker et al., 2009). For the current study, FaceGen Modeller Version 3.3 (Singular Inversions Inc.) was used to randomly produce 60 African, 60 East Asian, and 60 Caucasian (European, to be more specific) male faces.

FaceGen controls were set so that African and East Asian faces would be more objectively variable than Caucasian faces, and heighten the recognition of

African and East Asian faces. Given that face recognition is poorer with FaceGen faces compared to real faces (Crookes et al., 2015), the chance of other-race floor effects may be higher with FaceGen faces than real faces; this would be problematic for measuring effects of experience on recognition. Increasing other-race (for Experiment 1, and a large proportion of Experiment 2 participants, African and East Asian faces would largely be other-race) variability may counter this.

Faces had no hair either from the top of the head nor the jaws, and each face presented a neutral emotional expression. Stimuli were encapsulated by an oval, with the neck and ears still visible. The virtual light source was constant across faces. Greyscaled faces were presented from two visual orientations, one being frontal and the other being rotated 45° to the right (relative to the frontal placement) from the perspective of the participant. Research has featured learning and test phases in each trial of a face-matching task (e.g., Lindsay et al., 1991), and such a paradigm was used in the present study.

2.2.3 Procedure

All sessions of the experiments took place at City, University of London, and each session lasted around 60 minutes. Informed consent was obtained prior to the administering of questionnaires. Following the questionnaires, the face-matching task was completed. The task, which was written via the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), was delivered in MATLAB and consisted of 360 experimental trials (in a random order) which were preceded by six practice trials (randomly chosen from the 360). Each trial started with a central, red fixation point which appeared and disappeared twice (750 ms). Next, a frontal-face was presented centrally (500 ms), which was followed by a masked interval (3000 ms) and then a rotated face (500 ms), after which participants indicated whether faces were of the same identity, or different identities, by the use of response keys (Figure 2.1). On 180 of the experimental trials, there was an identity mismatch between the frontal and rotated stimuli; the rotated stimulus was randomly selected on these trials. After the face-matching task, participants were debriefed.

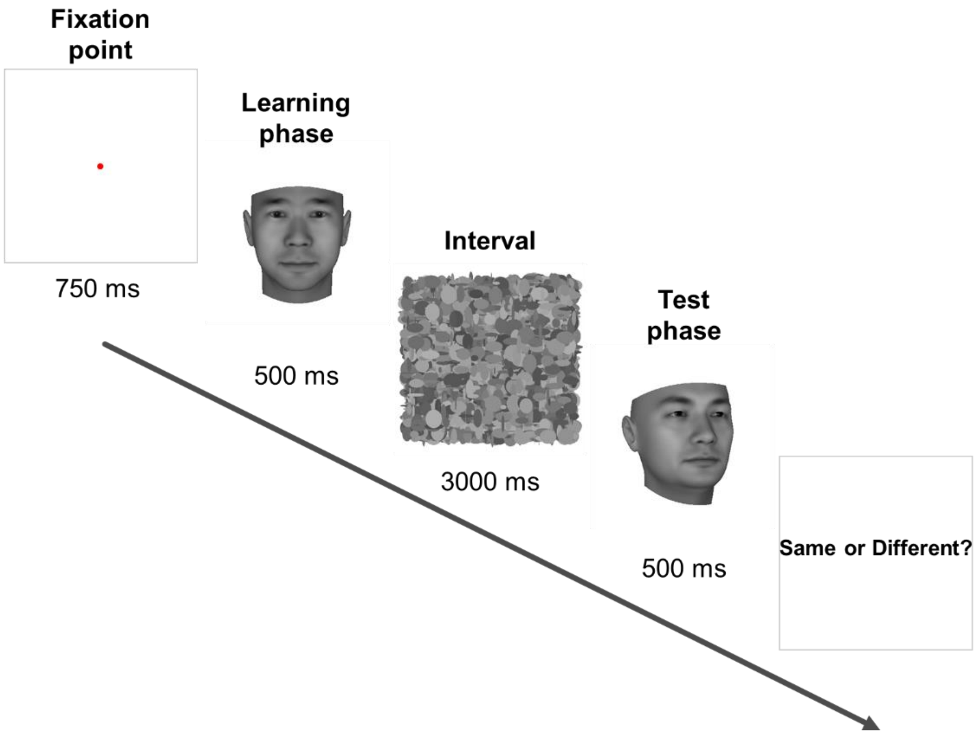


Figure 2.1. Depiction of a trial in the sequential face-matching task.

2.2.4 Data preparation

When calculating *d*', per participant for each race of face stimulus, any hits or false alarms which had rates of 0 or 1 were respectively substituted with .5/*n* or 1 - (.5/*n*), with *n* being the total number of relevant *same*/*different* trials (Stanislaw & Todorov, 1999), as has been performed on previous research on the other-race effect (Adams, Pauker, & Weisbuch, 2010; Crookes et al., 2015; Kloth, Shields, & Rhodes, 2014). In face recognition research, outliers have been defined, for instance, as values located outside of 1.5 × the interquartile range from the upper or lower quartile (Rhodes, Locke, et al., 2009). Across the current experiments, a differing number of variables was employed. In an analysis, a greater number of variables would increase the frequency of ostensible outliers (false positives). For this reason, a less conservative criterion was applied (and used across experiments for consistency). The criterion for defining an outlier was that a value exceeded 3 × the interquartile range from the upper/lower quartile. Additionally, where relevant, Studentized deleted residuals were used, with values surplus of |3| being noted as being outliers. Any participant who had a missing or illegible response on a relevant questionnaire item was removed from that particular analysis. Regarding questionnaires, outliers were defined in terms of totals, and not at the item-level. For the purpose of controlling the family-wise error rate, Holm-Bonferroni corrections were engaged (Gaetano, 2013; Holm, 1979).[20] SPSS (Versions 24 and 25) was used for the analyses, as were R Versions 3.3.0 (R Core Team, 2016).

2.3 Experiment 1

The first experiment intended to find if a difference between own- and other-race childhood experience correlates with the adulthood other-race effect (for Caucasian observers). Often, studies which investigate the potential link between experience and the other-race effect examine if there is a relationship

[20] Nonetheless, regarding *p*-value corrections in the context of multiple testing, i) it seems indeterminate how a family of tests ought to be defined (Feise, 2002), ii) it would appear commonplace for adjustments to be applied across a family of the total number of tests which occurred in an experiment, but not across a whole experiment, such as within the domain of face recognition (e.g., Horry et al., 2015; Michel et al., 2013), and iii) applying corrections increases the chance of Type II errors (not detecting an actual effect) (Rothman, 1990).

between *other-race experience* and the other-race effect (e.g., Young & Hugenberg, 2012; Zhao et al., 2014a, 2014b), rather than whether the own- vs. other-race *difference* in experience relates to the other-race effect (Wan et al., 2013). However, studies concerning the contributions of expertise to the other-race effect have commonly examined if the own- vs. other-race expertise difference has a relationship with the other-race effect (e.g., Zhao et al., 2014a, 2014b).

Figure 2.2 presents experience levels for two hypothetical observers. Observer A has a high level of experience for own-race faces and a low experience level for other-race, whilst Observer B has medium own-race experience and low other-race experience. Observer A would have a greater other-race experience disadvantage (experience for other-race faces subtracted from own-race experience) than Observer B. Under an experiential account of the other-race effect (e.g., Valentine, 1991), Observer A would have a larger magnitude of the other-race effect than Observer B, i.e., rather than other-race experience, it would be the (within-observer) disparity between own- and other-race experience which would be indicative of the other-race effect. Using other-race experience/expertise (instead of experience or expertise differences) could lead to unduly underwhelming impressions of the roles which experience and expertise play in the other-race effect (i.e., smaller effect sizes). Therefore, other-race experience disadvantages were used in the present experiment.

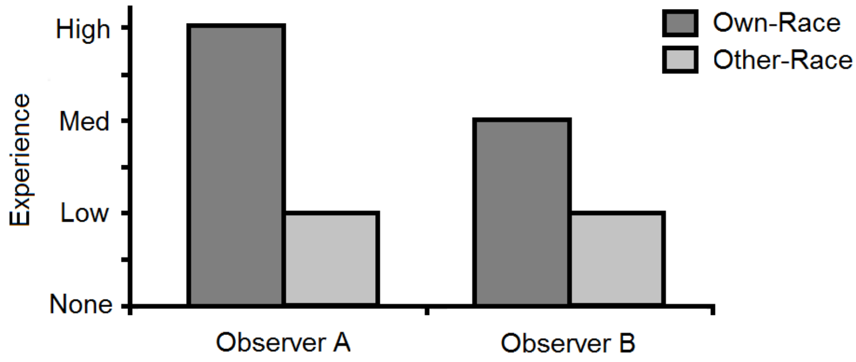


Figure 2.2. Observers A and B are similar in terms of their other-race experience, yet the size of the other-race experience disadvantage is greater for Observer A.

2.3.1 Results and discussion

Accuracy scores (d') were analysed by way of a 3 (race [Caucasian, East Asian, African]) × 1 within-subjects ANOVA. A main effect of race, $F(1.77, 85.02) = 12.34$, $p < .001$, $\eta_p^2 = .20$, arose from accuracy being higher with African faces ($M = 1.45$, $SD = .63$) than for either East Asian ($M = 1.25$, $SD = .59$), $p < .001$, or Caucasian faces ($M = 1.17$, $SD = .51$), $p < .001$, whilst accuracy was the same for faces of East Asians and Caucasians, $p = .17$ (Figure 2.3). Hence, the converse of the other-race effect took place regarding African faces, and the other-race effect did not occur regarding East Asian faces. This should be expected given the facial variability adjustment.

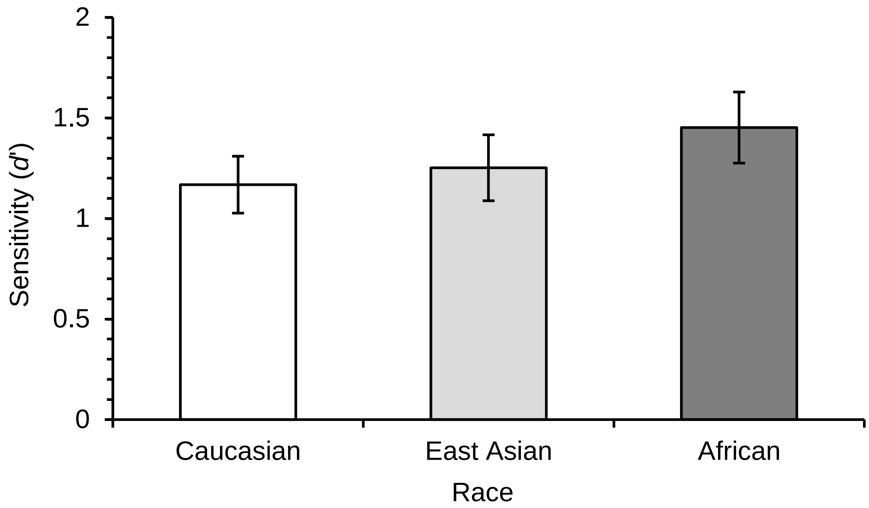


Figure 2.3. Accuracy (*d*') regarding the mean performance of Caucasian observers regarding own-race, East Asian, and African faces. The error bars represent the standard error of each relevant mean.

Ordinal alpha (polychoric correlations) has been preferred over Cronbach's alpha in the presence of Likert data, and ordinal alpha seems more tolerant than Cronbach's alpha to the skewness of items (Zumbo, Gadermann, & Zeisser, 2007). IRCQ items are Likert data, and a number of the items appeared skewed for Experiment 1 participants. Through the SPSS R-Menu (Basto & Pereira, 2012) ordinal alphas were derived;[21] IRCQ items had good internal consistency across the three races (Table 2.3).[22] Data for two Caucasian participants were removed on the basis of their PI20 responses; one had a PI20 score which would indicate moderate developmental prosopagnosia, and the other participant was removed due to their poor handwriting legibility on PI20 item responses. Another

[21] Each time the SPSS R-Menu was utilised in this thesis, it was deployed in SPSS Version 25, and the SPSS R-Menu made use of R Version 3.3.0.

[22] If anything, the very high ordinal alphas for experience with Caucasians would indicate redundancy between items.

participant was removed as raw responses indicated their non-engagement in the face recognition task (i.e., near-uniform responses, especially compared to other participants).

Table 2.3

Ordinal Alpha Values for Non-Valenced IRCQ Items

Life stage (years)	Black	Caucasian	East Asian
0-6	0.92	0.96	0.90
6-12	0.88	0.97	0.91
12-18	0.90	0.96	0.90
Mean (0-18)	0.90	0.96	0.90

Two sets of correlational analysis were used with respect to magnitudes of the other-race effect (*d*') and other-race experience disadvantages at the three life phases (0-6, 6-12, and 12-18 years). One set pertained to the relationship between Caucasian–East-Asian differences (in recognition and experience), and the other was for Caucasian–African/Black differences. Holm-Bonferroni adjustments (Gaetano, 2013; Holm, 1979) were applied (six tests). Regarding Caucasian–East-Asian differences, there was no correlation between the size of the other-race effect and the other-race experience disadvantage when contact was at 0-6 years [$r_s(44) = .01$, $p = 1.00$], 6-12 years [$r_s(46) = -.001$, $p = 1.00$], or 12-18 years [$r_s(46) = -.02$, $p = 1.00$]. Similar outcomes were evident concerning Caucasian-African/Black differences, as the magnitude of the other-race effect had no association with the experience disadvantage at 0-6 [$r_s(44) = .02$, $p = 1.00$], 6-12 [$r_s(45) = .04$, $p = 1.00$], or 12-18 years [$r_s(46) = .02$, $p = 1.00$]. Furthermore, benchmarks for effect sizes (Cohen, 1992), after converting r_s-values to r-values

(Walker, 2003), suggest that each of the effect sizes in Experiment 1 was very small.

In the one previous study, Wan et al. (2013), which explored whether the other-race experience disadvantage relates to the other-race effect in adulthood, for each childhood timespan used in that experiment (5-12 years, and 12-18 years) there were no associations, except for contact at 5-12 years with classmates (correlations did not occur regarding contact with friends or neighbours). In the current experiment, no correlations were apparent at any time span. This could suggest that contact during adulthood would determine the other-race effect, however, correlations between the other-race experience disadvantage in adulthood and the adulthood other-race effect were not evident in Wan et al. (2013). Overall, results from the current experiment are not supportive of a role of early experience in the adulthood other-race effect.

2.4 Experiment 2

This experiment sought to answer the question of whether the valence of other-race experience during childhood predicts the recognition of other-race faces in adulthood. To do this, two multiple regressions (Ordinary Least Squares) were run, one of which used recognition accuracy for East Asian faces as the dependent variable, whilst the independent variables were PI20 scores, positive East Asian contact (0-18 years), negative East Asian contact (0-18 years), IMS scores, EMS scores, and the IMS × EMS interaction; the other multiple regression featured recognition for African faces as the independent variable, with predictors being the same as the regression for East Asian faces, but with the relative positive and negative contact with Blacks substituted in place of East Asian contact. A multiple regression was not run with Caucasian faces as the independent variable due to the anticipated small number of non-Caucasian participants (considering the number of predictors). The intention of including PI20 scores was to control for overall face recognition skill, whilst IMS and EMS scores were envisioned to adjust for prejudice concerns (see Section 2.2.2.1).

2.4.1 Results and discussion

Internal consistencies of IRCQ and PI20 items were calculated with the SPSS R-Menu (Basto & Pereira, 2012). Generally, the valenced IRCQ items had acceptable internal consistency (Table 2.4), as did PI20 items (ordinal alpha = .91; N = 111) for Experiment 2 participants. In Plant and Devine (1998), the IMS and EMS scales were completed solely by Caucasians concerning prejudice motivations regarding Blacks. However, IMS and EMS in the current study inquired into prejudice motivations of persons from various races concerning *other-race* persons (i.e., more generally). Therefore, it seemed sensible to test the nature of other-race IMS and other-race EMS.

Due to other-race IMS and EMS items generally being skewed, the SPSS R-Menu (Basto & Pereira, 2012) was used for exploratory factor analysis (polychoric correlations, maximum likelihood; N = 110) and ordinal reliability analysis of the IMS and EMS items; the determinant was .002 (i.e., it did not suggest multicollinearity, Field, 2013), a Kaiser-Meyer-Olkin statistic of .80 indicated the suitability of factor analysis (Kaiser, 1974) for the current data, and examination of a scree plot (Cattell, 1966) suggested that two factors should be retained. Two factors collectively explained 59.53% of the variance prior to rotation. Following an oblimin rotation, IMS items each loaded highly onto one factor, whilst EMS items did so on the other factor. Internal consistencies were good for IMS and EMS items, ordinal alpha = .90 (N = 110) and ordinal alpha =

.84 (N = 111) respectively. IMS and EMS scales correlated with each other, $r_s(107) = -.20$, $p = .038$. Therefore, as in Plant and Devine (1998), i) two factors were evident, which separated IMS from EMS, ii) both scales had more than adequate internal consistency, and iii) scales were negatively correlated. This tentatively suggested that the other-race version of the IMS and EMS functioned similarly to the Black IMS and EMS of Plant and Devine (1998).

As in Experiment 1, data of three Caucasians were eliminated due to PI20 responses or lack of engagement in the face-matching task (Section 2.3.1). Two further participants were removed for not completing the face-matching activity, and another was eliminated as their raw responses indicated that they were not engaged in that task. The data of participants who gave ambiguous responses for their race/ethnicity was not included in the multiple regressions. Prior to the multiple regression featuring non-Black participants, two outliers were identified and eliminated. Two further participants were removed due to them having high Mahalanobis distance (24.81, 22.47) and leverage values (.30 and .27 respectively).

Table 2.4

Ordinal Alpha Values Pertaining to Valenced IRCQ Items

Life stage (years)	Valence	Black	Caucasian	East Asian
0-6	Positive	0.88	0.87	0.92
	Negative	0.68	0.74	0.81
6-12	Positive	0.88	0.89	0.88
	Negative	0.72	0.77	0.83
12-18	Positive	0.81	0.74	0.87
	Negative	0.67	0.62	0.81
Mean (0-18)	Positive	0.86	0.83	0.89
	Negative	0.69	0.71	0.82

Neither valence (positive or negative) of other-race contact predicted other-race recognition in either regression (Tables 2.5 and 2.6). This aligns with Jerovich (2017), wherein positive and negative other-race experience did not correlate with other-race recognition. It was speculated that the Walker and Hewstone (2006) finding of individuating other-race contact predicting other-race recognition may have been due to items reflecting positively-valenced experiences (Section 1.2.4.1); results in the current experiment suggest that it was not the valence of contact that was important. PI20 scores were not significantly predictive of other-race recognition in either analysis (although it could be said that scores were *marginally* predictive).

Table 2.5

Results of a Multiple Regression with Non-East-Asian Participants, Predicting their Recognition (d') of East Asian Faces

Independent variables	*B*	*SE B*	CI	*β*	*p*
Positive contact (0-18)	-.02	.07	-.16, .12	-.06	.73
Negative contact (0-18)	-.06	.13	-.31, .20	-.08	.66
PI20	-.02†	.01	-.03, .002	-.21	.089
IMS	-.04	.08	-.20, .12	-.07	.58
EMS	-.02	.05	-.12, .08	-.07	.62
IMS × EMS	.01	.06	-.11, .12	.02	.91

Note. $N = 72$; $R^2 = .07$; CI = 95% confidence interval; †$p < .10$.

Table 2.6

The Outcome of a Multiple Regression, Predicting Other-Race (African) Recognition Accuracy (d') Amongst Non-Black Observers

Independent variables	*B*	*SE B*	CI	*β*	*p*
Positive contact (0-18)	.05	.06	-.07, .16	.12	.44
Negative contact (0-18)	-.003	.08	-.17, .16	-.01	.97
PI20	-.01†	.01	-.03, .00	-.22	.055
IMS	-.08	.07	-.21, .05	-.15	.23
EMS	-.07*	.04	-.15, -.001	-.23	.047
IMS × EMS	.05	.03	-.02, .12	.17	.13

Note. $N = 82$; $R^2 = .14$; CI = 95% confidence interval; †$p < .10$, *$p < .05$.

Whether EMS related to other-race recognition was not a focus of this experiment. Previous research had found EMS to be predictive of neither the other-race effect (Wilson, 2010) nor a possible indicator of the other-race expertise disadvantage (Davis et al., 2016); these previous experiments featured Caucasian participants, with faces being Caucasian and Black, whilst EMS was asked with respect to the prejudice motivations of Caucasian participants regarding Blacks. If persons higher in EMS have poorer other-race recognition, it may be due to them diverting attention away from other-race faces or reduced individuation due to greater activation of the other-race category (Wilson, 2010). It is puzzling why EMS (at its average) emerged as a significant (negative) predictor for non-Blacks recognising Africans and not for non-East-Asians recognising East Asians; perhaps other-race EMS scores were more reflective of African EMS for non-Blacks than East Asian EMS regarding non-East-Asians.

2.5 General discussion

The present chapter aimed to determine i) if the other-race experience disadvantage (at different life stages during childhood) correlates with the other-race effect, and ii) whether the valence of other-race contact predicts other-race recognition.

2.5.1 Experience

In the first experiment, there were no associations between childhood experience (non-valenced) and the other-race effect. Similarly, valenced other-race contact did not predict other-race recognition in Experiment 2. In a previous study, there was a correlation between other-race experience and the other-race effect when motivation to individuate other-race faces was maximised via instructions to individuate such faces (in the absence of those instructions [their Experiment 1] there was no correlation) (Young & Hugenberg, 2012); motivation may have trumped experience in Experiments 1 and 2.

2.5.2 Variability

It is worth considering if the variability manipulation of the FaceGen stimuli could have resulted in there being no associations found between experience and the other-race effect (in Experiment 1) or other-race face recognition (in Experiment 2). Being wary of face recognition being more difficult with FaceGen faces than real faces (which could hinder the ability to find links between experience and face recognition), the diversity of FaceGen faces was manipulated such that African and East Asian faces would be more heterogeneous than the Caucasian faces (Section 2.2.2.2). The lack of other-race effects in Experiment 1 suggests a relative heterogeneity of the African and East Asian FaceGen faces over the Caucasian ones. This means that there were no other-race effects present to associate with childhood experience (Experiment 1). Consideration can be given to whether this absence of other-race effects signals a problem for finding whether experience and recognition are associated. Assuming that a relationship between experience and face recognition would be affected by changes in the diversity of FaceGen stimuli, too high (or too low) a level of facial diversity would disrupt an association between experience and face recognition (Figure 2.4).

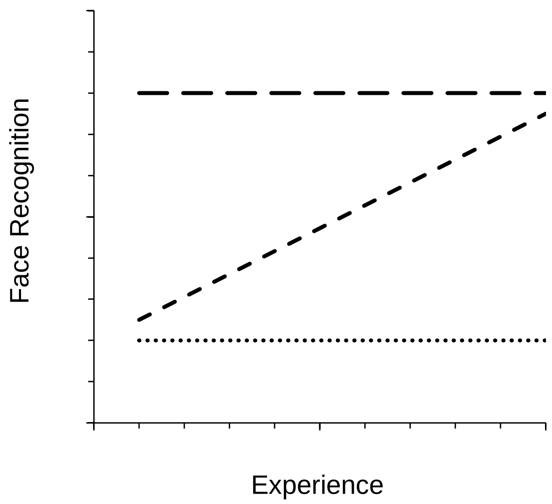


Figure 2.4. Scenarios of links between experience (*x*-axis) and face recognition (*y*-axis) at various levels of FaceGen facial diversity. There are three linear trend lines: large, medium, and small dashes represent facial diversity respectively at high, medium, and low levels. In this scenario, sufficiently high or low facial diversity levels eliminate an association of experience and face recognition.

The absence of other-race effects in the present study may be indicative of the diversity of the African and East Asian faces being set at such a high level that such a disruption occurred in Experiments 1 and 2, thereby leading to associations not being observed between experience and face recognition in either experiment. Ideally, a follow-up to the present experiments would occur using equal variability levels across African, Caucasian, and East Asian FaceGen stimuli (lower than the variability of the African and East Asian faces in the present study); the presence of the other-race effect would then allow for stronger conclusions to be made regarding whether associations occur between childhood experience and both the other-race effect and other-race face recognition.

2.5.3 Fixation point placement

The consequences of fixation cross manipulations in previous research (Hills et al., 2013; Hills & Lewis, 2011) could suggest that the position of the fixation point (in Experiments 1 and 2) disrupted relationships between experience and recognition. Relative to the lower section of the face, the upper portion may house relatively individuating details amongst Caucasian faces, whilst the opposite may be a reality for the faces of Blacks (Hills et al., 2013; Hills & Lewis, 2011; Hills & Pake, 2013). Indeed, a fixation stimulus draws first fixations (Hills et al., 2013), and the effects of fixation crosses on recognition indicate the importance of where attention is assigned for the manifestation of the other-race effect (Hills & Lewis, 2011).

In Experiment 1 of Hills and Lewis (2011), relative to when no fixation cross was employed, a fixation cross location which heralded the subsequent location of the nose tip[23] (at learning and test) increased the accuracy with which Caucasians recognised faces of Blacks but lowered performance for own-race faces. In the face-matching task of the current chapter, the fixation point was placed in the position which the nose tip subsequently occupied.

Prior research examining fixation cross placement in the context of the other-race effect had fixation crosses appear before *learning* and *test* faces (Hills et al., 2013; Hills & Lewis, 2011; Hills & Pake, 2013), however, in the present experiments, there was no fixation point at test. At the test phase, effects on face recognition have been found using manipulations of processing demands between face pairs, which suggests that expert processing at *test* has value for face recognition (Ho & Pezdek, 2016). In the current experiments, participants were free to use their default first fixation location (e.g., Hills & Pake, 2013) at test. It is not known which phase (learning or test) has been the driver for effects of fixation crosses in previous research on the other-race effect (e.g., Hills et al., 2013), therefore any assertions that the experience-recognition link was disrupted by the learning fixation cross remains speculative. Nonetheless, recognition of Caucasian faces is greater when a fixation cross proceeds the eye region (as opposed to the mouth region) at learning, and this effect of fixation cross positioning also occurs at test (Hills, Ross, & Lewis, 2011). This indicates that influences of fixation crosses on the other-race effect can occur at learning yet also at test; it remains a possibility that fixation point usage in Experiments 1 and 2 (i.e., at learning) may have disrupted experience-recognition links.

2.5.4 Computer-generated faces

Relationships between contact and face recognition might not have been

[23] This location is specified to have been used in Hills and Lewis (2011) by Hills and Pake (2013).

found due to the use of FaceGen stimuli.[24] The extent to which the morphology of FaceGen faces reflects reality is debateable, for instance, chin length is longer for Africans and East Asians than Europeans in real life, unlike in FaceGen (Holland, 2009), therefore a lack of (or weakened) experience-recognition relationships could be predicted. FaceGen stimuli (whether randomly-generated or versions of real identities) may "fail to fully reveal face expertise" perhaps due to a loss of surface information, or considering computer-generated faces as outgroup, or a lack of experience with computer-generated faces (Crookes et al., 2015, p. 15).

Accordingly, in Crookes et al. (2015), Caucasians recognised real East Asian faces less accurately than they did real own-race faces, yet East Asian and Caucasian FaceGen stimuli (randomly-generated) were recognised equally well. Nonetheless, with FaceGen faces, Caucasians have exhibited lesser recognition for African than own-races in previous studies (Matheson & McMullen, 2011; Pauker et al., 2009).

Still, FaceGen faces are representative of real-life face recognition to a tangible extent, as suggested by the presence of the other-race effect with the use of FaceGen faces (Matheson & McMullen, 2011; Pauker et al., 2009). However, given Crookes et al. (2015), some difference from reality should be expected when using FaceGen stimuli in examinations of the relationship between experience and either the other-race effect or other-race recognition. Interestingly, the relationship between contact questionnaire items and the other-race effect (or

[24] In FaceGen, a face can be generated by the program, however, it can alternatively be derived directly from an image of an identity, and both types have been used with respect to the other-race effect (Crookes et al., 2015). Unless stated otherwise, this thesis refers to the generated type.

other-race recognition) had not previously been explored using FaceGen stimuli.

2.5.5 Conclusion

Relationships between childhood experience and adulthood facial recognition were not evident, specifically between the non-valenced other-race experience disadvantage and the other-race effect, and between valenced other-race contact and other-race facial recognition in terms of magnitudes. However, bearing in mind the variability levels of African and East Asian faces (which likely resulted in other-race effects not occurring), and considering the use of fixation crosses and/or synthetic faces, it should not by any means be concluded that childhood experience and the other-race effect (or other-race face recognition) are actually not associated.

Chapter 3: OBJECTIVE AND SUBJECTIVE VARIABILITY

Section 1.2.2 described how the Rossion and Michel (2011) argument for Africans being heterogeneous in their facial diversity relative to Caucasians had not truly been addressed in previous research. Although Goldstein (1979a) established that Japanese East Asians and Blacks matched the variability of Caucasians, comparisons were solely descriptive in nature. Furthermore, Melanesians were placed in the Black group along with Africans. Melanesians are not African (e.g., Friedlaender et al., 2008; McEvoy et al., 2010). Comparisons in Phipps et al. (1988) were solely of Native Americans and Caucasians. Racial disparities in physical diversity do not ordinarily hold a place in theories of the other-race effect. Rossion and Michel (2011) suggested that the other-race effect could be moderated by racial differences in morphological variability; finding such differences would add weight to this idea.

As for subjective variability, in Section 1.2.5, it was detailed how the other-race expertise disadvantages have generally not being found to relate to the other-race effect, which is troublesome for ideas of the other-race effect having a perceptually-driven foundation. Moreover, Section 1.2.6 described how it was untested if other-race subjective homogeneity itself relates to the other-race effect; finding a relationship would bolster the idea of the other-race effect having a perceptually-driven route. The current chapter explores whether objective variability is equal across races/ethnicities, and it then considers if other-race subjective homogeneity is predictive of the other-race effect.

3.1 Objective variability

3.1.1 Introduction

As modern humans migrated farther away from their origin in Africa, their skulls became less heterogeneous in terms of mean variances (which aligns with the reduction in genetic diversity),[25] and this pattern of diminishing morphological variability occurs amongst males and females (Betti, Balloux, Amos, Hanihara, & Manica, 2009; Manica et al., 2007; von Cramon-Taubadel & Lycett, 2008). The decline in variability had been examined for individual facial and non-facial cranial dimensions (Manica et al., 2007), but not for the mean variance of the face within the skull (i.e., the *skeletal face*). Notably, the skeletal face does exhibit a strong association with the *full face* (i.e., the face as seen on a living person, with its soft tissues etc.) (Young et al., 2016), hence findings with the skeletal face should be relevant for the full face.

In addition to the mean variance, other measures of variability were also of interest in the current chapter, as were between-dimension correlations. The alternative variabilities of interest were pattern variability (Garn, Lavelle, & Smith, 1985) and the standardised generalised variance (SenGupta, 1987). The pattern variability index is calculated at the level of the individual person, and is defined as the standard deviation of their *z*-scores (Garn et al., 1985), thereby providing an indication of the within-face distribution of attributes. For instance, a person could have measurements from a narrow distribution across facial dimensions (e.g., near to the mean in each attribute), giving a lower index than if

[25] A serial founder effect occurred, i.e., "[s]ubsamples of established populations would move to new areas, founding new communities that would in time be the origin of further expansion" and "[w]ith each founding event, some genetic diversity is lost at random" (Betti, von Cramon-Taubadel, Manica, & Lycett, 2013, p. 2).

their measurements were from a wider distribution.[26]

According to a literature search, studies on the relationship between the variability of the cranium and migratory distance (e.g., Manica et al., 2007; von Cramon-Taubadel & Lycett, 2008) had not examined whether there is general reduction in cranial variability when associations between measurement dimensions are considered. In Sheehan and Nachman (2014), compared to the body, morphological measurements of the face/head were more variable in terms of coefficients of variation, and exhibited lower correlations. It could be that skeletal facial measurements become more correlated as migratory distance extends.

Although the mean variance would not capture between-dimension associations, covariances (standardised as correlations) would. Three populations from W. W. Howells' cranial data had been analysed in terms of the entire variance-covariance matrix, i.e. variances and covariances, by Petersen (2000), who used 10 dimensions, and also five of those 10. Petersen (2000) compared determinants of covariance matrices (generalised variances) and derived their relevant standardised generalised variances (determinant ratio raised to the power of the reciprocal of how many measurement types there were). Nevertheless, Petersen (2000) did not analyse the distance-variability relationship, and such a relationship had not been studied using the whole variance-covariance matrix.

Judging by results in previous studies (Betti et al., 2009; Manica et al., 2007; von Cramon-Taubadel & Lycett, 2008), one might expect a pair of

[26] Although pattern (and craniofacial) variability indices have been used in regard to morphological atypicality (Garn et al., 1985; Roelfsema, Hop, van Adrichem, & Wladimiroff, 2007; Ward, Jamison, & Farkas, 1998), this chapter in no way attempts to suggest a link between ethnicity/race and the abnormality of appearance.

populations to have unequal cranial variances when their difference in migratory distance from Africa was larger than a certain size (a *migratory distance threshold*). For example, African ethnicities might be as variable as Caucasian and East Asian ethnic groups, but be diverse relative to Native American ones.[27] This could apply to the face in isolation. However, studies were yet to assign migratory distance thresholds for morphology, either for the skull or any other section of the human skeleton. The present chapter sought to find not only whether a decline in variabilities and between-dimension correlations happen for the skeletal face, but to also pinpoint migratory distance thresholds pertaining to the types of variability.

Although the migratory distances of Africans and Caucasians might suggest that those two races would not differ in the morphological diversity of the face/head, a different prediction could be put forth if one was to consider admixture and a socially-influenced definition of race. In the context of the one-drop rule, it has been asserted that, in America, Blacks would be physically heterogeneous compared to Whites, such as in their skin colour (Fish, 2009);[28] [29] perhaps, regarding the structure of the face/head specifically, Blacks could be

---

[27] *Ethnicity* is used in Section 3.1 in an historic sense, regarding the migratory distances of indigenous groups in line with the gradual reduction in genetic diversity. An ethnic group would be either a subset of one race or be multiracial.

[28] Responses in the 2000 U.S. Census (Brittingham & de la Cruz, 2004) indicate that the vast majority of Caucasian Americans are European; from looking at Table 2, Figure 2a, and Figure 2b of Shriver et al. (2003), European Americans would appear to have a higher and narrower distribution of European ancestry than African Americans do regarding African ancestry, which might indeed lead to an expectation of a more diverse facial morphology for African Americans.

[29] Populations from Sub-Saharan Africa have a greater diversity of skin colour (inner upper arm) than ones from Europe in terms of variance and also coefficients of variation measured at the population level (Relethford, 2000) which may suggest that, in America, Blacks having heterogeneity in their skin colour relative to Whites would not just be driven by the one-drop rule.

more diverse than Caucasians. However, even with a social contribution to defining race categories (e.g., Davenport, 2016; Krosch & Amodio, 2014), the association between diversity and migratory distance could still be expected; Wang et al. (2007) found that the genetic diversity of Native Americans from across North and South America lowers as distance from the Bering Strait increases, despite extents of European Caucasian ancestry. Given migratory distances (Manica et al., 2007), one might expect no difference in within-race variability between Blacks and Caucasians in America, whilst Native Americans could be less variable than Blacks.

Using data from the 1988 Anthropometric Survey of U.S. Army Personnel (ANSUR) (Gordon et al., 1989), and the 2012 Anthropometric Survey of U.S. Army Personnel (ANSUR II) (Gordon et al., 2014), variability comparisons were undertaken between Native Americans, Caucasians, and Blacks with morphological measurements from full faces/heads. These comparisons were of the generalised variances of races. Statistical significance testing can be applied to generalised variance comparisons, with one such route being the nonparametric bootstrap in Petersen (2000), and this was utilised in the present chapter to compare races. Due to small sample sizes possibly obscuring bootstrapped variability comparisons (Stefan, 1999), and sizes for Native Americans being small in both the ANSUR and ANSUR II, Chapter 3 places emphasis on the comparisons of Blacks to Caucasians, whilst ones involving Native Americans are interpreted tentatively.

3.1.2 Method

3.1.2.1 Skeletal face

Measurements were of 28 populations from the Howells data, of which males are represented in each whilst females are only in 26 (Howells, 1996). Initial sample sizes (before any removals) are presented in Appendix C (Māori crania were not used as their sample sizes were small). The data were of anatomically modern humans (Howells, 1989, 1995) and are held at http://web.utk.edu/~auerbach/HOWL.htm. Howells (1989) gave descriptions of crania chosen in each population. Some descriptions refer to age (e.g., Tasmania and Arikara); it can be supposed that all populations would be representative of adults. Mandibular measurements were not available due to the incompleteness of the skulls in regard to mandibles (Howells, 1989). With information on how cranial measurements were specified (Howells, 1973), only 32 dimensions (best capturing the face) were analysed. These dimensions were from the facial skeleton and parts of the neurocranium, i.e., the front part of the skull (Table 3.1).

Von Cramon-Taubadel and Lycett (2008) used the Howells data of males alongside coordinates for i) each population, ii) the potential origin of modern humans, and iii) places between the African origin and the locations of populations. From these coordinates, they derived great circle distances for migratory travels. The coordinates featured in von Cramon-Taubadel and Lycett (2008), including coordinates for the onset of modern humans (southern African), were applied in the current analyses. Based on the literature (Jin & Su, 2000; Oppenheimer, 2012; Reyes-Centeno et al., 2014), migratory routes were selected. Great circle distances were calculated after Williams (2011) (distances are presented in Appendix C).

Mean variances were calculated across measurements for each population from *z*-scores (as in previous research, e.g., Betti et al., 2009; Manica et al., 2007). Those *z*-scores were then centred with respect to the mean of a gender in a population (e.g., male Zulu), and a pattern variability index was calculated from them for each face, which lead to mean pattern variability indices. Mean absolute Pearson's correlation coefficients (between dimensions) were calculated in MATLAB. To avoid issues with high dimensionality, principal component analysis can be used to lower the number of dimensions (Field, 2013; Relethford & Blangero, 1990; Slice, 2007). After *z*-scores of facial dimensions were submitted to principal component analysis, standardised generalised variances were calculated per population.

Table 3.1

Facial Dimensions

Abbreviation	Full name
BNL	Basion-nasion length
XFB	Maximum frontal breadth
STB	Bistephanic breadth
ZYB	Bizygomatic breadth
AUB	Biauricular breadth
WCB	Minimum cranial breadth
BPL	Basion-prosthion length
NPH	Nasion-prosthion height
NLH	Nasal height
OBH	Orbit height, left
OBB	Orbit breadth, left
JUB	Bijugal breadth
NLB	Nasal breadth
MAB	Palate breadth, external
ZMB	Bimaxillary breadth
SSS	Zygomaxillary subtense
FMB	Bifrontal breadth
NAS	Nasio-frontal subtense

EKB	Biorbital breadth
DKS	Dacryon subtense
DKB	Interorbital breadth
NDS	Naso-dacryal subtense
WNB	Simotic chord
SIS	Simotic subtense
IML	Malar length, inferior
XML	Malar length, maximum
MLS	Malar subtense
WMH	Cheek height
SOS	Supraorbital projection
GLS	Glabella projection
FRS	Nasion-bregma subtense
FRF	Nasion-subtense fraction

Note. Information from Howells (1989).

Confidence interval overlaps can be employed as an indicator of group differences, such as with 95% intervals at the .05 alpha-level (Cumming & Finch, 2005; Krzywinski & Altman, 2013). Here, they were used to indicate the minimum difference in migratory distance at which the variabilities of two groups would differ. In Cumming and Finch (2005), the smallest difference for 95% confidence intervals signifying a difference between a pair of independent groups at $p = .05$ was the top half of the confidence interval of one group and the bottom half for the second group covering each other by approximately 50% of their mean margin of error. This use of confidence intervals was extended in the current chapter by being applied to linear trend lines (Figure 3.1).

Margins of error were calculated via bootstrapping. This occurred with 1,000 resamples for each gender of a population in SPSS 23.0 for pattern variability to produce standard errors. To calculate the margins of error for the variance and standardised generalised variance, that same number of resamples was applied. MATLAB was used to produce bootstrapped total variances and determinant values which were transformed into mean variances and standardised generalised variances respectively; standard deviations of bootstrapped-derived values (standard errors) were used when calculating the margins of error. Holm-Bonferroni adjustments (Gaetano, 2013; Holm, 1979) were applied across correlations.

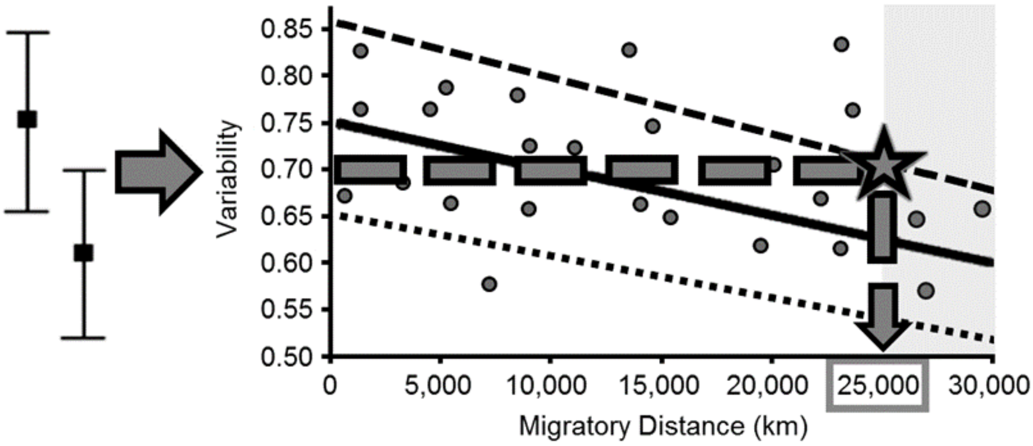


Figure 3.1. A scenario where groups are of unequal variability when their migratory distance from an African origin differs by at least 25,000 km. The positions of 95% confidence intervals portray a significant difference circa p = .05 (adapted from Cumming & Finch, 2005). In the graph, each data point (circle) is for a population. Of the three linear trend lines, the unbroken one represents a decline in variability (e.g., pattern variability) as migratory distance increases; the lines above and below are respectively for diminishing upper and lower limits.

3.1.2.2 Full face

Part of the overall ANSUR dataset (Gordon et al., 1989) was released to the public, and this data was previously available on the Defense Technical

Information Center website (e.g., Gordon & Bradtmiller, 2012). A copy of it, which was used in this chapter, is archived by Professor Matthew Reed, University of Michigan, at http://mreed.umtri.umich.edu/mreed/downloads.html (although this does not feature the set of measurements obtained via a headboard). ANSUR II data (Gordon et al., 2014) was accessed at https://insight.livestories.com/s/v2/ansur-ii/4a7623f2-62a0-4727-a984-98d8be712911/.

From the ANSUR (Gordon et al., 1989) and ANSUR II (Gordon et al., 2014), measurement data of self-categorised Caucasians, Blacks, and Native Americans were analysed. For the Caucasian category, this study combined some of the response options within the surveys. For example, as Middle Easterners are Caucasian (e.g., Risch et al., 2002; Rosenberg et al., 2002; Shriner, Tekola-Ayele, Adeyemo, & Rotimi, 2014), persons who categorised themselves as Middle Eastern were included in the Caucasian group alongside anyone who self-categorised as White. *Black* would be reflective of African racial lineage given that i) Black in the U.S. Census signifies African ancestry (United States Census Bureau, 2017), ii) African Americans have considerable African genetic heritage (e.g., Shriver, 2003), and iii) people represented in the ANSUR and ANSUR II datasets were in the *United States* Army.

With ageing, there are changes in facial morphology (Albert, Ricanek, & Patterson, 2007; Shaw et al., 2011); research on race/ethnicity and facial measurements has used sample age ranges of, for example, between 18 and either 30 (Porter & Olson, 2001) or 35 (Fang, Clapham, & Chung, 2011). Here, the range of 18 to 30 was applied.

Eight face/head measurement dimensions were of interest in the ANSUR dataset: bizygomatic breadth, interpupillary breadth, menton-sellion length, bitragion chin arc, bitragion crinion arc, bitragion frontal arc, bitragion submandibular arc, and bitragion subspinale arc. The first three were certainly of the face alone, and the latter five involved measurements that were across the face but from each tragion. The tragion is itself located on a forward point of the ear which approaches the face, and it is almost on the border between the ear and the face; dimensions involving the tragion were still mostly with respect to the face.[30] In the ANSUR II, the bitragion crinion arc, bitragion frontal arc, and bitragion subspinale arc are not featured. Consequently, only five dimensions were used from that data.

Manners in which variabilities can be compared were presented in Petersen (2000). These options were the Zhivotovsky *F*-test, Wishart bootstrap, and the nonparametric bootstrap. When raw data could be used, Petersen (2000) suggested the nonparametric bootstrap as the preferred method. In the bootstrap, after standardisation, a hypothesis category is tested for relative heterogeneity against a reference category (Petersen, 2000).[31] With the expectation of facial morphological diversity diminishing as distance from southern Africa increases, Blacks were the hypothesis group in each comparison involving them, whilst Caucasians were in their comparisons to Native Americans.

ANSUR samples sizes were small for Native Americans (6 men, 11 women), but, with sample sizes stated here after removing outliers, large for Blacks (324 men and 729 women) and also Caucasians (803 men, 870 women).

[30] The tragion has been used regarding facial measurements in previous research (e.g., Edler, Rahim, Wertheim, & Greenhill, 2010; Rhee, 2018).

[31] Dividing determinants of the hypothesis and reference categories gives the overall determinant ratio, and in each bootstrap a determinant ratio is calculated (Petersen, 2000).

The same applies to ANSUR II samples of Native Americans (6 men, 6 women), Blacks (247 men, 307 women) and Caucasians (1617 men, 627 women). Stefan (1999) expressed that small sample sizes could reduce the ability to statistically detect true differences in variability. For instance, one of the comparisons (group sizes of 19 and 12 crania) in Stefan (1999) had a determinant ratio of 19,722, yet a *p*-value of .19 using a similar manner to the nonparametric bootstrap; a different comparison, which had 36 and 25 crania, gave a determinant ratio of 83.39, and a *p*-value of .008. A small sample size would seem to impair the ability to find actual variability differences with nonparametric and Wishart bootstrapped determinant ratio comparisons, for example, with sizes of 57 for one group and 11 for another, Scherer and Wright (2015) observed what seems to be a considerably large determinant ratio of 20,766.71, yet *p*-values of .09 and .27 respectively for Wishart and nonparametric bootstraps. Similarly, the *p*-value for the Zhivotovsky *F*-test was .09.

Large sample sizes can increase the chance of finding actual effects of variables, but also *cause* significant results on statistical tests (Field, 2013). For bootstrapped tests, because increasing (re)sample size reduces the variability of estimates derived from bootstrapping (Ding, Bressler, Yang, & Liang, 2000; Hesterberg, Monaghan, Moore, Clipson, & Epstein, 2003), resampled determinant ratios would likely be more consistent with greater sample size, meaning the facility to detect differences increases, however, this indicates that a result apparently contrary the null hypothesis could indeed be directly driven by a large sample. With these points and given the sample sizes in previous determinant comparisons (Nystrom & Malcom, 2010; Petersen, 2000; Scherer and Wright, 2015; Stefan, 1999; Weisensee, 2001) a few steps were undertaken to minimise the negative impact of large samples.

In bootstrapping, rather than having samples of n, subsamples have been used (e.g., Bickel, Götze, & van Zwet, 1997). The nonparametric bootstrap can be modified to specifically select a certain number of items from a category. For variability comparisons between Africans and Caucasians, per bootstrap and still using replacement, a random 50 Africans and 50 Caucasians were selected. Sample sizes of Caucasians were equalised with that of Blacks by random number generation in SPSS 22.0. Not initially matching sample sizes, but using bootstrap subsamples of equal size per group, would cause a bias in favour of a larger sample (Caucasians for both genders) being more variable. When comparing either of these races to Native Americans, 50 faces/heads were chosen as a random sample from those two races, and the usual nonparametric bootstrap was applied. Regarding the ANSUR data, because the number of Native American men between 18 and 30 years was fewer than the number of dimensions, comparisons with Native Americans only occurred for women.

3.1.3 Results

3.1.3.1 Skeletal face

3.1.3.1.1 Variance

For the skeletal data, univariate outliers were determined via z-scores in SPSS 23.0 (computed within each population and gender combination) as values exceeding |3.29|, and they were removed. Analysis in SPSS 23.0 showed a negative correlation between mean variance and migratory distance for the male and female populations, $r_s(26) = -.50$, $p = .038$ and $r_s(24) = -.52$, $p = .038$ respectively. Migratory distance thresholds (derived from linear trend lines) were indicated to be at 25,000 km for males, whilst beyond 30,000 km for females

(Figure 3.2). Prior to using the bootstrapping method described in Section 3.1.2.1, migratory distance thresholds were calculated using the 32 dimension variances (for each population and gender combination) as the sample of variances on which bootstrapping was applied in order to calculate margins of error; thresholds were found of approximately 20,000 km for males and 23,000 km for females. However, it was felt that bootstrapping at the level of crania (thereby generating a variance-covariance matrix per bootstrap) would be more robust that merely using the same 32 values (the dimension variances) as the sample for bootstrapping. See Appendix A for the levels of each variability and interdimensional correlations in populations.

a)

b)

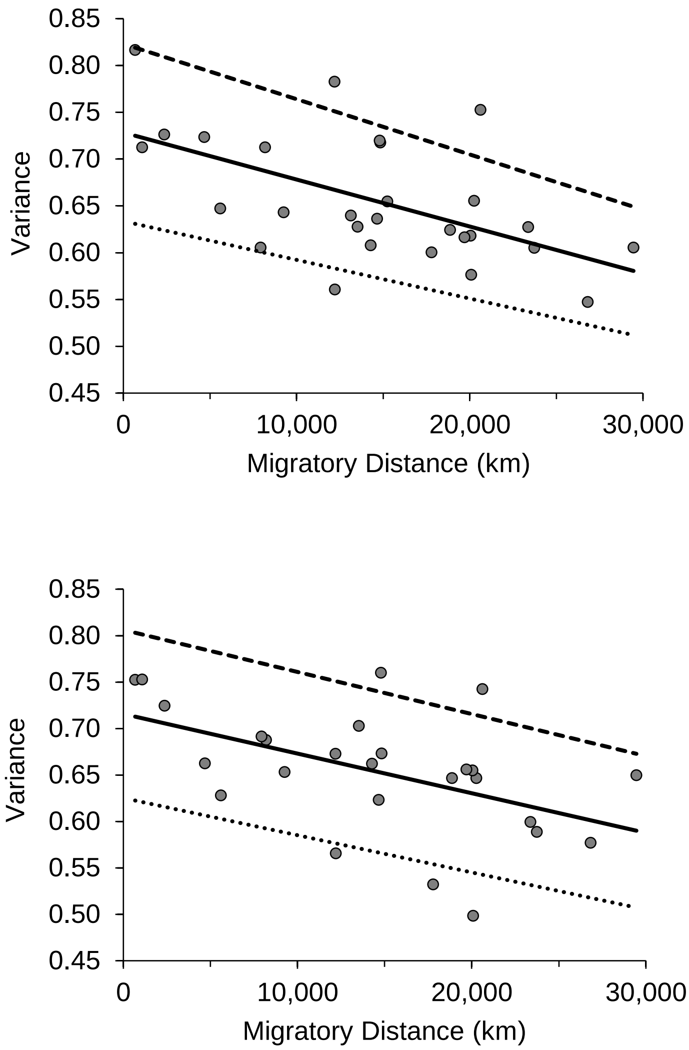
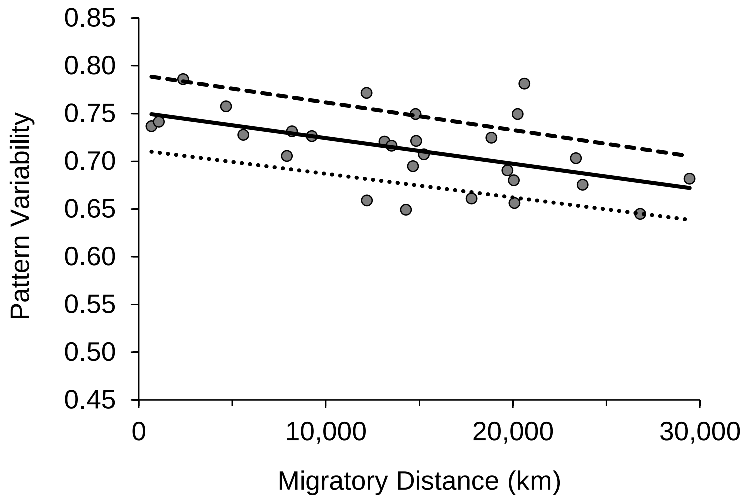


Figure 3.2. For males (a) and females (b), the graphs display the relationship between facial skeletal variance and southern-African migratory distance. Each circle represents the mean variance of a population. Per graph, there are three linear trend lines. The middle one is the linear trend line regarding pattern variability and migratory distance. The lines above and below it were generated from calculating the 95% confidence interval for each population, plotting upper and lower confidence limits per group, and then fitting linear trend lines, the higher one being for the upper limit and the other the lower limit.

3.1.3.1.2 Pattern variability

Migratory distance was also associated with mean pattern variability for males, $r_s(26) = -.50$, $p = .038$, and females, $r_s(24) = -.52$, $p = .038$, with the correlations occurring in the expected direction. Linear trend lines (Figure 3.3) showed migratory distance thresholds of around 20,000 km and 21,000 km respectively for male and female faces.

a)



b)

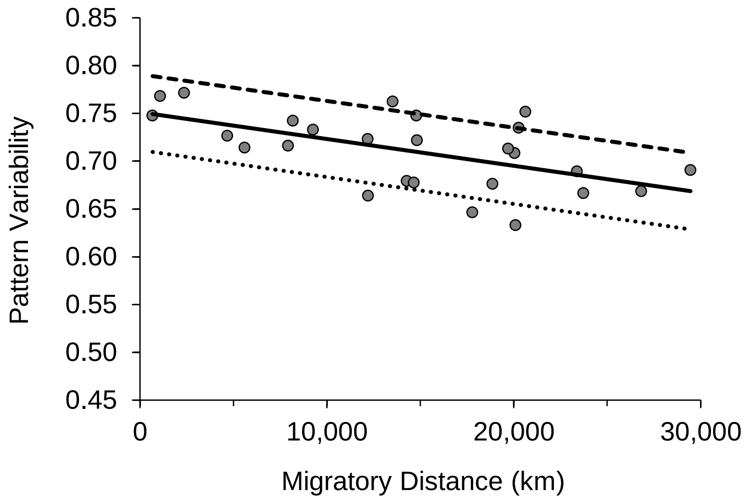


Figure 3.3. As in Fig. 3.2, but for pattern variability rather than variance regarding males and females (a and b respectively).

3.1.3.1.3 Interdimensional correlations

After excluding any face which had at least one univariate outlier, the ppcor package (Kim, 2015) was used in R Version 3.5.3 (R Core Team, 2019) to find whether there was an association between migratory distance and mean absolute correlation values when controlling for the ability of sample size to account for mean absolute correlations; there was no semi-partial correlation for either males, $sr_s(25) = .17, p = .39$, or females, $sr_s(23) = .27, p = .37$.

3.1.3.1.4 Standardised generalised variance

Screening for multivariate outliers via Mahalanobis distances (e.g., Cousineau & Chartier, 2010) was performed on *z*-scores in SPSS 23.0 within each population, however, small sample sizes in several populations precluded the utility of both this and the (preceding) univariate outlier discovery. Checks were made for univariate normality, and using guidelines in Arifin (2015), multivariate normality in SPSS 22.0. There was a focus on outliers being defined at the population-level rather than generally in order to retain population distributions as much as possible. However, because outliers for a whole dataset can affect principal component analysis with their removal being a solution (Wold, Esbensen, & Geladi, 1987), any multivariate outlier within a gender overall was eliminated.

The data was found to exhibit a slight deviation from multivariate normality in the male, but not the female, dataset when each was analysed as a whole regardless of population, but this was not unexpected given that different population groups constituted the samples. Using the normality-assumed route in parallel analysis for females but the permutation method for males (O'Connor, 2000), principal component analysis was run to specify eight components for males and seven for females. Within each gender, permutation and normally distributed options gave the same number of components. For the males, 72.13% of the variance was retained, whilst 68.64% was for females. An oblique rotation (direct oblimin) was used in SPSS 22.0, which resulted in eight and seven regression-derived variables for males and females respectively. Per population for those variables, a covariance matrix determinant was calculated in MATLAB, and, from each, a standardised generalised variance was obtained.

Spearman's correlation outcomes showed that, for males, there was a diminution in the standardised generalised variance with the increase in migratory distance, $r_s(26) = -.53$, $p = .030$, as there also was for females $r_s(24) = -.55$, $p = .030$. With a power transformation converting the generalised variance into the standardised generalised variance (SenGupta, 1987), the present Spearman’s correlations also show the relationship between the generalised variance and migratory distance. Linear trend lines in Figure 3.4 suggest the migratory distance threshold was 27,000 km for females, yet beyond 30,000 km for males.

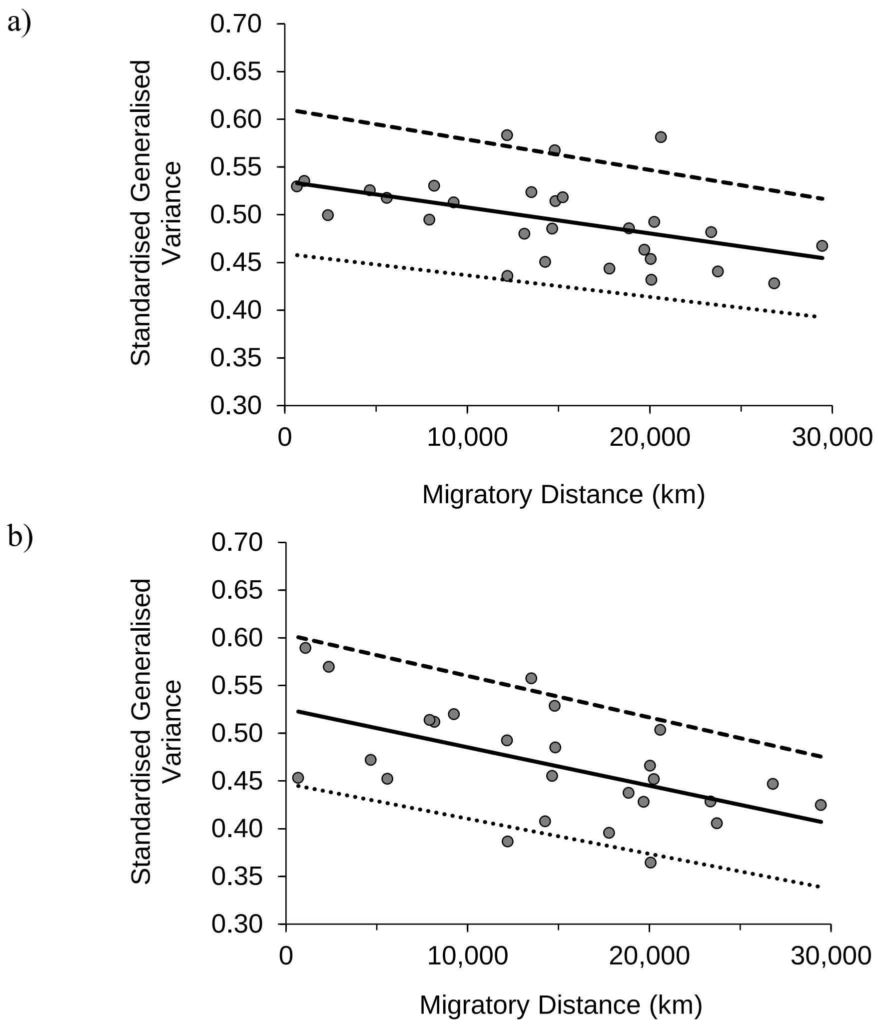


Figure 3.4. The relationship between standardised generalised variance and distance from a starting point in southern Africa for a) male and b) female of populations, including 95% confidence interval linear trend lines.

3.1.3.2 Full face

ANSUR and ANSUR II data were analysed separately. Faces/heads with any missing values were removed. As with the skeletal facial analysis process, univariate and multivariate outliers were removed for Blacks and Caucasians, but not Native Americans because the sample size for the latter was small for ascertaining whether any value was an outlier. The nonparametric bootstrap from Petersen (2000) was run in MATLAB with 9,999 resamples. Comparisons showed no racial differences in generalised variances, *p*s > .05 (Table 3.2). In the comparison column of Table 3.2, the population on the left is the hypothesis sample, and the one to the right is the reference sample. Because comparisons were regarded as two-tailed, and all determinant ratios were above one (i.e., in the expected direction), it can be inferred that reference categories were not more variable than hypothesis categories (e.g., Caucasians were not heterogeneous relative to Blacks).

Table 3.2

Racial Comparisons of Generalised Variances Using ANSUR and ANSUR II Data

		Migratory	ANSUR		ANSUR II	
		distance	Determinant		Determinant	
Gender	Comparison	difference (km)	ratio	*P*-value	ratio	*P*-value
Female	B – C	6,000	1.85	1.00	1.39	1.00
	C – NA	10,000	66.92	1.00	18.80	1.00
	B – NA	16,000	132.22	1.00	89.80	1.00
Male	B – C	6,000	1.30	1.00	2.04	1.00
	C – NA	10,000			161.34	1.00
	B – NA	16,000			205.49	1.00

Note. B = Black; C = Caucasian; NA = Native American. For significance testing, regarding each ANSUR dataset, tests were treated as being two-tailed (initially, generated *p*-values were one-tailed, and they were then doubled), and doubled once more with respect to the genders used (Bonferroni correction) and adjusted with Holm-Bonferroni corrections (Gaetano, 2013; Holm, 1979) within each gender.

3.1.4 Discussion

The objective of Section 3.1 was to find whether races/ethnicities are of equal morphological variability with respect to their faces. Previous research had found variability to be constant using coefficients of variation (Goldstein, 1979a; Phipps et al, 1988). This chapter approached the study of variability in a number of ways: i) seeing if there is a correlation between types of skeletal facial variability (and interdimensional correlations) and migratory distance, ii) discovering migratory distance thresholds, and iii) (for Blacks, Caucasians, and Native Americans) assessing whether there is unequal generalised variance of the full face/head.

3.1.4.1 Skeletal face

Following prior research which has shown that mean cranial variances dwindle as migratory distance increases (Betti et al., 2009; Manica et al., 2007; von Cramon-Taubadel & Lycett, 2008), results in this chapter demonstrated that such a decline in the diversity of the skeletal face is a reality for males and females, and not just for mean variance but also for mean pattern variability and the standardised generalised variance. Although a face samples attributes from narrower distributions as migratory distance increases (as shown by results with mean pattern variability), no partial correlation was found between migratory distance and between-attribute Pearson's correlation coefficients. The latter finding, in tandem with results concerning the variance, indicates that the decline found with the standardised generalised variance was likely driven by diminishing variances.

Confidence intervals indicated that ethnic groups will be of unequal skeletal facial variance when differences in migratory distance are larger than certain magnitudes, with the group nearer to the southern African origin being the more heterogeneous one. Migratory distance thresholds, per gender and type of variability, were 20,000 km or upwards; ethnic groups would largely be of equal variability, even when they are not of the same race. Actual differences would occur with considerable gulfs in migratory distances. Average migratory distances for Africans and Native South Americans respectively average about 4,000 km and 28,000 km; the Africans can be characterised as more heterogeneous than the Native Americans (Native North and Native South Americans together would average 25,000 km approximately as their migratory distance) regarding pattern variability for females and males. Considering the clear relationship between skeletal and full faces (Young et al., 2016), correlations and migratory distance thresholds can reasonably be assumed to apply to the morphology of the full face.

Whilst Rossion and Michel (2011) stated that African faces would be heterogeneous relative to Caucasians, linear trend lines in this chapter indicated no difference in the variability of Caucasians (their estimated average migratory distance would be 10,000 km) and Africans. Rossion and Michel (2011) held that Africans would be more variable than Aboriginal Australians. Considering that Aboriginal Australians have a migratory distance of approximately 20,000 km, results would indicate equal variability.

As for Goldstein (1979a) and Phipps et al. (1988), because they used coefficients of variation and the current chapter did not, the migratory distance thresholds presented earlier in this chapter would likely not be directly relevant to those studies.

3.1.4.2 Full face

The racial comparisons of the full face/head highlighted that Blacks and Caucasians do not differ in their generalised variances. This is contrary to the Rossion and Michel (2011) thought of African heterogeneity relative to Caucasians, and the perspective that Blacks may be more variable in facial morphology than Caucasians via hypodescent. However, considering the relatively close migratory distances of Africans and Caucasians, an absence of differences would not be unexpected.

Comparisons involving Native Americans were not significant. Looking at the ethnicities of persons in the ANSUR (Gordon et al., 1989), and the locations of Native American tribes (Waldman, 2006), measurements were from Native North Americans, and their approximate migratory distance would be of around 20,000 km. No significant differences would appear to be in alignment with Figures 3.4a and 3.4b, although it should be noted that Figure 3.4 concerns the standardised generalised variance, and this was calculated after principal components analysis. The issue of small sample sizes potentially leading to a Type II error (Stefan, 1999) could be a factor in explaining why Native Americans were not less variable than Africans and Caucasians, as there were few Native Americans in the data. Determinant ratios became numerically greater as migratory distance differences increased. This signals that something like the correlation between the generalised variance and migratory distance (found with the skeletal face) could also be evident in the full face, but, once more, the number of Native Americans leaves this idea far from settled.

3.2 Subjective variability

3.2.1 Introduction

Although the other-race effect is considered to have a perceptual basis (e.g., Hugenberg et al., 2010), previous research collectively casts doubt over whether there truly is a perceptual basis to the other-race effect (see Sections 1.2.5 and 1.2.6). The idea underlying the perceptual perspective of the other-race effect is that the other-race effect is a product of other-race faces subjectively seeming more similar than own-race faces (e.g., Goldstein, 1979). Therefore, one would expect that a poorer perceptual expertise with other-race faces (vs. own-race faces) would be associated with the other-race effect (DeGutis et al., 2013). Research concerning the other-race effect has explored whether other-race disadvantages in configural, featural, and holistic perceptual expertise are related to the other-race effect (DeGutis et al., 2013; Michel, Caldara, & Rossion, 2006; Zhao et al., 2014b). They are not (e.g., Zhao et al., 2014b), aside from one study when the other-race holistic disadvantage was calculated by regression (DeGutis et al., 2013), although another study using that same method found no such relationship (Zhao et al., 2014b).

Therefore, it may appear unlikely that the other-race effect is perceptually-driven (Section 1.2.5.3). Research has tested if there is a greater perceptual (i.e., subjective) similarity amongst other-race faces than own-race faces in paradigms (e.g., Goldstein & Chance, 1979; Lorenzino et al, 2018), for instance, with observers deciding whether two simultaneously-presented faces have the same identity (e.g., Prioetti et al., 2019). If the extent to which faces seem more different to each other when they are own-race than when they are other-race (i.e., other-race subjective homogeneity) relates to the other-race effect, then a perceptual basis of the other-race effect would be signalled.

Fortunately, the dataset of Bate, Bennetts et al. (2018b), which was used in Experiments 1 and 2 of Bate, Bennetts et al. (2018a), contains face recognition and face matching performance data for Caucasian and East Asian observers regarding Caucasian and East Asian faces. This data was used in Section 3.2 to determine whether other-race subjective homogeneity is related to the other-race effect for Caucasians and East Asians.

Regarding the other-race effect, Bate, Bennetts et al. (2018a) used two tasks: the CFMT long form (Russell, Duchaine, & Nakayama, 2009) and the East Asian CFMT (McKone et al., 2012). As for other-race subjective homogeneity, Bate, Bennetts et al. (2018a) employed another two tasks: the pairs matching test (Bate, Frowd et al., 2018) which features Caucasian faces (Bate, Bennetts, Murray, & Portch, 2020) and a new variant of that test featuring East Asian faces.

The original CFMT (Duchaine & Nakayama, 2006) has experimental trials that present Caucasian faces (McKone et al., 2012). In the CFMT, there are three different experimental phases (totalling 72 trials) in which observers attempt, per trial, to recognise a target facial identity in the presence of two distractor facial identities, with all three faces being presented simultaneously (Duchaine & Nakayama, 2006). In phase one, per trial, a target face image is shown, which is then followed by the same image amongst two distractor facial identities (Duchaine & Nakayama, 2006). For phase two, six target faces are initially shown head-on simultaneously, then, in each trial, observers decide which one of three faces (rotated from head-on or/and lit differently) has an identity matching any of the six targets (Duchaine & Nakayama, 2006). Phase three is similar to phase two, although Gaussian noise is used on the three faces, of which one face is the target (Duchaine & Nakayama, 2006). In a variant of the CFMT, called the CFMT long form, an additional phase (30 experimental trials) proceeds the third phase in order to make a more difficult task which separates typical face recognition ability from the above-typical (Russell et al., 2009). This fourth phase, for instance, presents more noise in the three simultaneously-presented faces (Russell et al., 2009). As with the CFMT, this lengthier variant has Caucasian faces (Bate, Bennetts et al., 2018a). An East Asian variant of the CFMT has been produced using Chinese faces (McKone et al., 2012).

As for the pairs matching test (Bate, Frowd et al., 2018), this task uses Caucasian faces (Bate et al., 2020), and 48 face pairs (24 male pairs, 24 female) are featured (Bate, Frowd et al., 2018). In 24 of these pairs, faces are of the same identity (*same* pairs), and, in the other 24 pairs, faces are of different identities (*different* pairs) (Bate, Frowd et al., 2018). To construct the *different* pairs, faces of different identities "were paired according to their perceived resemblance to each other" (Bate, Frowd et al., 2018, p. 6). On each trial, a face pair is shown, and observers make a *same*/*different* decision as to whether pairs share or differ in identity. Bate, Bennets et al. (2018a) produced a version of the pairs matching test featuring Asian faces. Regarding this version, faces were described as Asian (Bate, Bennetts et al., 2018a), and, in the dataset, data regarding these faces was labelled with *EA* (Bate, Bennets al., 2018b), which is an abbreviation for East Asian (e.g., Hedrick, 2008), therefore the Asian faces were specifically supposed to be East Asian.

Bate, Bennetts et al. (2018a) did not examine if other-race subjective homogeneity and the other-race effect are associated. Their research was focussed on super-recognisers, with super-recognisers being persons of strong facial recognition ability for own-race faces (Bate, Bennetts et al., 2018a). Alongside Caucasian super-recognisers, there were also Caucasian control, and Asian control participants. Controls were described as 35 Caucasians, and 28 Asians aged between 18 and 50 years (Bate, Bennetts et al., 2018a); from descriptions in Bate, Bennets et al. (2018a), it could be surmised that the Asian group was at least largely East Asian. In the dataset, the data for Asian participants was denoted with *EA* (Bate, Bennetts et al., 2018b); as with the faces in the pairs matching test, Asian participants were East Asian.

Results indicated that Caucasian super-recognisers were not better than East Asian controls at recognising East Asian faces (Bate, Bennets et al., 2018a). Bate, Bennets et al. (2018a) found that both the other-race effect and other-race subjective homogeneity were of a similar size for Caucasian super-recognisers as they were for Caucasian controls. Therefore, for Section 3.2, it may have seemed reasonable to combine Caucasian super-recognisers with Caucasian controls into an overall Caucasian group. However, whilst for controls (whether Caucasian or East Asian) the CFMT long form and East Asian CFMT were deployed in a counterbalanced fashion, Caucasian super-recognisers undertook the CFMT long form prior to the East Asian CFMT (Bate, Bennetts et al., 2018a). Therefore, for Caucasian super-recognisers, the magnitude of the other-race effect may have been attenuated (Bate, Bennetts et al., 2018a), which alludes to the possibility of an order effect (e.g., see Harris, 2008, pp. 156–157). This suggests that the present study (Section 3.2), in finding whether other-race subjective homogeneity and the other-race effect are associated, ought not to combine Caucasian super-recognisers and Caucasian controls into a singular group. The number of

Caucasian super-recognisers was eight (Bate, Bennetts et al., 2018a, 2018b), which would be too small for any meaningful analysis of whether other-race subjective homogeneity is associated with the other-race effect amongst Caucasian super-recognisers.

The dataset used in Bate, Bennetts et al. (2018a) contains various performance measures for the CFMT long form, East Asian CFMT, and both Caucasian and East Asian variants of the pairs matching test (Bate, Bennetts et al., 2018b); for Section 3.2, specifically the percentage of experimental trials in which decisions were correct was used from each Caucasian and East Asian control participant within each of the four tasks.

3.2.2 Results

Analysis occurred in SPSS 24.0 and 25.0. In order to calculate magnitudes of the other-race subjective homogeneity and the other-race effect, recognition accuracies for other-race faces (percent correct decisions) were subtracted from own-race accuracies per participant. These data were analysed in separate linear regressions for each participant race, with other-race subjective homogeneity and the other-race effect respectively as predictor and dependent variables. For the regression using Caucasian participants, the data of one participant was removed due to a considerable Studentised deleted residual value. Other-race subjective homogeneity predicted the other-race effect for Caucasian, but not East Asian, observers (Table 3.3 and Figure 3.5).

Table 3.3

Two Simple Regressions Testing whether Other-Race Subjective Homogeneity Predicted the Other-Race Effect for Caucasian and East Asian Participants

Participant race	*B*	*SE B*	CI	*β*	*p*
Caucasian	.68*	.19	.36, 1.12	.47	.003
East Asian	-.02	.17	-.37, .32	-.03	.89

Note. Caucasian $N = 34$; Caucasian regression $R^2 = .22$; East Asian $N = 28$; East Asian regression $R^2 = .001$; CI = 95% confidence interval; *$p < .05$ (Holm-Bonferroni-corrected, Gaetano, 2013, Holm, 1979). Concerning the regression with Caucasian participants, standardised predicted and residual values indicated that the assumption of homoscedasticity did not hold; bootstrapped estimates were generated for the standard error, confidence interval, and *p*-value (9,999 resamples, bias-corrected) and are presented.

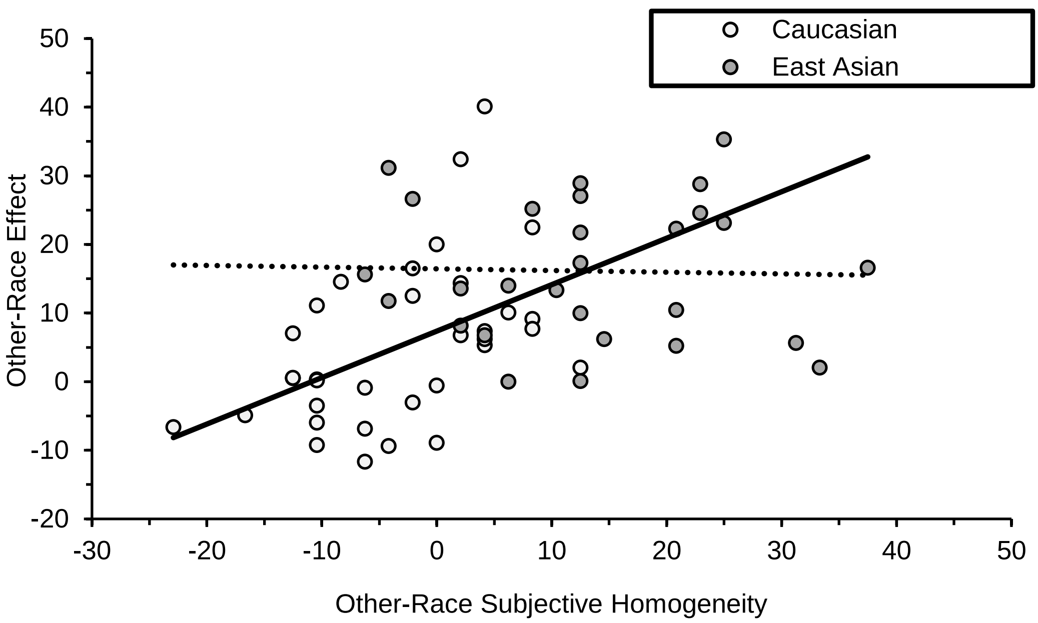


Figure 3.5. Magnitudes of other-race subjective homogeneities and other-race effects for Caucasian and East Asian observers. The unbroken line is the linear trend line for Caucasian observers; the dashed line is the linear trend line regarding East Asian participants.

3.2.3 Discussion

Section 3.2 examined if there is a relationship concerning other-race subjective homogeneity and the other-race effect. Indeed, despite other-race subjective homogeneity being a key tenet of theorisations on the other-race effect, previous research had not directly explored their supposed link (Section 1.2.6.2). Furthermore, a general lack of relationships between other-race expertise disadvantages and the other-race effect would suggest that the other-race effect is not perceptually-based (Section 1.2.5). Whilst a connection regarding other-race subjective homogeneity and the other-race effect was present amongst Caucasian observers, it was not evident for East Asian observers. This suggests that there is a considerable perceptually-based component to the other-race effect for Caucasian observers, unlike for East Asian observers. Therefore, the other-race effect for East Asian observers may be more driven by processes occurring beyond perceiving the face (i.e., during storage in memory or retrieval from memory) than the other-race effect regarding Caucasian observers.

3.3 General discussion

All in all, Chapter 3 made novel contributions concerning variability in the context of the other-race effect. Analysis concerned objective variability (i.e., facial morphological variability analysis of the skeletal face and the full face) and subjective variability (i.e., testing whether other-race subjective homogeneity predicts the other-race effect). There are clear declines in different measures of morphological variability as migratory distance advances for males and females, although thresholds for unequal variability are large. In that context, it should not be surprising that Blacks, Caucasians, and Native Americans were equal in the generalised variance of the full face, yet, as previously mentioned, the sample sizes for Native Americans were slight. Regarding subjective variability, for Caucasian observers, other-race subjective homogeneity did predict the other-race effect in terms of magnitudes, yet the same cannot be said for East Asian observers. This opens the possibility of the weighting of perceptually- vs. storage- vs. retrieval-driven contributions to the other-race effect differing with observer race.

3.3.1 Routes to variability

It is worth considering whether the facial variability for a race calculated from an estimated average of migratory distances of ethnic groups (the analyses with Howells' data by and large) actually relates to variability calculated directly at the race level (essentially, ANSUR and ANSUR II face/head comparisons). The former can be called the indirect route, the latter being the direct route. With continental groupings, each comprised of three populations in the Howells cranial data, the indirect route (a mean of within-population variances) can differ from the direct continental variance, because the direct route is influenced by differences between populations, and these differences would enlarge with the distance separating populations (Relethford, 2001). Such a pattern could happen in faces for not only the variance, but also for the generalised variance and pattern variability as well. Hence the direct route would be affected by differences between populations representing the race, and some discrepancy should certainly be expected between the routes.

3.3.2 Magnitudes

Inspection of Figure 3.5 could suggest that with a higher magnitude of other-race subjective homogeneity or other-race effect, the association regarding other-race subjective homogeneity and the other-race effect diminishes; indeed, magnitudes of other-race subjective homogeneity were smaller for Caucasian ($M = -3.25$, $SD = 8.29$) than East Asian observers ($M = 13.32$, $SD = 11.61$), $t(60) = -6.54$, $p < .001$, and the same pattern occurred regarding the other-race effect for Caucasians ($M = 5.17$, $SD = 11.88$) and East Asians ($M = 16.12$, $SD = 10.04$), $t(60) = -3.87$, $p < .001$. Therefore, other-race subjective homogeneity not predicting the other-race effect for East Asians could possibly be explained by their stronger other-race subjective homogeneity and other-race effect.

Results in Hancock and Rhodes (2008) would align with there being a perceptually-driven contribution to the other-race effect for Caucasians and East Asians (Section 1.2.6). It is not clear why there would appear to be some discrepancy between their results and those of Section 3.2.2 concerning East Asian observers. In Hancock and Rhodes (2008), East Asian participants were described as having "good variation in contact" with Caucasians (p. 48), whereas, in Bate, Bennetts et al. (2018a), East Asians in the Bate, Bennetts et al. (2018b) data were stated to have grown up in an Asian country and remained there most of their life so far. If the connection pertaining to other-race subjective homogeneity and the other-race effect diminishes at stronger levels of either phenomenon, it could be speculated that East Asians in Bate, Bennetts et al. (2018b) may have had less other-race experience (and a greater other-race experience disadvantage) than East Asians in Hancock and Rhodes (2008). Consequently, East Asians in Bate, Bennetts et al. (2018b) data may have exhibited a greater other-race subjective homogeneity and other-race effect than those of Hancock and Rhodes (2008), which caused there to be evidence favouring a perceptual contribution to the other-race effect for East Asians (other-race disadvantage in the face inversion effect predicted the other-race effect) in Hancock and Rhodes (2008), whilst other-race subjective homogeneity and the other-race effect were not linked when using the Bate, Bennetts et al. (2018b) data.

3.3.3 Conclusion

Regarding objective variability, the presence of declines (and thresholds) point to the possibility of racial/ethnic differences in the morphological diversity of the face serving a moderating role in the other-race effect. As for subjective variability, results indicated that whilst other-race subjective variability is a determiner of the other-race effect for Caucasian observers, it is not for East Asian observers, thereby implying that the other-race effect may not universally be a perceptually-driven phenomenon.

Chapter 4: GENDER CATEGORISATION

4.1 Introduction

The purpose of Experiment 3 was to explore the other-race gender effect, particularly to see whether patterns that are evident concerning the other-race effect and its mechanisms also occur for the other-race gender effect and its underlying processes. Generally, the magnitude of the other-race holistic disadvantage has no association with the other-race effect (Horry et al., 2015; Michel, Caldara, & Rossion, 2006; Michel, Rossion, et al., 2006; Zhao et al., 2014b), which implies that something aside from the other-race holistic disadvantage leads to the other-race effect. Nonetheless, previous research has found that alternative other-race expertise disadvantages (featural and configural) have no connection to the other-race effect (DeGutis et al., 2013; Zhao et al., 2014b). As for gender categorisation, an other-race disadvantage for featural, but not holistic, processing has been present (Zhao & Hayward, 2010). Still, prior research had not determined whether other-race expertise disadvantages have associations with the other-race gender effect. For a better understanding of the other-race gender effect, Experiment 3 aimed to find if the sizes of other-race expertise disadvantages relate to the other-race gender effect; no relationships would suggest some alternative factor produces the other-race gender effect.

It is unknown if the difficulties in categorising the gender of other-race faces arise due to other-race faces seeming to be more male or female than own-race faces (as, perhaps, other-race faces may hold a stronger resemblance to one own-race gender over another own-race gender) (Section 1.3.2). Therefore, the subjective boundaries between male and female genders were explored for own-and other-race faces; finding no correlation between the other-race gender categorisation bias and the other-race gender effect would suggest that the bias has no bearing on the effect.

Regarding identity, it seems uncertain if experience relates to other-race expertise disadvantages, whether that expertise be holistic or non-holistic (see Section 1.2.3.2.2). Previous research was yet to examine relationships between experience and other-race expertise disadvantages (or categorisation biases) concerning gender. Furthermore, whilst prior research has demonstrated a link between experience and the other-race effect (e.g., Hancock & Rhodes, 2008), none had explored whether experience correlates with, or predicts, the other-race gender effect.

Experiment 3 did not focus on *other-race experience*, but instead (as in Experiment 2) the *other-race experience disadvantage* (Figure 2.2). Experiment 3 used non-valenced IRCQ items in order to measure the other-race experience disadvantage, although participants did also complete valenced items. Non-valenced items were analysed over valenced items as results in Experiments 1-2 would suggest that non-valenced items had greater internal consistency than the valenced items.

The aperture paradigm (Murphy & Cook, 2017) was tailored to quantify the other-race holistic disadvantage and the other-race local processing[32] disadvantage for gender categorisation. A gender categorisation task provided a metric of the other-race gender categorisation bias and the other-race gender

[32] Local face regions would contain featural, and some configural, information. Aperture viewing could perhaps disrupt larger configural relations (e.g., the distance between the eyes and the mouth).

effect. Correlational analyses were conducted between experience, expertise, the other-race gender categorisation bias, and the other-race gender effect to shed light on mechanisms underlying the other-gender effect. Interestingly, previous research suggested that the other-race gender effect was determined by males being more dissimilar to females within the Caucasian race than amidst Chinese East Asians (Zhao & Bentin, 2008); in Experiment 3 not only were faces and participants Caucasian and East Asian, but they were also Black. This was a step to finding whether observers *generally* find it easier/difficult to distinguish between male and female genders within a race than other races as results across participant groups would suggest if there is higher/lower physical facial dissimilarity between males and females in one race compared to others in general.

4.2 Method

4.2.1 Participants

Sample size was set bearing in mind previous research which examined face processing via psychometric functions (e.g., Brewer, Biotti, Bird, & Cook, 2017; Shah, Bird, & Cook, 2015), the aperture paradigm (Murphy & Cook, 2017), and research which explored links between experience, expertise, and the other-race effect (e.g., Bukach et al., 2012; Hancock & Rhodes, 2008; Davis et al., 2016; Zhao et al., 2014b); the aim was to have a final group sample size per race (after the removal of non-engaged participants and outliers) of 16 to 20 for the categorisation and aperture tasks, and a minimum of 48 across the three races (East Asian, Black, and Caucasian) for the correlation analyses.

There is not only an effect of participant age on facial identity processing (Megreya & Bindemann, 2015) but also on facial gender discrimination (Carbon, Grüter, & Grüter, 2013), hence an age range for participation was set conservatively at 18-50 years old. There was a total of 73 participants (31 males, 42 females; $M_{\text{age}} = 26.93$, $SD_{\text{age}} = 8.40$), of whom 70 were monoracial (22 Caucasians, 24 East Asians, and 24 Blacks). The experiment was approved by an ethics committee of City, University of London, which was the institution at which all testing occurred.

4.2.2 Materials

4.2.2.1 Questionnaires

Two questionnaires were utilised. The PI20 was used in its capacity as an indicator of developmental prosopagnosia (Shah et al., 2015). Relative to controls, poorer facial gender categorisation has been found regarding developmental prosopagnosic group samples in one experiment (Esins, Schultz, Stemper, Kennerknecht, & Bülthoff, 2016) but not in four others (Chatterjee & Nakayama, 2012; DeGutis, Chatterjee, Mercado, & Nakayama, 2012; Dobel, Bölte, Aicher & Schweinberger, 2007; Le Grand et al., 2006). Nonetheless, in one of those latter four, which featured 18 developmental prosopagnosics, it was observed that five developmental prosopagnosics "were considerably lower than normal" (Chatterjee & Nakayama, 2012, p. 494) (these were lower than 1.5 *SD*s from the mean of typically developing persons) and that "it is conceivable that a minority of" developmental prosopagnosic "subjects may have facial gender perception that is quite below normal" (p. 494). In another of the four, gender categorisation of developmental prosopagnosics and controls alike was at ceiling (100% correct), i.e., the task may not have been sufficiently arduous for a difference to be apparent (Dobel et al., 2007). In the current experiment, conservatively, any Caucasian who scored 65 or above was to be removed (none scored at least that number). The IRCQ (Chapter 2) was employed to measure contact with Caucasians, East Asians, and Blacks during the life stages (years of age) of 0-6, 6-12, and 12-18.

Ordinal alphas (in place of Cronbach's alpha due to Likert data usage and skew, Zumbo et al., 2007) were derived with the SPSS R-Menu (Basto & Pereira, 2012) using the monoracial participants (N = 70); there was strong internal consistency amongst the PI20 items, ordinal alpha = .92, and sets of IRCQ non-valenced items (Table 4.1). Analysis in Experiment 1 indicated redundancy amongst items for Caucasian experience; for Experiment 3, item redundancy would seem to have occurred more generally.

Table 4.1

Ordinal Alphas of the Non-Valenced IRCQ Items for Monoracial Participants

Life stage (years)	Black	Caucasian	East Asian
0-6	0.94	0.97	0.98
6-12	0.95	0.97	0.98
12-18	0.96	0.94	0.95
Mean (0-18)	0.95	0.96	0.97

4.2.2.2 Face stimuli

Ninety faces of persons who self-categorised as White, Black, or Asian were selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), http://faculty.chicagobooth.edu/bernd.wittenbrink/cfd/index.html. Thirty faces were chosen from each of the three categories. In each category, half of the faces (in terms of gender) were female and the remainder were male. Pictures of those faces were taken in the USA (Ma et al., 2015). The Chicago Face Database contains various ratings/perceptions from observers regarding the photographed faces, for instance on the perceived age (averaged across observers), race (proportion who rated the face as White, Black, etc.), and gender (proportion that categorised the face as belonging to a male, or a female, with respect to gender).

In the instructions given to the observers who rated the faces, certain judgements were stipulated to be made with respect to a face in the context of other faces in the USA (these instructions are found in the Chicago Face Database spreadsheet). Therefore, it seems reasonable to assume that raters were in (or of) the USA. It has been noted that, in the USA, Asian "is mostly used to denote people of far Eastern origins, for example, Chinese, Japanese, and Filipinos" (Bhopal, 2004, p. 442), i.e., East Asian (e.g., Risch et al., 2002). Furthermore, “[i]n popular and informal understanding in the US today, an ‘Asian’ … is someone with a particular phenotype … this phenotype is most closely associated with East Asians” (Kibria, 1998, p. 949); inspection of the faces and racial ratings in the Chicago Face Database strongly indicates that observers were generally using *East Asian* as the basis for *Asian* categorisations. Moreover, interviews with second-generation Asian (Chinese, Indians, etc.) Americans (Park, 2008) would indicate that the subjective Asian typicality of Asian faces in the Chicago Face Database would predominantly reflect the extent of perceived East Asian racial appearance: in Park (2008), it was found that "East Asian ethnic groups, namely Chinese, Japanese, and Korean, are often the first groups to come to mind when reflecting on the term *Asian American*" and "most respondents usually continued their explanation and expanded the number of groups to include, first, Southeast Asian groups (such as the Vietnamese or Filipino)" (Park, 2008, p. 548) (i.e., still including the East Asian race, Risch et al., 2002) "followed by South Asians (such as Indian and Pakistani)" and "[t]his pattern was fairly consistent across responses, even across ethnic groups" (Park, 2008, p. 548). Additionally, in Experiment 4, participants were able to racially categorise (Caucasian and East Asian categorisations) faces from morph continua between norms of prototypical White and Asian faces within male and female genders, and the faces which were used to make those norms were from the Chicago Face Database (see Chapter 5); for faces which were 5% away from an Asian norm on a White-to-Asian continuum, East Asian categorisations (for engaged participants) were above chance for faces of males (Mdn = 20), $T(47)$ = 1,176, $p < .001$, and females (Mdn = 20), $T(46)$ = 1,128, $p < .001$.

Selected faces had average age ratings of between 18 and 40 years old, and each face had race-ratings which matched self-categorised *race* on at least 82% of occasions (for each face) for males and females (M_{males} = 95.39, SD_{males} = 4.89; $M_{females}$ = 94.94, $SD_{females}$ = 4.87),[33] and gender-ratings were congruent with self-categorised gender on at least 95% and 85% of instances for males and females respectively (M_{males} = 99.49, SD_{males} = 1.12; $M_{females}$ = 98.72, $SD_{females}$ = 2.64).

The 90 faces were used to construct greyscaled stimuli for the categorisation and aperture tasks. For Experiment 3, a programme in MATLAB (Adams, Gray, Garner, & Graf, 2010) was employed for the purpose of deriving

---

[33] In the context of a 3 (self-categorised race: Asian, Black, White) x 2 (self-categorised gender: male, female) ANOVA run on race-ratings, there was a racial difference, $F(2, 84) = 8.53$, $p < .001$, $\eta_p^2 = .17$, with those ratings being lower for Asians (M = 92.38, SD = 5.04) than Blacks (M = 96.33, SD = 3.61), p = .002, and Whites (M = 96.79, SD = 4.67), $p < .001$; given the general heterogeneity of an *Asian* categorisation (Bhopal, 2004; Park, 2008), this racial difference should not be surprising. There was neither an effect of gender $F(1, 84) = .22$, $p = .64$, $\eta_p^2 = .003$, nor an interaction between race and gender, $F(2, 84) = .19$, $p = .83$, $\eta_p^2 = .004$.

morph continua from the Chicago Face Database faces. Within each race, a prototypical male and female face were each formed from 15 identities. Per race, faces from 20%, 30%, 40%, 50%, 60%, 70%, and 80% on a male-female continuum were selected. These faces were then bounded within an oval to occlude the ears and head hair, and they were the stimuli which were presented in the categorisation task (see Figure 4.1 for an example).

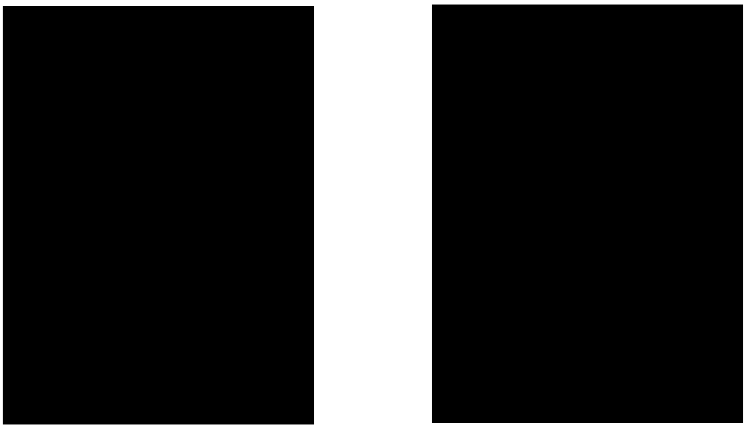


Figure 4.1. Example stimuli. The face on the left is 20% along the Caucasian male-female morph continuum, whilst the face on the right is 80% along. (Faces removed due to copyright.)

Race norms (three) were created from 15 male faces and 15 female faces per race. Morph continua were then formed between each of the original 90 identities and their race norm. The point at 15% along each continuum (from the race norm to the identity) was the stimulus face which was presented in the aperture task. Stimulus faces were surrounded by an oval. As in Murphy and Cook (2017), on aperture trials, the aperture (12.50% of the height of the full stimulus) gradually moved down the face from top to bottom. The time taken for the aperture to traverse the face was 8,000 ms.

4.2.3 Procedure

Each testing session occurred at City, University of London. After giving informed consent, and then having completed questionnaires (PI20, then the IRCQ), participants engaged in the categorisation task, and then the aperture task. Programmes for both face tasks were written in MATLAB via the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Any participant who missed any items (or otherwise did not follow instructions) on the PI20 or IRCQ was requested to correctly complete those items after the categorisation or aperture tasks.

Trials within both face tasks were presented in randomised orders. In the categorisation task, per trial, a face from a male-female morph continuum was presented for 500 ms, after which participants categorised the face as male, or as female. Six practice trials were followed by 420 experimental trials. Of the experimental trials, per race (three races), each morph level (seven levels) was presented 20 times.

The aperture task consisted of full-view and aperture-view trials. Before each type of trial, one of two symbols appeared which cued participants to the trial being a full-view trial (a circle) or an aperture trial (an arrow pointing downward). There were six practice trials and subsequently 360 experimental trials. Half of the experimental trials were full-view trials (180 trials), whilst the remainder were aperture-view trials. Within the experimental trials, each face (90 identities) was displayed four times, two times in the full-view condition and twice in the aperture-view condition. Presentation durations of a face were 2,000 ms and 8,000 ms in full-view and aperture-view conditions respectively. After the presentation, participants categorised the gender of that face (male or female) (see Figure 4.2). Following this second computer-based task, participants were debriefed. The session had a duration of approximately 80 minutes.

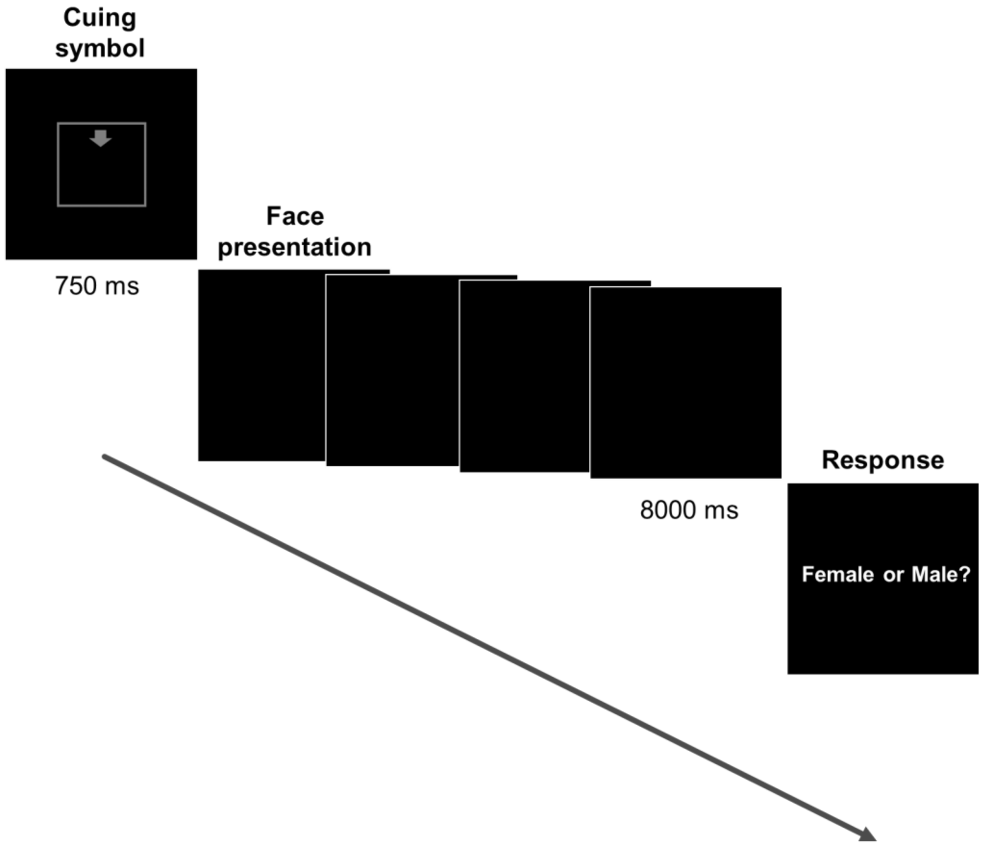


Figure 4.2. An illustration of an aperture-view trial in the aperture task. (Faces removed due to copyright.)

4.2.4 Data preparation

Data analysis was limited to monoracial persons. As in Experiments 1 and 2, outliers were defined as values outside 3 × the interquartile range from outside the upper and lower quartiles. Outliers were with respect to values within a participant race (e.g., for Caucasians, East Asians, and Blacks separately) except for when participant race groups were pooled together in correlational analyses of own- versus other-race differences. Additionally, across races of participants, Studentized residuals were employed to define outliers (values exceeding |3|).

For the categorisation task, a psychometric function (cumulative Gaussian) was derived per participant regarding each stimulus race (and the associated decision noise and PSE) in MATLAB using code from Yarrow (2018). As in Murphy and Cook (2017), lower decision noise (i.e., a steeper function slope) was interpreted as demonstrating a greater ability to perceive gender differences. Whilst a gender PSE of less than 50% indicated a bias for perceiving faces as being male, a PSE above 50% suggested a bias for seeing faces as female. For the aperture task, response sensitivity (d') was used as the measure of accuracy. Hits were defined as categorisations of the faces as *male* when the faces were male, and false alarms were *female* categorisations to male faces.

For the categorisation task, the identification of non-engaged participants was performed in MATLAB and from procedures in Yarrow (2018): model fit was compared (as the difference in deviance) between a straight line (i.e., participants who merely guessed or solely gave one response in a condition) and the more complex model (i.e., the one used to generate the PSE and decision noise), and the threshold for the more complex model being a better fit (indicating engagement) was found using the χ^2 distribution with one degree of freedom at an alpha-level of .05. As for the aperture task, non-engaged participants were defined as persons who responded uniformly in any combination of race and viewing condition, e.g., responding *female* to all Black faces which were presented through an aperture. Maximum hits (100% hit-rate) or minimum false alarms (i.e., 0% false-alarm-rate) were replaced with .5/n or 1 - (.5/n), with n as the number of relevant trials (Stanislaw & Todorov, 1999).

Magnitudes of other-race disadvantages (in experience, local processing, holistic processing, and categorisation) were calculated by taking the average of other-race performances (or experience) and subtracting them from own-race performance/experience. To calculate the other-race gender categorisation bias, absolute differences in the categorisation boundaries (PSEs) between the own-race group and each of the two other-race faces were averaged. Experience was measured as the average of non-valenced contact items.

Data were analysed in SPSS 24.0 and 25.0. Holm-Bonferroni corrections were applied to regulate the family-wise error rate (Gaetano, 2013; Holm, 1979). Any significant interaction involving the race of participants was followed by one ANOVA per participant race, with a Bonferroni correction applied to each ANOVA (i.e., *p*-values multiplied by 3),[34] with Holm-Bonferroni corrections then used on subsequent pairwise comparisons.

4.3 Results

The data of multiracial participants (three) were not analysed. Regarding non-engaged monoracial participants, there were 16 in the categorisation task [35] and nine in the aperture task, each of whom was removed. Due to technical problems, three monoracial participants were unable to complete the aperture task. As for outliers, there was one outlier concerning the PSE and one for the decision noise on the categorisation task, and, for the correlations (Section 4.3.4), one concerning the other-race gender effect; outliers were removed.

4.3.1 Other-race gender effect

A 3 (stimulus race [Caucasian, East Asian, Black]) × 3 (participant race [Caucasian, East Asian, Black]) mixed ANOVA was run on decision noise. A main effect of stimulus race was apparent, $F(2, 100) = 33.40$, $p < .001$, $\eta_p^2 = .40$.

[34] Although Type II errors are less likely with the Holm-Bonferroni correction than Bonferroni (Abdi, 2010), the Bonferroni adjustment was used for practicality.

[35] The relatively high number is addressed in Section 6.3.2.1.

Performance was better with Caucasian faces (*M* = 16.02, *SD* = 6.99) than East Asian faces (*M* = 27.75, *SD* = 14.26), $p < .001$, and Black faces (*M* = 28.03, *SD* = 15.42), $p < .001$, with performance for the latter two races not differing, $p = .88$. Two participants were outliers, however they were retained as their data had no impact on outcomes.

The interaction between stimulus race and participant race, $F(4, 100) = 4.03$, $p = .004$, $\eta_p^2 = .14$ (Figure 4.3) was explored with a one-way ANOVA for each race of participant. Amongst Caucasian participants a main effect of stimulus race was evident, $F(2, 34) = 14.42$, $p < .001$, $\eta_p^2 = .46$; other-race gender effects occurred regarding decision noise being lower for own-race faces (*M* = 16.93, *SD* = 6.95) than East Asian faces (*M* = 30.76, *SD* = 14.88), $p = .002$, and Black faces (*M* = 33.39, *SD* = 20.77), $p = .003$; decision noise was equal for East Asian and Black faces, $p = 1.00$. As for East Asian participants, there was also a main effect of stimulus race, $F(1.19, 21.38) = 9.68$, $p = .01$, $\eta_p^2 = .35$ (Greenhouse-Geisser $\varepsilon = .59$), however, no other-race gender effect occurred as performance with own-race faces (*M* = 21.99, *SD* = 9.19) matched that of Black faces (*M* = 28.03, *SD* = 14.19), $p = .28$, and gender categorisation was worse for own-race faces than it was for Caucasian faces (*M* = 16.92, *SD* = 7.42), $p = .050$ ($< .05$). As with Caucasian participants, gender categorisation was more proficiently with Caucasian faces than Black faces, $p < .001$. For Black participants, a main effect of stimulus race occurred, $F(1.13, 16.89) = 16.64$, $p = .002$, $\eta_p^2 = .53$ (Greenhouse-Geisser $\varepsilon = .56$). Gender categorisation was better with Caucasian faces (*M* = 14.20, *SD* = 6.41) than own-race faces (*M* = 22.67, *SD* = 7.60), $p < .001$, and East Asian faces (*M* = 30.49, *SD* = 17.93), $p = .003$, and an other-race gender effect occurred, with gender categorisation performance being greater with own-race faces than East Asian faces, $p = .049$.

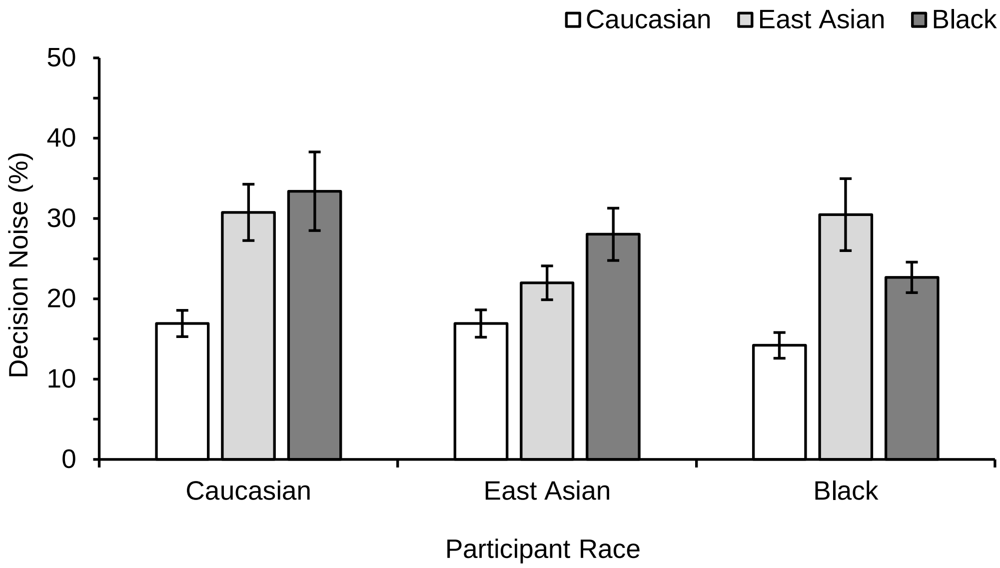


Figure 4.3. Mean decision noise for own- and other-race faces for each participant group with respect to gender categorisation. Each bar represents decision noise for a participant race (Caucasian $N = 18$; East Asian $N = 19$; Black $N = 16$) with respect to a stimulus race. Errors bars are for the standard error of the mean.

4.3.2 Other-race expertise disadvantage for gender

Results of the aperture task were analysed via a 2 (viewing condition [full-view, aperture-view]) × 3 (stimulus race [Caucasian, East Asian, Black]) × 3 (participant race [Caucasian, East Asian, Black]) mixed ANOVA. An effect of viewing condition was clear, $F(1, 55) = 210.02$, $p < .001$, $\eta_p^2 = .79$, with gender being categorised more accurately in the full-view condition ($M = 1.26$, $SD = .49$) than the aperture-view condition ($M = .56$, $SD = .44$). There was a main effect of stimulus race, $F(2, 110) = 45.66$, $p < .001$, $\eta_p^2 = .45$; gender was categorised more accurately for Caucasian faces ($M = 1.19$, $SD = .60$) than for East Asian faces ($M = .75$, $SD = .39$), $p < .001$, and Black faces ($M = .79$, $SD = .45$), $p <$

.001. Gender categorisation was as accurate for East Asian faces as it was amongst Black faces, $p = .43$.

There were no interactions between viewing condition and stimulus race, $F(2, 110) = .51$, $p = .60$, $\eta_p^2 = .01$, and between viewing condition, stimulus race, and participant race, $F(4, 110) = .50$, $p = .73$, $\eta_p^2 = .02$, which suggests that the advantage for Caucasian faces occurred in both viewing conditions, and that only Caucasian participants exhibited the other-race local disadvantage for gender. Given that the magnitude of the aperture effect was the same across stimulus races, and that there was no three-way interaction, results indicate that the aperture effect was of a similar size across stimulus races for participant race groups, i.e., there was no other-race holistic disadvantage for gender (Figure 4.4).

There was no interaction concerning race of stimulus and participant race, $F(4, 110) = 1.36$, $p = .25$, $\eta_p^2 = .05$. In the context of there being no three-way interaction (stimulus race × participant race × aperture), this suggests that gender processing of local face areas was better for Caucasian faces than the other two races across participant race groups, i.e., Caucasian participants exhibited other-race local processing disadvantage for gender, unlike East Asian and Black participants.

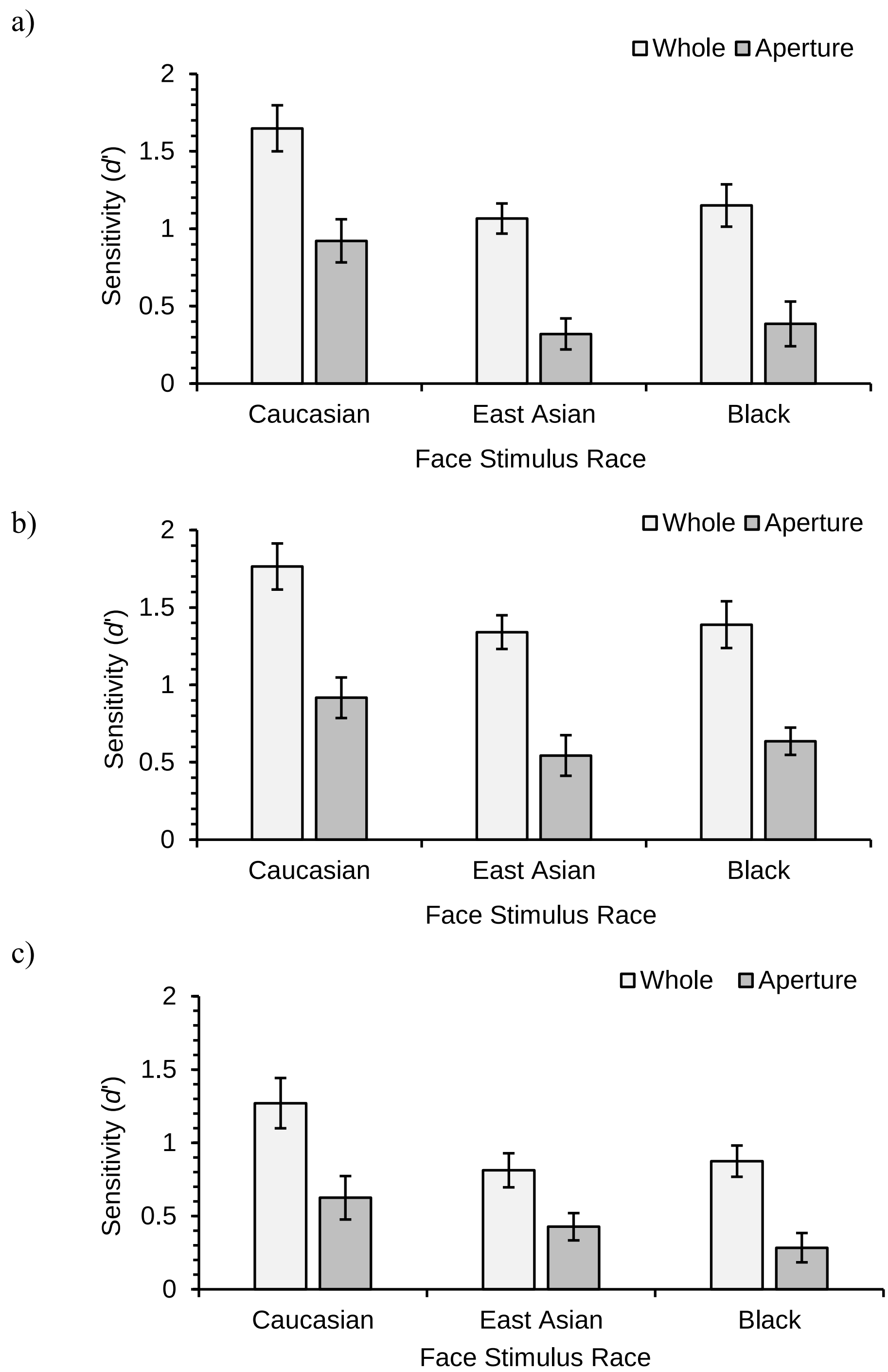


Figure 4.4. Mean gender categorisation accuracies (*d*’) for a) Caucasian participants ($N = 18$), b) East Asians ($N = 20$), and c) Blacks ($N = 20$), in whole and aperture conditions. Error bars are in regard to the standard error of the mean.

4.3.3 Other-race gender categorisation bias

A 3 (stimulus race [Caucasian, East Asian, Black]) × 3 (participant race [Caucasian, East Asian, Black]) mixed ANOVA was employed regarding the PSEs. A main effect of stimulus race was apparent, $F(1.81, 90.32) = 16.43$, $p < .001$, $\eta_p^2 = .25$ (Huynh-Feldt $\varepsilon = .90$). Pairwise comparisons showed that PSEs were higher (more female) for Black faces ($M = 62.56$, $SD = 13.99$) than Caucasian faces ($M = 48.63$, $SD = 8.38$), $p < .001$, and higher for East Asian faces ($M = 62.73$, $SD = 19.47$) than Caucasian faces, $p < .001$, whilst PSEs were the same for Black and East Asian faces, $p = .96$.

An interaction between stimulus race and participant race, $F(3.61, 90.32) = 2.59$, $p = .047$, $\eta_p^2 = .09$, (Figure 4.5) was proceeded by one-way ANOVAs at each participant race. For Caucasian participants there was a main effect of stimulus race, $F(2, 36) = 5.07$, $p = .034$, $\eta_p^2 = .22$; they exhibited an other-race gender categorisation bias with respect to Black faces ($M = 65.85$, $SD = 19.69$) being perceived as more female than own-race faces ($M = 48.36$, $SD = 7.72$), $p = .023$, yet no other-race categorisation bias occurred concerning East Asian faces ($M = 60.48$, $SD = 21.24$), $p = .15$ (Figure 4.4). PSEs of Caucasian observers did not differ between other-race face categories, $p = 1.00$. Regarding East Asian participants, there was no main effect of stimulus race, $F(2, 34) = 2.92$, $p = .20$, $\eta_p^2 = .15$, and indeed there was not any other-race gender categorisation bias or difference between other-race groups ($M_{\text{Caucasian}} = 49.37$, $SD_{\text{Caucasian}} = 7.74$; $M_{\text{East Asian}} = 54.34$, $SD_{\text{East Asian}} = 15.52$; $M_{\text{Black}} = 57.69$, $SD_{\text{Black}} = 11.02$), ps > .05. As for Black participants, a main effect of stimulus race was evident, $F(1.22, 18.24) = 12.01$, $p = .005$, $\eta_p^2 = .44$ (Greenhouse-Geisser $\varepsilon = .61$). They perceived East Asian faces ($M = 73.37$, $SD = 21.05$) as equally male/female as they did own-race faces ($M = 64.12$, $SD = 6.76$), $p = .32$. Interestingly, own-race faces appeared more female than Caucasian faces ($M = 48.17$, $SD = 9.66$) (an *own-race* gender categorisation bias), $p < .001$. Additionally, Black participants perceived East Asian faces as being more female than Caucasian faces, $p = .011$.

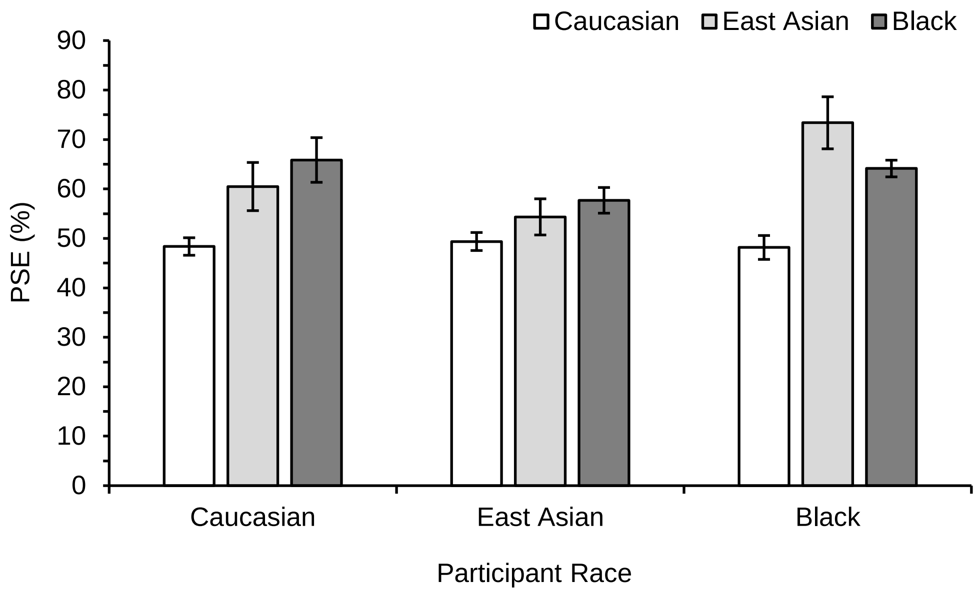


Figure 4.5. Gender categorisation PSEs for each participant race (Caucasian $N = 19$; East Asian $N = 18$; Black $N = 16$) per stimulus race. Error bars pertain to the standard error of the mean.

One-sample t-tests were used find whether PSEs for own-race faces differed from the point which is 50% of the distance between male and female stimulus norms. PSEs for own-race Caucasian faces ($M = 48.36$, $SD = 7.72$) were not different from 50%, $t(18) = -.93$, $p = .50$, and neither were PSEs for own-race East Asian faces ($M = 54.34$, $SD = 15.52$), $t(17) = 1.19$, $p = .50$, however, PSEs for own-race Black faces ($M = 64.12$, $SD = 6.76$) were greater than 50%, $t(15) = 8.35$, $p < .001$, i.e., they exhibited a bias for perceiving own-race faces as female.

4.3.4 Correlations

Correlational analyses were run between magnitudes of the other-race disadvantages of interest. The other-race gender effect had an association with the other-race local disadvantage for gender, $r(44) = -.55$, $p < .001$, and the other-race gender categorisation bias, $r(51) = -.55$, $p < .001$, yet not the other-race holistic disadvantage for gender, $r(44) = -.10$, $p = 1.00$. The other-race experience disadvantage was not correlated with the other-race holistic disadvantage for gender, $r(56) = -.02$, $p = 1.00$, the other-race local disadvantage for gender, $r(56) = .22$, $p = .41$, the other-race gender categorisation bias, $r(52) = .17$, $p = .66$, or the other-race gender effect, $r(51) = -.32$, $p = .091$.

4.4 Discussion

In Experiment 3, the other-race gender effect was exhibited by Caucasian participants (with respect to Black, and East Asian faces) and Black participants (for East Asian faces), but not East Asian participants. Previous research on the other-race gender/sex effect had exclusively used Caucasian and East Asian stimuli and participants (O'Toole, et al., 1996; Zhao & Bentin, 2008; Zhao & Hayward, 2010), and suggested that the presence of the other-race gender effect is stimulus-driven, with there potentially being larger dimorphism within Caucasians than Chinese East Asians (Zhao & Bentin, 2008). In the current experiment, gender was indeed categorised better amongst Caucasian faces than the other two races, and this occurred across participant race groups. This indicates that Caucasians faces could be characterised as having a larger distance between their male and female genders, rather than characterising East Asians as having less male-female facial dissimilarity in comparison to other races. Interestingly, Black participants had less decision noise for Black faces than East

Asian faces (i.e., an own-race gender effect) unlike Caucasian and East Asian participants, which suggests that their particular other-race gender effect was not stimulus-driven.

4.4.1 Expertise

Aligning with previous research (Zhao & Hayward, 2010), the other-race holistic disadvantage for gender did not occur. Regarding processing outside of holistic, prior research (using Caucasian and East Asian faces) demonstrated an other-race featural deficit for gender categorisation amidst East Asian participants (Zhao & Hayward, 2010). In the current experiment, however, results indicated that the gender of Caucasian faces was categorised best on the basis of local regions, regardless of the race of participants (i.e., this may be stimulus-driven, as with the overall face). Therefore, results suggest that Caucasians, but not non-Caucasians, exhibited the other-race local processing disadvantage for gender. Hence, the other-race local processing disadvantage for gender would still seem to be determined (at least partially) by Caucasians having a larger objective distance between males and females (in local facial traits) than other races.

Interestingly, the other-race gender effect did not correlate with the other-race holistic disadvantage for gender, but it did correlate with the other-race local processing disadvantage; assuming a causal path from expertise to perception, this would favour a theory of the local processing disadvantage contributing to the size of the other-race gender effect.

4.4.2 Bias

Caucasian observers exhibited an other-race gender categorisation bias for Black faces (perceiving them as more female than own-race faces). However, Blacks perceived own-race faces as being more female than Caucasian faces, and they had a bias for perceiving own-race faces as female (as show by their own-race PSE exceeding 50%).

As for why Blacks exhibited this bias concerning own-race faces, stimulus construction possibly caused a loss of information which is ordinarily useful for gender categorisation. For instance, the jaw affects gender categorisation (Brown & Perrett, 1993), yet the oval removed part of the jaw, including the jaw line. Therefore, perhaps regarding Black faces in particular, the way in which facial information is weighted for categorising gender may have been affected, and resulted in a general bias and potentially heightened difficulty in gender categorisation. It is worth noting that oval occlusions have been used in prior research on the other-race gender effect using Caucasian and East Asian faces (Zhao & Hayward, 2010), and in research on gender categorisation outside of the other-race gender effect (e.g., Cellerino, Borghetti, & Sartucci, 2004; Prete, Fabri, Foschi, & Tommasi, 2016) such as with Caucasian faces (e.g., DeGutis et al., 2012); the occurrence of Blacks having a bias for perceiving own-race faces as female would suggest caution in applying ovals when studying the gender categorisation of Black faces.

Nonetheless, there was a correlation between magnitudes of the other-race gender categorisation bias and the other-race gender effect, which aligns with the idea that bias may influence the magnitude of the other-race gender effect.

4.4.3 Experience

No association between the other-race experience disadvantage and the other-race gender effect was evident. Whilst it has been suggested that the dimensions which are the most valuable for gender categorisation are not constant across races (Yamaguchi et al., 1995), the lack of correlations involving experience could suggest that the same facial traits are of the greatest use for differentiating between genders regardless of the race of a face. However, adulthood experience was not quantified (only childhood experience was), therefore it remains untested whether adulthood experience relates to the other-race gender effect.

The final subsample size (53) may have been small for exploring correlations between experience and gender categorisation proficiency. Although there was no significant correlation between the other-race experience disadvantage and the other-race gender effect, it should be noted that the 99% confidence interval (generated from 9,999 bootstrapped bias-corrected resamples) had a lower limit of -.60 and an upper limit of -.02, i.e., zero was outside of this confidence interval thereby indicating that the correlation coefficient was different from zero. The Pearson's *r* itself (-.32) would suggest a medium effect size (Cohen, 1992) in the expected direction (the non-corrected *p*-value was .018).

If one uses *r*-values of .1, .3, and .5 for defining small, medium, and large effect sizes respectively (Cohen, 1992), some previous research which has applied correlation analyses between experience and the other-race effect has found significant correlations alongside small-to-medium effect sizes in the presence of larger samples sizes of, for instance with samples sizes of 172 (Wan et al., 2013)[36] and 146 (Zhao et al., 2014a). Still, the subsample size used in Experiment 3 exceeded the sizes used in several other explorations of the link between experience and the other-race effect (e.g., Davis et al., 2016; Hancock & Rhodes, 2008; Young & Hugenberg, 2012, in each correlation analysis).

---

36 Using Walker (2003), the tau-value of Wan et al. (2003) was converted to a Pearson's *r*-value.

Nonetheless, the size was perhaps small given the application of the Holm-Bonferroni correction (i.e., a Type II error may have occurred given the conjunction of subsample size and the correction). Furthermore, the Holm-Bonferroni adjustment (as with Bonferroni) would be conservative when tests are not independent (Abdi, 2010), and correlations in Experiment 3 were likely non-independent.

4.4.4 Gender/sexual dimorphism

Results in Experiment 3 suggested that the dimorphism of the face is subjectively larger in Caucasians than East Asian and Blacks regardless of observer race, and therefore this pattern may also occur in terms of objective facial dimorphism. The Gabor-jet model (Margalit, Biederman, Herald, Yue, & von der Malsburg, 2016) was used to quantify the psychophysical dissimilarity between males and females amongst stimuli using the faces from 20% and 80% along male-female morph continua; at least numerically, there was a bigger dissimilarity within Blacks (220 units) than the other two races, and Caucasians (161 units) had a larger difference than East Asians (130 units). Still, whether dimorphism in the face is similarity across races is unresolved.

Appendix B does present a supplementary analysis which found that the sexual size dimorphism of the cranium (adjusted for absolute latitude) increased as *Homo sapiens* migrated farther from their African start. However, faces in the Chicago Face Database were adjusted "such that the size of the core facial features as depicted in the photo was roughly equivalent across targets" (Ma et al., 2015, p. 1125). Size dimorphism (Samal et al., 2007; Steyn & İşcan, 1998) would have been minimised by this change in size. Nonetheless, dimorphism is apparent beyond size, for instance in shape (Mydlová et al., 2015).

The higher discriminability within Caucasian faces may apply to some local facial area. Indeed, the correlation between the other-race disadvantage for local processing (and not the other-race holistic) disadvantage for gender and the other-race gender effect suggests that some (or several) local areas were particularly important.[37]

4.4.5 Conclusion

The presence of the other-race gender effect would seem likely to be, in part, driven by gender discernment being generally easier for Caucasian faces than East Asian and Black faces. Regarding identity recognition, previous research has largely found no relationship between other-race expertise disadvantages and the other-race effect (DeGutis et al., 2013; Horry et al., 2015; Michel, Caldara, & Rossion, 2006; Michel, Rossion, et al., 2006; Zhao et al., 2014b); for gender, the current experiment demonstrated a correlation regarding the other-race local disadvantage for gender and the other-race gender effect. Furthermore, an association between the other-race gender categorisation bias and the other-race gender effect was apparent. These correlations, of course, do not mean that the local expertise disadvantage and categorisation bias lead to the other-race gender effect. Nonetheless, based on the results of this chapter, it is posited that the local disadvantage and bias are influential.

[37] Eyebrows affect gender categorisation (Brown & Perrett, 1993), and eyebrow-plucking by females has been suggested as a possible reason for better sex categorisation with Caucasian faces than East Asian faces (O'Toole et al., 1996). Yet, for the stimuli of the current study, the extent of eyebrow-plucking would appear to be reasonably consistent across females of each race.

Chapter 5: RACE CATEGORISATION AND GENDER

Following Sections 1.4.1, 1.4.2, and 1.5.4, Experiment 4 explored a number of themes regarding race category boundaries and race categorisation proficiency. It aimed to find if facial gender affects (monoracial) race categorisation, whether any effect of gender is of the same magnitude in a Caucasian-to-Black continua as it is in a Caucasian-to-East-Asian continua, and if gender categorisation proficiency influences the effects of gender on race categorisation. Regarding the impact of gender on the Caucasian/Black race category boundary, Experiment 5 considered a possible effect of facial orientation, whilst Experiment 6 concerned a potential contribution of facial luminance.

5.1 Experiment 4

5.1.1 Introduction

As in previous research (e.g., Krosch & Amodio, 2014), Experiment 4 utilised PSEs as monoracial race category boundaries. Unlike preceding research (Carpinella et al., 2015; Ge et al., 2009; Li et al., 2018), Experiment 4 did not use response latencies to measure race categorisation proficiency; regarding accuracy, results in some previous research would seem to be in the vicinity of ceiling (Ge et al., 2009; Li et al., 2018, Experiment 1A), however, ceiling (and floor) effects would likely not be an issue in the present experiment as decision noise (arising from responses to morph continua spanning 5% to 95% of a race) were utilised.

5.1.2 Method

5.1.2.1 Participants

There was a total of 50 participants (19 males, 31 females). The target sample size of 50 was calculated before data collection assuming a low participant non-engagement rate, and using G*Power Version 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009) under the suppositions of small-to-medium effect sizes and accounting for *p*-value adjustments to control the family-wise error rate. Given that the weighting of facial information in race categorisation changes between childhood and adulthood (Dunham, Stepanova, Dotsch, & Todorov, 2015), and that ageing affects gender categorisation ability (Carbon et al., 2013), the participation age limit was placed between 18 and 50 years; participants were between the ages of 18 and 50 years ($M_{\text{age}} = 25.24$, $SD_{\text{age}} = 8.09$).[38] Ethics approval was granted at City, University of London, which is where each experimental session took place.

5.1.2.2 Stimuli

Faces derived from the Chicago Face Database (Ma et al., 2015) were subject to morphing. Initial facial averages of Blacks, Caucasians, and East Asians for males and for females from Experiment 3 were morphed in Morpheus Photo Morpher Version 3.17. Within each gender, Caucasian-to-Black and Caucasian-to-East-Asian morph continua were created (i.e., four continua in total). The final seven stimuli selected for the experiment in each continuum were from between 5% to 95% along a continuum at 15% steps (i.e., 5, 20, 35, 50, 65, 80, and 95%). These stimuli were each enclosed in an oval to occlude the scalp hair and ears. Faces were presented in greyscale.

For the gender categorisation task, Caucasian male-to-female faces from

---

[38] Essentially, in case gender categorisation ability affected gender differences in category boundaries or proficiency, the age range was limited in order to maximise the chance of finding gender differences, as the main focus of Experiment 4 was on seeing if there were effects of gender. The possible influence of gender categorisation ability was treated as a secondary focus. However, limiting the age range may have limited the potential of the experiment to find impacts of gender categorisation ability.

Experiment 3 (i.e., individual faces from the Chicago Face Database, Ma et al., 2015, yet morphed) were employed (20 to 80% at 10% increments). Caucasian faces were used due to the stimuli having the lowest non-engagement level in Experiment 3 (see Section 6.3.2.1), Caucasian faces being employed in each racial morph continuum, and a commonality in gender categorisation ability for Caucasian faces with Black and East Asian faces as suggested by correlations using data from Experiment 3: engaged monoracial and multiracial participants were identified (using methods described in Chapter 4) and pooled together, and, after outlier identification (as in Experiment 3) and removal, Spearman's correlation analyses were applied in SPSS 25.0. Following Holm-Bonferroni corrections (Gaetano, 2013; Holm, 1979), there was a correlation between decision noise for faces of Caucasians and Blacks, $r_s(54) = .65$, $p < .001$, and Caucasians and East Asians, $r_s(57) = .55$, $p < .001$. These correlations would imply that an observer attempts to apply the same gender cues across races when categorising gender.

5.1.2.3 Procedure

Informed consent was given before the tasks commenced. There were three tasks, each of which was created in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997): a race categorisation task featuring the Caucasian-to-Black continua (where participants categorised each face as either Caucasian or Black), a race categorisation task comprised of the Caucasian-to-East-Asian continua (Caucasian or East Asian categorisation decisions), and a gender categorisation task formed of the male-to-female Caucasian faces (male or female categorisations). Faces were solely presented at an upright orientation. Half of the participants completed the Caucasian/Black task first of all, whilst the other half first finished the Caucasian/East-Asian task.

Each race categorisation task was comprised of six practice trials and 280 experimental trials. On each trial, a face was presented for 500 ms, after which participants categorised the race of the face. Regarding practice trials, presented stimuli were randomly selected from the stimulus set. Of the experimental trials, stimulus presentation order was randomised, and each stimulus was presented 20 times. Both of the race categorisation tasks were completed before the gender categorisation task. This final task followed the same procedure as in the categorisation task of Experiment 3, except that only Caucasian faces were presented (i.e., six experimental trials, and 140 experimental trials within which each stimulus was displayed 20 times). At the end of the session, participants were debriefed. The session lasted around 40 minutes.

5.1.2.4 Data preparation

As in the categorisation task of Experiment 3, per participant and continuum, a psychometric function was fit, both the PSE and decision noise were calculated, and non-engaged participants were identified using manners from Yarrow (2018) in MATLAB. As in Experiments 1 to 3, outliers were identified as values further than thrice the interquartile range from either upper or lower quartiles, and, regarding the multivariate linear regression analysis, studentized residuals over |3|. Analysis was undertaken in SPSS Versions 24.0 and 25.0. *P*-values were Holm-Bonferroni-corrected (Gaetano, 2013; Holm, 1979).

5.1.3 Results and discussion

One participant was removed due to non-engagement (which occurred across all tasks). Another was removed as raw data indicated that this participant responded *East Asian* with the assigned Caucasian response key, and *Caucasian* with the East Asian response key, e.g., responses on either end of the morph continua were largely in the opposite direction to what was expected.

5.1.3.1 Race category boundary

PSEs from race categorisation tasks were analysed within a 2 (racial ancestry [Caucasian/Black, Caucasian/East-Asian]) × 2 (facial gender [male, female]) within-subjects ANOVA. There was an effect of racial ancestry on PSEs, $F(1, 47) = 4.11$, $p = .048$, $\eta_p^2 = .08$, with the PSE being relatively nearer to the Caucasian norm in Caucasian-to-Black continua ($M = 52.93$, $SD = 9.63$) than in Caucasian-to-East-Asian continua ($M = 50.05$, $SD = 9.30$). Facial gender affected PSEs, $F(1, 47) = 20.43$, $p < .001$, $\eta_p^2 = .30$, as the PSE was relatively closer to the Caucasian prototype for male faces ($M = 53.73$, $SD = 7.70$) than female faces ($M = 49.25$, $SD = 9.75$). There was no interaction between racial ancestry and facial gender, $F(1, 47) = .20$, $p = .65$, $\eta_p^2 = .004$, which suggested that the effect of facial gender occurred amongst both racial ancestry continua. Accordingly, planned comparisons showed this to be the case amongst Caucasian-to-Black continua ($M_{male} = 55.37$, $SD_{male} = 9.97$; $M_{female} = 50.49$, $SD_{female} = 10.41$), $t(47) = 5.07$, $p < .001$, and Caucasian-to-East-Asian continua ($M_{male} = 52.09$, $SD_{male} = 8.87$; $M_{female} = 48.00$, $SD_{female} = 12.46$), $t(47) = 2.56$, $p = .014$ (Figure 5.1). Therefore, for Caucasian/Black and Caucasian/East-Asian racial ancestry continua, the race category boundary was a phenotype which had more evident Caucasian ancestry amongst male faces than female faces.

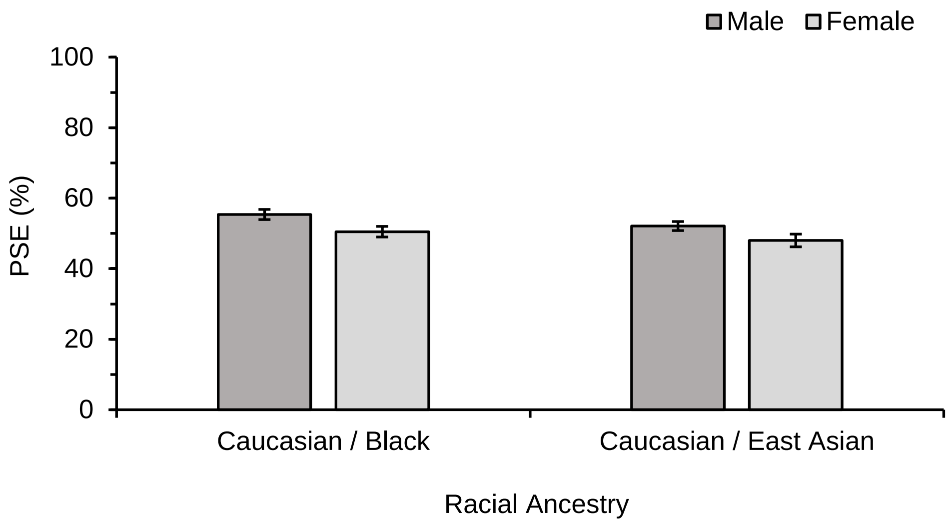


Figure 5.1. PSEs for race categorisations, i.e., race category boundaries. Error bars are for standard errors.

5.1.3.2 Race categorisation proficiency

A 2 × 2 within-subjects ANOVA was run on facial gender and racial ancestry pertaining to decision noise for race categorisation. Race categorisation was more proficient in the Caucasian-to-Black continua (M = 11.70, SD = 4.83) than the Caucasian-to-East-Asian continua (M = 17.75, SD = 10.55), $F(1, 47) = 25.36$, $p < .001$, $\eta_p^2 = .35$, and for male faces (M = 13.56, SD = 5.84) than female faces (M = 15.89, SD = 8.85), $F(1, 47) = 10.58$, $p = .002$, $\eta_p^2 = .18$. However, there was an interaction between racial ancestry and facial gender, $F(1, 47) = 18.40$, $p < .001$, $\eta_p^2 = .28$; t-tests demonstrated that amongst Caucasian-to-Black faces there was no difference in precision between male (M = 11.98, SD = 5.10) and female faces (M = 11.43, SD = 5.49), $t(47) = .87$, $p = .39$, although, for Caucasian-to-East-Asian faces, males (M = 15.15, SD = 8.58) were racially categorised more proficiently than females (M = 20.34, SD = 13.63), $t(47) = -4.21$, $p < .001$ (Figure 5.2). It is speculated that the lack of a gender difference for

Caucasian-to-Black faces suggests that the similarity between racial prototypes is the same within male faces as it is amongst female faces; for Caucasian-to-East-Asian faces, the greater proficiency for male faces may indicate that racial prototypes are less similar amongst males than females.

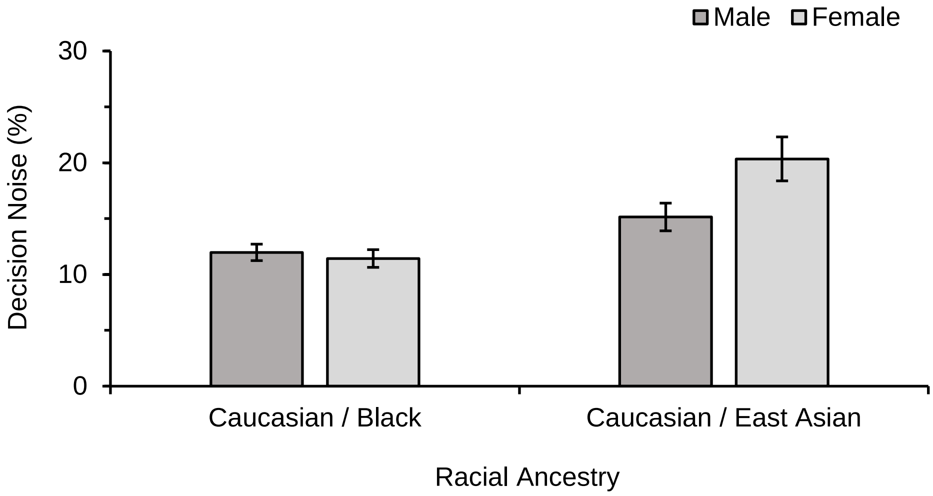


Figure 5.2. Decision noise for race categorisations, with standard errors represented in the error bars.

5.1.3.3 Gender categorisation ability

Multivariate linear regression was employed to find whether gender categorisation ability (decision noise) predicted gender effect sizes in race PSEs and decision noise; decision noise from the gender categorisation task was used as the predictor variable, whilst magnitudes of gender differences (i.e., four magnitudes per participant) were the dependent variables. Four participants were removed due to them being univariate outliers (one with respect to gender categorisation, two regarding decision noise differences for Caucasian/East-Asian categorisations [interquartile ranges], and one pertaining to the Caucasian/Black decision noise difference [studentized residual]). All in all, gender categorisation ability predicted magnitudes of gender effects, $F(4, 39) = 2.86$, $p = .036$, $V = .23$. More specifically, there was an effect on decision noise in Caucasian-to-Black continua (gender categorisation ability made race categorisation more proficient for males relative to females), $F(1, 42) = 10.92$, $p = .008$, , $\eta_p^2 = .21$, unlike in regard to the other magnitudes, *p*s > .05.[39]

5.2 Experiment 5

5.2.1 Introduction

Although Experiment 4 demonstrated an effect of gender on the race category boundary for faces of Caucasian and Black ancestry, and for faces of Caucasian and East Asian ancestry, faces were solely presented in an upright orientation. In real life, however, faces are seen in a variety of visual orientations. The effect of facial orientation on the racial categorisation of faces has rarely been studied (e.g., Cloutier & Macrae, 2007), and experiments had yet to examine whether race category boundaries are affected by orientation, and indeed

[39] Experiment 4 concentrated on the possible effects of facial gender rather than participant gender. Interestingly, participant gender affects the male/female gender category boundary of faces such that the boundary is in the direction of being nearer to the own-gender prototype (male, or female) on the gender morph continuum, although it is unknown why there is such a difference (Webster et al., 2004). Effects of participant sex on faces occur for accuracy (*d*') for sex categorisation, which is greater for female participants than males (O'Toole et al., 1996). This may reflect that face processing abilities are generally greater for female than male participants, which has, for instance, been noted as occurring in face recognition (O'Toole et al., 1996). As effects of participant gender were not a focus of Experiment 4, no attempt was made to balance sample size between males and females, hence unequal variances would be a reasonable expectation. An analysis using the engaged participants with respect to the gender categorisation task ($N = 49$) found that gender categorisation PSEs were the same for female participants ($M = 50.32$, $SD = 4.96$) as they were for male participants ($M = 45.64$, $SD = 11.05$), $t(21.04) = -1.70$, $p = .10$, unlike in Webster et al. (2004). Decision noise for gender categorisation was lower for females ($Mdn = 12.34$) than males ($Mdn = 16.93$), $U(47) = 133.00$, $p = .002$, which aligns with the greater sex categorisation proficiency of females in O'Toole et al. (1996).

whether facial orientation affects differences between males and females with respect to race category boundaries.

5.2.1.1 Facial orientation

5.2.1.1.1 Morphology

Morphology and skin tone influence how faces are perceived in terms of race (Dunham, Dotsch, Clark, & Stepanova, 2016; Dunham et al., 2015; Strom, Zebrowitz, Zhang, Bronstad, & Lee, 2012). Regarding identity, recognition decreases as orientation changes from upright and through degrees to inversion (Ashworth, Vuong, Rossion, & Tarr, 2008); compared to an upright orientation, face inversion disrupts the expert processing of morphology (e.g., Yovel & Kanwisher, 2004). The use of configural information in race categorisation has been evident (Zhao & Bentin, 2011), therefore it would not be surprising if inversion disrupted the processing of morphological race cues. Indeed, in Cloutier and Macrae (2007), orientation (from upright to inverted at 45° steps) reduced racial categorisation proficiency for monoracial faces of Caucasians and Blacks, with response latencies becoming lengthier, but this only occurred when skin tone was matched between the races using a colour manipulation; when skin tone was not manipulated, there was no influence of orientation on response latencies (Cloutier & Macrae, 2007). Furthermore, in Colombatto and McCarthy (2017), who used luminance-matched FaceGen stimuli, inversion lengthened response latencies and rendered accuracy poorer.

In Montalan et al. (2013), in which the mean luminance of pixels was equated across faces, orientation did not affect response latencies for the racial categorisation of Caucasian faces and Black faces. Montalan et al. (2013) were exploring the other-race categorisation advantage, which they found to not be apparent, perhaps due to only a few stimulus faces being used per race (eight Black, and eight Caucasian faces) allowing participants to have "performed on the basis of the idiosyncratic properties of individual stimuli" (Montalan et al., 2013, p. 365). The lack of a main effect of orientation on reaction times may have also been due to few faces being used, which allowed for racial responses for individual identities to be learned and applied.

5.2.1.1.2 Skin tone

Interestingly, facial luminance (lightness) perception has been proposed to be unaffected by face inversion (Willenbockel, Fiset, & Tanaka, 2011); face inversion increases the effect of facial luminance on race categorisation, which suggests that inversion increases the relative weight of *luminance* in race categorisation decisions, in place of disrupted processing of facial structure (Willenbockel et al., 2011). This strongly indicates that inversion increases the weighting of *skin tone* in race categorisation decisions, and would align with orientation leaving race categorisation response latencies unaffected in Cloutier and Macrae (2007) when skin tone differences between Blacks and Caucasians were not removed; in Cloutier and Macrae (2007), skin tone was given more weight as orientation changed from upright, and skin tone usage consequently made up for the weaker perception of structural racial information.

5.2.1.2 Gender

Regarding the skin of the inner arm, females are lighter than males (Jablonski & Chaplin, 2000), which would presumably apply to the face, and may factor into why gender affects race category boundaries.[40] It is not known if

[40] Nevertheless, the possible influence of skin tone may not be a reality as gender categorisation proficiency was not related to gender differences in race category boundaries in Experiment 4.

observers account for gender differences in skin tone (to even some extent) when racially categorising; for instance, regarding faces of Caucasian and Black ancestry, the darker skin tone of males than females may bias the Caucasian/Black race category boundary i) toward the Caucasian norm on the morph continuum for male faces and ii) nearer the Black norm for female faces. Observers may use gender categorisation to limit these biases.

Given that face inversion hampers male-female categorisation in terms of accuracy (Zhao & Hayward, 2010) and response latency (Cloutier & Macrae, 2007; Zhao & Hayward, 2010), inversion may reduce any adjustment for gender differences in skin tone when racially categorising. In tandem with the possible greater weight given to skin tone under inversion, inversion may shift race category boundaries in opposite directions for males and females. Hence, regarding Caucasian and Black ancestry under inversion, the Caucasian/Black race category boundary on the morph continuum may shift in the direction of the Caucasian norm for males, and the Black norm for females. Therefore, inversion could magnify the gender difference in race category boundaries.

Although previous research had not directly tested whether facial orientation affects race category boundaries, prior research (Willenbockel et al., 2011) may suggest that race category boundaries amongst male faces would be unaffected by inversion; in Willenbockel et al. (2011), the facial luminance and orientation of male Caucasian-to-Black faces was manipulated, and a main effect of orientation on White categorisations did not occur. Still, Figure 2b of Willenbockel et al. (2011) could suggest that inversion may have caused male faces to generally be perceived as more Caucasian (compared to an upright facial orientation) when faces were morphed to be 50% between Caucasian faces and

Black faces; for male faces, inversion may have moved the race category boundary nearer the Black norm on the race morph continuum.

Experiment 5 focussed on determining if facial orientation does indeed affect gender differences in a race category boundary (using upright and inverted faces), and if it does so by affecting the race category boundary for male faces, female faces, or both. Effects of inversion on race categorisation proficiency had been studied in terms of response latencies to monoracial faces (Cloutier & Macrae, 2007; Colombatto & McCarthy, 2017; Montalan et al., 2013), and percentage accuracy with respect to monoracial FaceGen faces (Colombatto & McCarthy, 2017); Experiment 5 explored whether decision noise (for *real* Caucasian-to-Black faces varying in skin tone and morphology) would be affected by orientation.

5.2.2 Method

5.2.2.1 Participants

A sample size of 16 engaged participants was decided on bearing in mind sample sizes of prior psychophysics experiments (e.g., Murphy & Cook, 2017; Shah et al., 2015), and using G*Power Version 3.1.9.2 (Faul et al., 2009) to calculate the necessary sample size to replicate the effect of gender on the race category boundary using data from the 49 engaged participants of Experiment 4 with respect to the Caucasian-to-Black continua (with power at .8 and an alpha-level of .05, the G*Power calculation suggested that 16 participants were required). There were 17 participants (eight males, nine females), none of whom participated in Experiment 4 (one participant was not engaged, see Sections 5.2.2.4 & 5.2.3). The minimum age for participation was 18-years-old (M_{age} = 31.76, SD_{age} = 11.30) given that race categorisation decisions are weighted more for skin tone (and less for structure) in childhood compared to adulthood (Dunham et al., 2015), and that the magnitude of the face inversion effect (for identity) increases during childhood (Hills & Lewis, 2018) then remains stable across adulthood (Boutet & Faubert, 2006). All sessions of the experiment took place at City, University of London, which was the institution where ethics was approved.

5.2.2.2 Stimuli

The Caucasian-to-Black continua used in Experiment 4 (ultimately derived from faces in the Chicago Face Database, Ma et al., 2015) were also employed in Experiment 5, but presented at an inverted orientation (180°) as well as upright (0°). To maximise participant engagement, it was decided that one racial type of morph continuum be used; from Experiment 4, Caucasian-to-Black continua were employed rather than Caucasian-to-East-Asian continua.

Of the two types of racial ancestry continua, Caucasian-to-Black was used in order to maximise the chance of finding an effect of orientation on a race category boundary in the experiment. Median luminance measurements (derived in Adobe Photoshop Version 2015.0.0) of stimuli for each stimulus at 5% and 95% along each racial morph continuum at 5% away from Black or East Asian norms in Caucasian-to-Black and Caucasian-to-East-Asian continua respectively (Table 5.1) showed that the difference between genders was numerically greater 5% from the Black norm (12 units) than 5% from the East Asian norm (5 units). This may indicate a greater gender difference in skin tone in the Caucasian-to-Black continuum than the Caucasian-to-East Asian continuum. This outcome could hint that any effect of facial orientation on race category boundaries would perhaps be more forthcoming with the morphed faces of Caucasian and Black ancestry than those of Caucasian and East Asian ancestry.

Table 5.1

Median Luminance of the Endpoints of Morph Continua used in the Race Categorisation Tasks of Experiment 4

		Gender	
Racial continuum	Percent morph	Male	Female
Caucasian-to-Black	5	158	162
Caucasian-to-Black	95	102	114
Caucasian-to-East-Asian	5	161	165
Caucasian-to-East-Asian	95	158	163

Note. A higher morph percentage represents greater non-Caucasian ancestry.

5.2.2.3 Procedure

Through the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), a race categorisation task (Caucasian/Black race categorisation decisions) was modified from Experiment 4 and deployed in MATLAB. After informed consent was attained, the task began. Six practice trials were followed by 560 experimental trials. The practice trials were selected randomly from the experimental trials, and the trial order of the experimental trials was randomised. Per trial, a face was displayed for 500 ms, and then participants monoracially classified the face as Caucasian or Black. In the 560 experimental trials, each stimulus face was intended to be shown upright 20 times and inverted 20 times. However, a programming error meant that inverted male faces of between 15% and 95% Caucasian ancestry were shown at least 19 times; a trial intended for each of those levels (six trials) was allocated at random to any condition. Once the task had ended, participants were debriefed. Experiment 5 had a duration in the region of 30 minutes.

5.2.2.4 Data preparation

For simplicity, the aforementioned erroneous six trials were removed from analysis. The same methods of defining outliers and non-engaged participants were employed as in Experiment 4, i.e., Yarrow (2018) for non-engagement in MATLAB, and the distance from upper/lower quartiles (exceeding three times the interquartile range) for defining outliers. Analysis was performed in SPSS 24.0, with Holm-Bonferroni corrections being used (Gaetano, 2013; Holm, 1979).

5.2.3 Results and discussion

In Experiment 5, the data of one participant was removed as they were non-engaged when categorising female upside-down faces.

5.2.3.1 Race category boundary

PSEs were analysed with a 2 (gender of face [male, female]) × 2 (facial orientation [upright, inverted]) within-subjects ANOVA. There was a main effect of gender, with PSEs being higher for male faces (M = 54.94, SD = 6.74) than female faces (M = 48.58, SD = 10.78), $F(1, 15) = 20.68$, $p < .001$, $\eta_p^2 = .58$. A main effect of facial orientation was apparent, $F(1, 15) = 5.35$, $p = .035$, $\eta_p^2 = .26$, with PSEs being greater for upright faces (M = 53.96, SD = 7.79) than inverted faces (M = 49.55, SD = 10.69). Nonetheless, these main effects need to be interpreted in the context of an interaction between gender and facial orientation, $F(1, 15) = 10.25$, $p = .006$, $\eta_p^2 = .41$; PSEs were higher for male faces than female faces amongst upright faces (M_{male} = 55.67, SD_{male} = 6.65; M_{female} = 52.26, SD_{female} = 9.55), $t(15) = 2.58$, $p = .021$ (as in Experiment 4), and inverted faces (M_{male} = 54.21, SD_{male} = 8.12; M_{female} = 44.90, SD_{female} = 13.92),

$t(15) = 4.74$, $p < .001$, and the interaction would suggest that this difference was greater amongst inverted faces. Furthermore, whilst inversion did not affect PSEs for male faces, $t(15) = .94$, $p = .36$, inversion lowered PSEs for female faces, $t(15) = 2.87$, $p = .023$. Therefore, it would seem that the increase in the magnitude of the gender effect under inversion was due to a change in the race category boundary amongst female faces, and not within male faces.

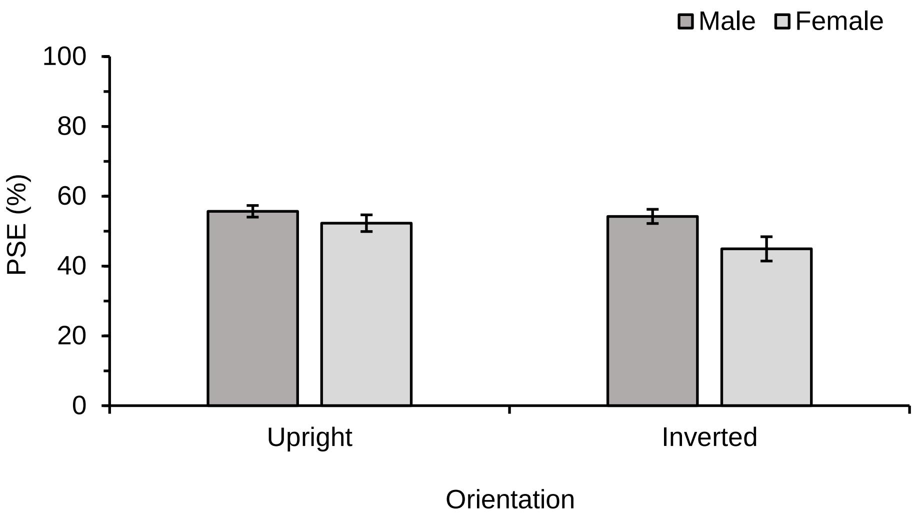


Figure 5.3. Race categorisation PSEs for Caucasian-to-Black continua of male and female faces under upright and inverted facial orientation. Standard errors are portrayed in the error bars.

5.2.3.2 Race categorisation proficiency

One outlier was identified; its removal did not affect results, therefore, given the sample size, it was retained in the analyses. Decision noise was entered into a within-subjects ANOVA which used the same independent variables and levels as the analysis of PSEs. There was a main effect of gender: racial categorisations were more proficient for male faces ($M = 10.08$, $SD = 6.16$) than females faces ($M = 16.28$, $SD = 11.11$), $F(1, 15) = 13.97$, $p = .002$, $\eta_p^2 = .48$.

Categorisation was affected by facial orientation, with proficiency being greater for upright faces (*M* = 10.03, *SD* = 7.30) than inverted faces (*M* = 16.33, *SD* = 9.83), $F(1, 15) = 30.50$, $p < .001$, $\eta_p^2 = .67$. There was no interaction between gender and facial orientation, $F(1, 15) = 1.74$, $p = .21$, $\eta_p^2 = .10$, which suggests that the effect of facial orientation occurred across male faces and female faces to corresponding extents (Figure 5.4).

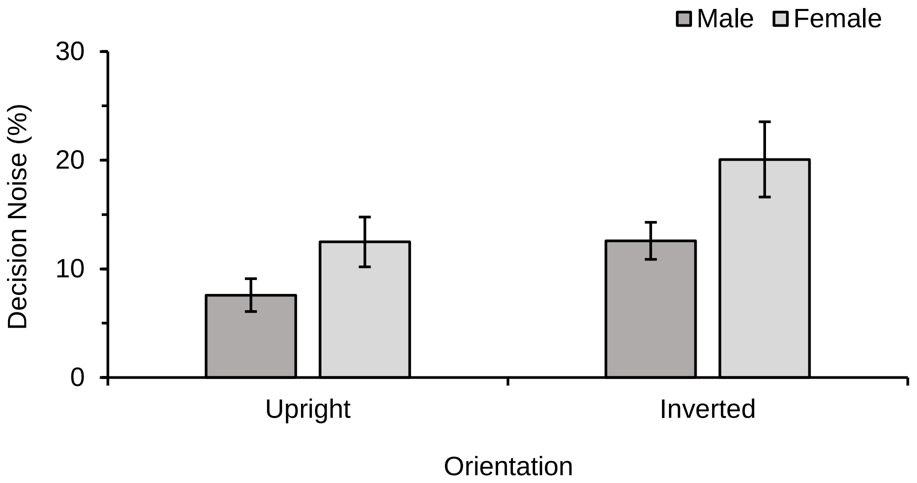


Figure 5.4. Proficiency (decision noise) for race categorisation decisions under the different conditions. Error bars represent standard errors.

Therefore, unlike reaction times for monoracial Caucasian and Black faces (when skin tone is not manipulated) (Cloutier & Macrae, 2007), inversion is detrimental to the race categorisation proficiency (decision noise) of Caucasian-Black racial ancestry continua. In comparison to Experiment 4, amongst upright faces, male faces (*M* = 7.58, *SD* = 6.05) were categorised more proficiently than female faces (*M* = 12.48, *SD* = 9.17), $t(15) = -3.68$, $p = .002$; there is no clear reason for why there was a gender difference here and not in Experiment 4.

5.3 Experiment 6

5.3.1 Introduction

Experiments 4 and 5 showed that facial gender affects the Caucasian/Black race category boundary. However, those experiments did not test from where this effect arose. For instance, it could have arisen due to a gender difference in skin tone, and, consequently, males and females differing with respect to facial luminance. On the arm, females do have a lighter skin tone than males (Jablonski & Chaplin, 2000); median facial luminance of self-categorised Blacks and Whites from the Chicago Face Database (Ma et al., 2015) is greater for females than males amongst Blacks ($M_{\text{female}} = 111.75$, $SD_{\text{female}} = 16.49$; $M_{\text{male}} = 101.46$, $SD_{\text{male}} = 20.56$), $t(176.20) = 3.84$, $p < .001$, $d = .55$, and Whites ($M_{\text{female}} = 167.17$, $SD_{\text{female}} = 10.99$; $M_{\text{male}} = 159.62$, $SD_{\text{male}} = 11.67$), $t(181) = 4.50$, $p < .001$, $d = .66$ (Holm-Bonferroni-corrected, Gaetano, 2013; Holm, 1979), with *d*s calculated via Lakens (2013). This difference in luminance would presumably reflect a difference in facial skin tone.

Experiment 6 tested whether a gender difference in facial luminance contributes towards the gender difference regarding the Caucasian/Black boundary. The present experiment replicated the Experiment 4 Caucasian/Black race task (i.e., with typical luminance), but additionally with trials which equalised the facial luminance of males and females within each individual morph level. If luminance-equalisation reduced the effect of gender on the Caucasian/Black race category boundary, this would suggest that luminance is a contributor to the gender effect.

5.3.2 Method

5.3.2.1 Participants

The goal was to have 16 engaged participants, who were 18-years-old upwards (see Section 5.2.2.1). There were 16 participants (2 males, 14 males; $M_{\text{age}} = 25.13$, $SD_{\text{age}} = 3.10$). Participation occurred at City, University of London, where ethics approval was given.

5.3.2.2 Stimuli

The Caucasian-to-Black stimuli of Experiment 4 were utilised. Additionally, using the SHINE toolbox (Willenbockel et al., 2010) in MATLAB, the luminance within the male continuum and the female continuum was equalised (in terms of the mean and standard deviation, aside from the colour black which was treated as a background) inside each morph level (not across), e.g., at 20% along the Caucasian-to-Black continuum, the luminance of males and females was equalised.

5.3.2.3 Procedure

After informed consent was attained, the Caucasian/Black race categorisation task commenced. The procedure of this task matched that of the Caucasian/Black task of Experiment 4 but also with luminance-equalised trials (i.e., Experiment 5 but with luminance-equalised upright faces instead of luminance-typical inverted faces); there were six practice trials, and, subsequently, 560 experimental trials. Of the experimental trials, each face was presented 20 times, except for the 15% to 95% Caucasian-ancestral male luminance-equalised faces regarding four participants, in which faces were presented 19 or more times due to the same programming error which afflicted Experiment 5 (Section 5.2.2.3). Once the task was completed, participants were debriefed. A session of the experiment lasted about 30 minutes.

5.3.2.4 Data preparation

The data was readied in the same manner as described in Experiment 4. Analysis was undertaken in SPSS 24.0.

5.3.3 Results and discussion

5.3.3.1 Race category boundary

On the PSEs, a within-subjects ANOVA, 2 (gender [male, female]) × 2 (luminance [typical, equalised]), was run. A main effect of gender occurred, with PSEs for male faces (M = 55.96, SD = 4.80) exceeding those of female faces (M = 51.56, SD = 7.70), $F(1, 15) = 13.94$, $p = .002$, $\eta_p^2 = .48$. Regarding luminance, there was not a main effect, as PSEs for luminance-typical faces (M = 53.65, SD = 6.17) were the same as luminance-equalised faces (M = 53.87, SD = 6.02), $F(1, 15) = .12$, $p = .74$, $\eta_p^2 = .01$. A gender × luminance interaction was not apparent, $F(1, 15) = 2.28$, $p = .15$, $\eta_p^2 = .13$, which demonstrates that facial luminance (and possibly skin tone) was not a driver behind gender affecting the race category boundary, and suggests that facial morphology was the crucial factor.

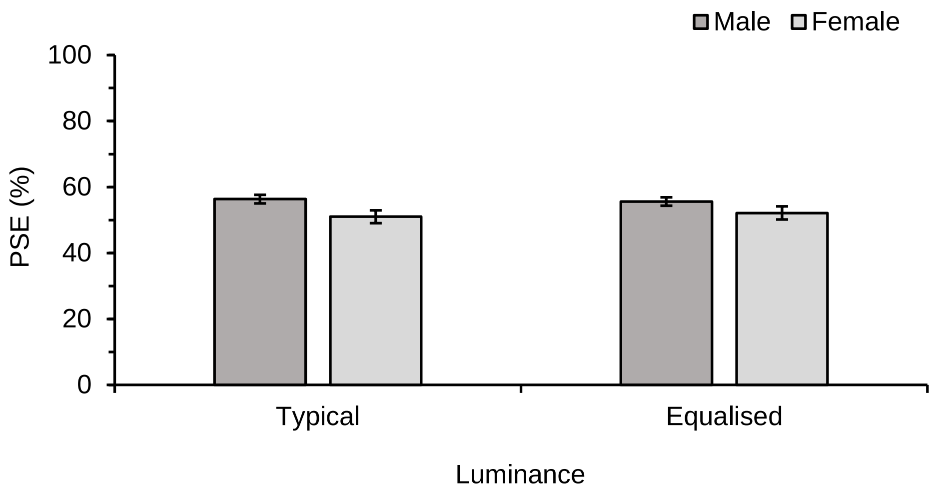


Figure 5.5. PSEs (Caucasian/Black) for male and female faces of typical and equalised luminance.

5.3.3.2 Race categorisation proficiency

Given that gender did not affect decision noise in Experiment 4, decision noise was not of central interest in Experiment 6. Nonetheless, regarding decision noise, a 2 (gender [male, female]) × 2 (luminance [typical, equalised]), within-subjects ANOVA, demonstrated a main effect of gender, as proficiency was greater regarding male (M = 8.72, SD = 5.72) than female (M = 12.48, SD = 8.49) faces, $F(1, 15) = 7.36$, $p = .02$, $\eta_p^2 = .33$. A main effect of luminance was not apparent, with proficiency for luminance-typical faces (M = 10.22, SD = 6.16) matching that of luminance-equalised faces (M = 10.98, SD = 7.64), $F(1, 15) = .66$, $p = .43$, $\eta_p^2 = .04$. An interaction between gender and luminance did not occur, $F(1, 15) = .002$, $p = .96$, $\eta_p^2 = .0002$ (Figure 5.6). As the removal of an outlier left the preceding analysis unaffected, results include that outlier. Regarding luminance-typical faces, with that outlier kept, decision noise was lower for males (M = 8.33, SD = 4.33) than females (M = 12.11, SD = 9.03) (which is contrary to Experiment 4, but aligns with results from Experiment 5), $t(15) = -2.17$, $p = .046$, although, with removal of that outlier, males ($M = 8.31$, $SD = 4.48$) and females ($M = 10.75$, $SD = 7.44$) did not differ in decision noise, $t(14) = -2.06$, $p = .058$ (agreeing with Experiment 4, unlike Experiment 5), yet it should be noted that removal reduces power and that the *p*-value was still marginal. The effect of removing the outlier may indicate that the experiment itself was underpowered. However, this should not imply that there actually is an effect of gender. Therefore, regarding luminance-typical faces in Experiment 6, it is not clear if the gender of faces affects the proficiency with which faces of Caucasian and Black ancestry are categorised by race.

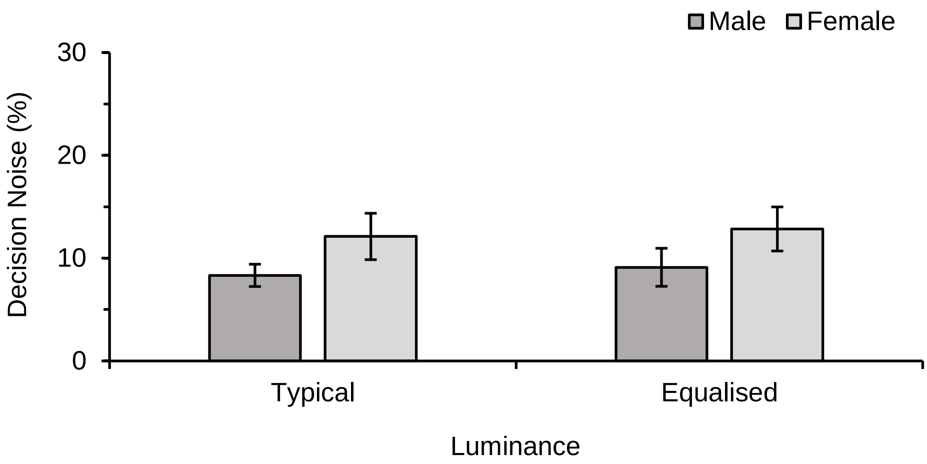


Figure 5.6. Decision noise for Caucasian/Black race categorisations for faces of males and females at typical and equalised levels of luminance.

5.4 General discussion

Chapter 5 examined whether the gender and visual orientation of faces affect race categorisation. Experiment 4 enquired into whether facial gender affects the race categorisation of upright faces, whilst Experiment 5 determined to find if facial orientation affects a gender difference in a race category boundary, and if orientation influences categorisation proficiency, and Experiment 6 determined whether a gender difference in facial luminance is a factor regarding gender affecting a race category boundary.

5.4.1 Upright faces

5.4.1.1 Boundary

As covered in Section 1.5.1, previous research had not directly explored whether gender affects monoracial race category boundaries of the real faces of others, although effects of gender on race category boundaries are implied by gender affecting i) the conceptual categorisation of non-faces (Ho et al., 2011), and ii) the categorisation of objectively ambiguous biracial FaceGen faces (midway between Caucasian and either Black or East Asian) (Carpinella et al., 2015).[41] Nonetheless, there had been some inconsistency concerning whether gender affects how non-face stimuli are racially categorised (Ho et al., 2011), and although facial femininity/masculinity influences the categorisation of biracial FaceGen faces (Carpinella et al., 2015), there are divergences between FaceGen and real faces in terms of racial differences (Holland, 2009) and race categorisation (Gaither et al., 2019); the present research demonstrated an effect of gender on Caucasian/East-Asian (Experiment 4), and Caucasian/Black race category boundaries (Experiments 4, 5, and 6).

The gender difference concerning PSEs could be explained through

[41] For the current experiments, norms and racial morph continua were derived from real faces (of high racial prototypicality based on ratings from the Chicago Face Database), therefore stimuli could be assumed to be representative of reality. However, the phenotype of biracials (of equal ancestry from two races) on a facial quality would not necessarily be equidistant between two races on average; the skin tone of biracials who have an African parent and a European Caucasian parent has been reasoned to be, on average, located between the midway point and the European Caucasian parent (Khan, 2008).

previous research on response latencies. Such research (Carpinella et al., 2015; Li & Tse, 2016; Thomas et al., 2014) would suggest that Caucasian females are closer than Caucasian males to the overall Caucasian norm, whilst Black males are nearer than Black females to the overall Black norm (see Section 1.5.2). Therefore, the objective midpoint between (general) race norms may be expected to transfer to a point relatively nearer to the Caucasian gender norm on the Caucasian-to-Black morph continuum amongst male faces than female faces (Figure 5.7). This would introduce a *bias* into PSEs, encouraging PSEs to be closer to the Caucasian gender norm for male faces than for female faces. Reaction times concerning East Asian faces (Carpinella et al., 2015; Li & Tse, 2016) indicate that East Asian male and female norms are equidistant to the overall East Asian race norm (Figure 5.7); gender differences amongst Caucasian faces may cause a bias in the Caucasian/East-Asian race category boundary. As a gender difference would be occurring in one monoracial category rather than two, one could expect the effect of gender on race category boundaries to be greater for Caucasian-to-Black faces than for Caucasian-to-East-Asian faces at least *numerically* if not significantly; effect sizes in Experiment 4 (calculated using a spreadsheet from Lakens, 2013) would indicate this to be the case, $d_{\text{Caucasian-to-Black}} = .73$, $d_{\text{Caucasian-to-East-Asian}} = .37$. Nonetheless, whilst outcomes in Davenport (2016) could lead to the expectation that the type of biracial ancestry would change the extent to which gender affects race category boundaries (Section 1.5.1), in Experiment 4 the magnitude of gender effects did not differ between Caucasian-to-Black and Caucasian-to-East-Asian continua.

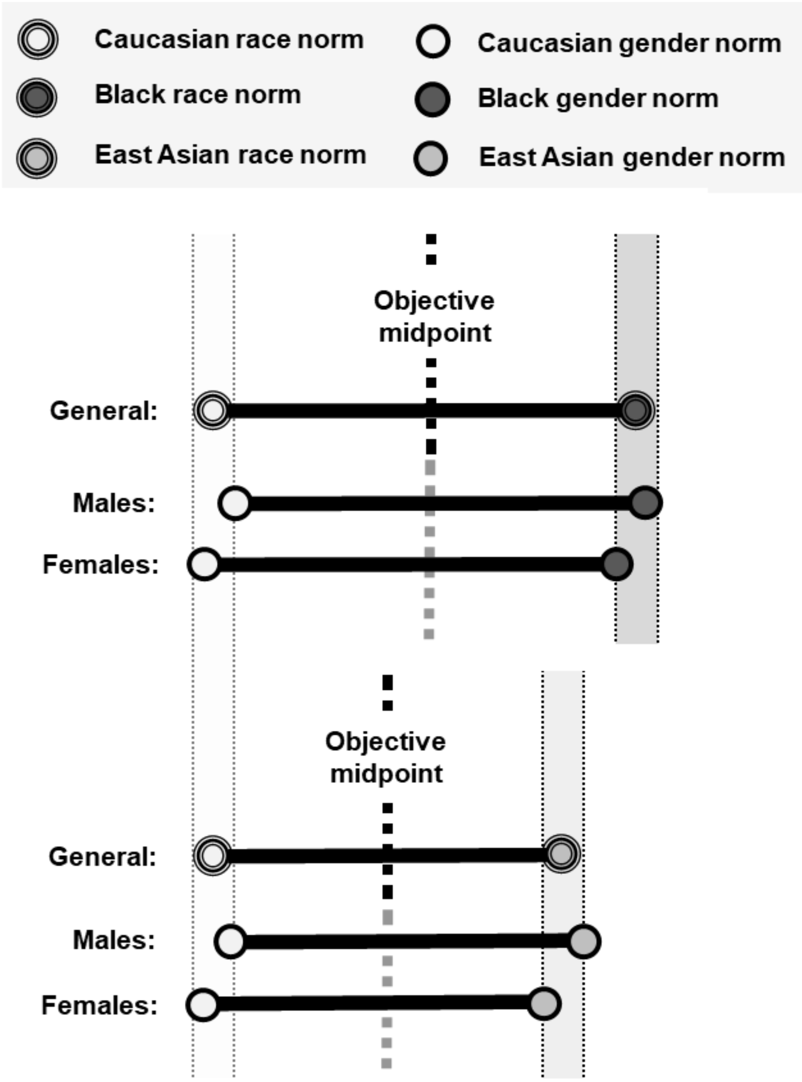


Figure 5.7. Caucasian/Black and Caucasian/East-Asian race category boundaries, and biases. Results with decision noise suggest i) greater physical dissimilarity between Caucasian and Black race norms than between Caucasian and East Asian race norms (Experiment 4), ii) that dissimilarities between race/gender norms are larger for males than females concerning Caucasian and East Asian ancestry (Experiment 4), but it is unclear if this gender difference also applies regarding Caucasian and Black ancestry (Experiments 4, 5, and 6).

5.4.1.2 Proficiency

As explained in Section 1.5.2, research had seldom explored the effect of gender on the proficiency with which race is categorised beyond response latencies. Regarding decision noise in Experiment 4, race categorisations were generally more difficult for female faces than male faces. In the context of the gender × ancestry interaction and gender comparisons within racial ancestries, this would seem to be driven by a relative difficulty categorising faces on the Caucasian-to-East-Asian female continuum (compared to Caucasian-to-East-Asian male faces). Interestingly, regarding accuracy, an interaction between gender and ancestry had been apparent in previous research (Li & Tse, 2016), yet it was not fully explored therein; in Experiment 4, there was no difference between genders in Caucasian-to-Black continua, yet there was a gender difference concerning Caucasian-to-East-Asian continua. As stated in Section 5.1.3.2, this could suggest that dissimilarities between Caucasians and East Asians are objectively greater amongst males than females. Finding a gender difference with Caucasian-to-East Asian faces alone cannot be attributed to a ceiling effect concerning Caucasian-to-Black faces, as performance with such faces was not approaching ceiling (i.e., not near 0%); indeed, previous research using decision noise (Murphy & Cook, 2017) has found experimental manipulation effects (aperture-viewing) concerning decision noises of around 10-15% (i.e., similar percentages to the decision noise found with the Caucasian-to-Black faces) for the *worst*-performing condition in a comparison. In contrast to Experiment 4, Experiment 5 found that proficiency was higher for males than females for Caucasian-to-Black faces, and the outcome in Experiment 6 (luminance-typical faces) is equivocal. Whilst variety in *p*-values can be expected across replications of experiments (e.g., Cumming, 2008), it is nevertheless unresolved whether there is a gender difference in race categorisation proficiency for faces of Caucasian and Black ancestry.

It may be useful to consider if a difference in the procedure between experiments could have led to the uncertainty regarding whether gender influences race categorisation proficiency for Caucasian-to-Black faces. Due to the inconclusive outcome within Experiment 6, only Experiments 4 and 5 will be considered initially. In their Caucasian/Black race categorisation tasks, whilst both experiments had the same number of experimental trials featuring upright faces, the total number of experimental trials (regardless of facial orientation) in Experiment 4 (280 trials) was half that of Experiment 5 (560 trials). The mean decision noise seems similar for females in both Experiment 4 (M = 11.43; SD = 5.49) and Experiment 5 (M = 12.48; SD = 9.17), but perhaps higher for males in Experiment 4 (M = 11.98; SD = 5.10) than Experiment 5 (M = 7.58; SD = 6.05). This raises the possibility that the discrepancy between Experiments 4 and 5 could be due to a greater familiarity with the stimuli in Experiment 5 boosting performance with male faces. Like Experiment 5, Experiment 6 also featured a total of 560 experimental trials, and (in luminance-typical trials) decision noise for males (outlier retained: M = 8.33; SD = 4.33; outlier retained: M = 8.31; SD = 4.48) may be more similar to Experiment 5 than Experiment 4.

Nonetheless, whether the discrepancy between Experiments 4 and 5 is due to the number of trials is speculative. Indeed, it was Experiment 3 which had the greater number of participants, and indeed more so than both Experiments 4 and 5 combined, so it could be argued that more weight could be placed on Experiment 3. Additional research would be required to bring some resolution to whether gender affects Caucasian/Black race categorisation proficiency, and part of that research could include finding whether the number of trials improves the proficiency with which male Caucasian-to-Black faces are racially categorised.

An ancestral difference in decision noise (main effect of ancestry) was not of primary interest in Experiment 4. Still, the lower decision noise for Caucasian-to-Black continua than Caucasian-to-East-Asian continua could be explained by there being less physical dissimilarity between Caucasians and East Asians than between Caucasians and Blacks. Indeed, regarding race, observers are sensitive to the morphology and skin tone of the face (e.g., Dunham, et al., 2015; Stepanova & Strube, 2012), and, if a dendrogram of morphological distances between populations of skull measurements (Relethford, 2009) applies to the face, Caucasians may have a greater facial morphological similarity with East Asians than Blacks, plus, regarding the lightness/darkness of facial skin, Figures 2a and 2b of de Rigal (2010) would indicate that Caucasians are more similar to East Asians than they are to Blacks.

5.4.1.3 Gender categorisation

It is not forthcoming why gender categorisation decision noise predicted gender differences concerning Caucasian/Black decision noise, yet not the other three gender differences. Results could mean that the facial traits which are most useful in Caucasian/Black categorisations are different for Caucasian-to-Black male faces than female faces, whilst they are the same for Caucasian-to-East-Asian males and females in Caucasian/East-Asian categorisations.

However, racial differences in the value of facial dimensions for gender differentiation have been suggested (Yamaguchi et al., 1995); if observers switch how facial gender cues are weighted according to the race which is being viewed, solely using Caucasian faces in the gender categorisation task may have weakened relationships between gender categorisation ability and magnitudes of gender differences in PSEs and decision noise; Caucasian gender cues would lose representativeness as continua became closer to Black or East Asian norms. Still, the interracial correlations in gender categorisation ability (Section 5.2.2) would suggest that the gender categorisation proficiency found with Caucasian faces would be representative of abilities with Black and East Asian faces. Still, it may have been more prudent to have included Black and East Asian gender morph continua in addition to the Caucasian continuum, and used averages (e.g., an average of gender categorisation proficiencies for Caucasian faces and East Asian faces with respect to the gender difference in Caucasian-to-East-Asian PSEs).

5.4.2 Orientation

5.4.2.1 Boundary

Section 5.2.1 considered whether perceivers make adjustments for skin tone when racially categorising, given that females have a lighter skin tone than males. Results of Experiment 5 may indicate that perceivers do attempt to adjust for gender differences in skin tone; as inversion disrupts gender/sex categorisation (Cloutier & Macrae, 2007; Zhao & Hayward, 2010), adjustment occurs more for upright faces than inverted faces. The effect of orientation on the Caucasian/Black race category boundary for females (under inversion, moving nearer the Black norm on the morph continuum) and not males tentatively indicates that adjustments are made specifically for the lighter skin tone of females rather than the darker skin tone of males; the skin tone of males may be perceived as the default for race categories.

5.4.2.2 Proficiency

Regarding race categorisation proficiency, whilst Cloutier and Macrae (2007) found that facial orientation left response latency unaffected when skin colour was kept within stimulus faces, Experiment 5 demonstrated a clear worsening in terms of decision noise. That outcome of Experiment 5 suggests that despite inversion increasing a reliance on skin tone in race categorisation (Willenbockel et al., 2011), the impaired processing of facial structure under inversion (Cloutier & Macrae, 2007) negatively impacted on proficiency (decision noise). This supports the importance of facial morphology in race categorisation. However, the stimuli of Experiment 5 were in greyscale rather than in colour; compared to presenting faces with their actual skin colour, a greyscale presentation has been considered to attenuate the perception of dissimilarities in skin colour (Anzures, Pascalis, Quinn, Slater, & Lee, 2011). Therefore, the negative effect of inversion on proficiency in Experiment 5 might not have occurred (or being as large) had faces been presented in colour.

It could be that the non-effect of orientation in Cloutier and Macrae (2007) (skin colour retained) was due to task ease. Cloutier and Macrae (2007) presented Caucasian faces and Black faces, and response latencies were lengthier when racial skin tone differences were removed; in that condition, orienting away from upright increased response latencies (Cloutier & Macrae, 2007), i.e., an effect of orientation on response latencies may only be apparent when race categorisations are generally more challenging. In Experiment 4, Caucasian/East-Asian categorisations were more difficult than Caucasian/Black categorisations; perhaps inversion would increase response latencies for Caucasian/East-Asian categorisations to monoracial Caucasian faces and East Asian faces.

5.4.3 Luminance

Experiment 6 found that the lighter luminance of females (in comparison to males) was not a cause of males and females differing in the Caucasian/Black boundary. This could seem surprising given that skin tone is a determiner of racial perception, such as for Caucasian-to-Black (Dunham et al. 2015;

Stepanova & Strube, 2012) and Caucasian-to-East-Asian faces (Dunham et al., 2016), however, even when controlling for skin tone, effects of masculinity/femininity on the racial categorisation of racially ambiguous (FaceGen) faces remain (Carpinella et al., 2015). This suggests that the difference between genders in the race category boundary (and proficiency) was due to morphology. [42] Therefore, whilst observers may successfully adjust for the lighter luminance and skin tone of females when categorising upright faces, and this adjustment is hampered under inversion, luminance (and likely skin tone) has no bearing on gender difference regarding upright faces, although (given Experiment 5) an influence could yet occur when faces are inverted.

5.4.4 Conclusion

The present experiments demonstrated that the Caucasian/East-Asian and Caucasian/Black race category boundaries are affected by the gender of a face, the effect of gender (Caucasian/Black boundary) regarding upright faces is not attributable to males and females differing in facial luminance, and that this gender effect is enhanced by inversion. Given that gradually increasing the angle of rotation away from upright to inverted incrementally reduces the expert

[42] Differences would seem unlikely to be due to gender differences in the racial ancestry of self-categorised monoracial persons. For instance, the higher PSE for male faces than female faces amongst Caucasian-to-Black faces is likely not attributable to self-categorised Black women having less African ancestry than self-categorised Black men; regarding the African ancestry of African Americans, data from two studies showed no difference between males and females, whilst examination of data from another study showed females to have a higher percentage than males (Cheng et al., 2012). Furthermore, the result concerning the Caucasian/Black PSE is doubtful to have arisen from Caucasian men having less Caucasian ancestry than Caucasian women; the extent of European Caucasian ancestry does not differ between European American males and females (Halder et al., 2012). A literature search did not find genetic ancestry percentages for East Asian Americans stated separately for males and females.

processing of facial structure (e.g., Cloutier & Macrae, 2007), the effect of orientation on the gender difference can be supposed to gradually increase as orientation departs further from upright. The effect of inversion on the gender difference in the Caucasian/Black boundary was driven by the race category boundary for female faces becoming a less prototypically Caucasian phenotype, thereby indicating that the skin tone of monoracial males could be perceived as more racially prototypical than the skin tone of monoracial females. Whilst race categorisation proficiency was greater for male faces than female faces for Caucasian-to-East-Asian faces (Experiment 4), there was conflicting evidence concerning whether such a difference is present amongst Caucasian-to-Black faces (Experiment 4 vs. 5, and perhaps 6). Race categorisation became less proficient when faces were inverted, thereby aligning with previous research (e.g., Dunham et al. 2015) which indicated a substantial weighting of facial structure in race categorisation when facial orientation is upright.

Chapter 6: GENERAL DISCUSSION

This thesis presented research on the face and race across two areas: i) identity recognition, and ii) gender and race categorisation. For identity recognition, the thesis explored whether experience during childhood relates to the other-race effect and other-race recognition in adulthood (Chapter 2: Experiments 1 and 2), racial/ethnic inequality in objective facial variability (Chapter 3: Section 3.1), and if other-race subjective homogeneity relates to the other-race effect (Chapter 3: Section 3.2). In the area of gender and race categorisation, the other-race gender effect was studied (Chapter 4: Experiment 3), as was a possible effect of gender on race categorisation (Chapter 5: Experiments 4, 5, and 6). The final chapter presents a summary of findings from these experiments and dataset analyses, alongside implications arising from the research, limitations, and potential directions for subsequent research. Chapters 2 through 5 are addressed sequentially in Sections 6.1 to 6.4.

6.1 Experience

6.1.1 Recap and implications

Chapter 1 explored whether the other-race experience disadvantage relates to the other-race effect (Experiment 1), and if the valance of other-race contact predicts other-race recognition (Experiment 2). Participants completed a number of questionnaires (PI20, IRCQ, and IMS/EMS) and a facial identity recognition task which featured East Asian, Caucasian, and African faces.

It was reasoned that the other-race effect may relate more strongly to the other-race experience disadvantage than other-race experience (Section 2.3, particularly Figure 2.2). Nonetheless, in Experiment 1, there were no associations between the other-race experience disadvantage (at any of the three childhood timespans of 0-6, 6-12, and 12-18 years) and the other-race effect. Consequently, results in Experiment 1 would suggest that childhood experience is not a factor in the other-race effect in adulthood.

For Experiment 2, it was suggested that previous research (Jerovich, 2017) may not have found relationships between other-race valenced contact and other-race recognition due to i) not accounting for face recognition ability and ii) using only one item each for positive and negative contact (Section 1.2.4.1). Experiment 2 used multiple regressions to examine the relationship, with the PI20 being used to control for general face recognition ability, and multiple items (rather than one) contributing to each valence score. Furthermore, IMS and EMS scores were included to control for motivations concerning prejudice. Despite this, positive and negatively valenced other-race contact were not predictive of other-race recognition. According to experiential views on the other-race effect (e.g., Rossion & Michel, 2011), experience improves other-race recognition, thereby minimising the other-race effect. Therefore, outcomes in Experiment 2 may indicate that neither positively nor negatively valenced childhood other-race contact would reduce the adulthood other-race effect.

6.1.2 Limitations and future research

6.1.2.1 Experience and eye-movements

As explained in Section 2.5.3, the placement of the fixation point may have attracted first fixations to the nose tip; previous research indicated that first fixations (reflective of dimension weightings) would ordinarily be driven by experience (e.g., Hills et al., 2013: Hills & Pake, 2013), therefore the fixation point may possibly have disrupted relationships between experience and recognition, and hence be why relationships were not found in Experiments 1 and 2. Accordingly, it is suggested that future research be conducted in the vein of Experiments 1-2, but without the use of a fixation point.

Subsequent research could find whether the first fixation location at *learning* or *test* drives the other-race effect, and such research could use East Asian faces. Fixation cross manipulations (in the context of the other-race effect) have been applied only at *both* learning and test phases, and, as noted in Section 2.5.4, it is unknown at which phase the fixation cross needs to appear in to have their effects, although previous research (Hills et al., 2011) would be suggestive of influences arising from either phase; future research on first fixations and the other-race effect could manipulate whether the fixation cross appears at learning or test.

The utility of a first fixation location for recognition may be driven by how good a location is for holistic processing (Hsiao & Cottrell, 2008) rather than being determined by relative variability. Given previous research on fixation locations and recognition (e.g., Hills & Pake, 2013), the location for optimal holistic processing may differ between races (of faces). Future research could manipulate first fixations to find whether first fixations influence face recognition via their effect on holistic processing. A composite task (e.g., Avidan, Tanzer, & Behrmann, 2011) could be employed, but researchers would need to be attentive to how faces would be divided. Bearing in mind the literature on fixation and identity (e.g., Hills et al., 2013; Hills & Lewis, 2011), and the link between recognition and holistic processing (Richler et al., 2011a), one would expect that first fixations to the upper face would be superior to lower face fixations for the holistic processing of Caucasians faces, and for the opposite pattern to occur for the faces of Blacks.

Section 2.5.4 considered if the use of FaceGen stimuli may have been a factor in not finding relationships concerning experience and recognition; as the use of FaceGen faces in research on the other-race effect is not uncommon (e.g., Chang et al., 2015; Pauker et al., 2009), it would be desirable to know whether any relationships between experience and recognition with real faces are of the same strength when FaceGen faces are used.

6.2 Objective and subjective facial variability

6.2.1 Recap and implications

Chapter 3 explored whether races/ethnicities differ in their facial morphological diversity, and whether other-race subjective homogeneity is a predictor of the other-race effect. Four datasets were utilised: W. W. Howells' cranial measurements (Howells, 1996) with respect to the skeletal face, and both the ANSUR and the ANSUR II (Gordon et al., 1989, 2014) pertaining to the full face/head, i.e., with tissues additional to bone, and the dataset of Bate, Bennetts et al. (2018b) with respect to exploring the potential bond between other-race subjective homogeneity and the other-race effect.

For the skeletal face, types of variability (mean variance, mean pattern variability, and standardised generalised variance) each negatively correlated with migratory distance, whilst there was no relationship between inter-dimension correlations and migratory distance after accounting for sample size. For any body part (let alone the face), no previous study had looked into associations between migratory distance and either pattern variability, standardised generalised variance, or between-trait correlations. The amount by which confidence intervals overlapped (Cumming & Finch, 2005) (regarding linear trend lines of confidence intervals) was used order to estimate the smallest dissimilarity in migratory distance where unequal facial diversity occurs (migratory distance thresholds). For the variance, pattern variability, and the standardised generalised variance, these thresholds were so great (20,000 km or in excess) that for each type of variability, the morphological diversity of ethnic groups would not generally differ, even when groups are from different races to each other. The lack of differences in diversity should also occur at the level of races. Generalised variances were compared with a nonparametric bootstrap (Petersen, 2000) concerning the full face/head for three races (Blacks, Caucasians, and Native Americans), and generalised variances were equivalent across them.

Rossion and Michel (2011) suggested that racial disparities in morphological variability can moderate the other-race effect. Outcomes in Chapter 3 indicated that races/ethnicities are, for the most part, of equal diversity in their facial structure. Nevertheless, trends of declining variability occur as migratory distance increases, and, furthermore, disparities do happen; racial differences in the structural diversity of the face may very well be a moderating variable in the other-race effect.

Outcomes in Chapter 3 may be of some importance for the future construction of police lineups. Principal component analysis could allow for police lineups to be formed with a level of likeness specified (Tredoux, 2002). An actual face can be transformed into a computer-generated 3D structure (such as from a 2D image) (e.g., Crookes et al., 2015); the similarity between structures can be controlled, and it has been envisioned that lineups could eventually feature well-tailored computer-generated faces (Segovia, Bailenson, & Leonetti, 2012).

With morphology in mind, equal variability for ethnicities/races could result in bias, depending on the value used for variability. If a universal (or default) variability level had to be set, the diversity found at a migratory distance of approximately 15,000 km (around that of East Asians) could be applied. This is because confidence interval overlaps in Chapter 3 would predict that an ethnic/race group at that distance would not differ from any others in regard to variability.

Whilst an effect of objective variability on the other-race effect remains speculative, some resolution was found in Chapter 3 concerning whether other-race subjective variability influences the other-race effect. An influence was found amongst Caucasian observers, however, it was not apparent for East Asian observers. These findings are crucial for understanding the mechanisms which underpin the other-race effect. The other-race effect may be considered to be a perceptually-driven occurrence (e.g., Hugenberg et al., 2010), yet the overall dearth of relationships between other-race expertise disadvantages and the other-race effect (e.g., DeGutis et al., 2013; Michel, Caldara, & Rossion, 2006; Zhao et al., 2014b) would suggest otherwise; based on analysis in Chapter 3, whilst it cannot be stated that the other-race effect is *primarily* perceptually-driven for Caucasian observers, a perceptual contribution is now clear. For East Asian observers however, it can be speculated that the other-race effect is principally storage- or retrieval-driven.

6.2.2 Limitations and future directions

6.2.2.1 The number of faces and dimensions

For the full face/head comparisons, there are drawbacks concerning sample sizes, and the number of structural dimensions. As covered in Chapter 3, few 18-

to-30-year-old Native American men and women featured in either the ANSUR or ANSUR II which leaves uncertain any conclusions drawn from their comparisons to Blacks and Caucasians. As not all skeletal facial elements exhibit a reduction in variability as migratory distance lengthens (Manica et al., 2007), dimensions in the full face/head would vary in how much they each decline with migratory distance (with some presumably not displaying the trend). In Chapter 3, eight and five dimensions were used; in the event that measurement dimensions used in that chapter were unrepresentative of the general facial variability decline, it is encouraged that the nonparametric bootstrap be applied to racial comparisons using a greater number of morphological dimensions. For instance, the ANSUR headboard-derived data (Gordon et al., 1989) or the National Institute for Occupational Safety and Health anthropometric dataset (Zhuang & Bradtmiller, 2005) could be analyzed.

6.2.2.2 Dimension comparisons

Anthropometric research might also compare the diversity of individual facial dimensions (or regions) of the face *within* different races. The relative variability of a facial trait could determine its utility in face recognition, and the emphasis placed on dimensions when encoding facial identity would typically be based largely on own-race experience (Ellis, 1975; Valentine, 1991). If races differ in the relative usefulness of dimensions for individuation, applying dimension selection strengths which are optimal for own-race faces onto other-race faces could be a path to the other-race effect (Hills et al., 2013; Hills & Lewis, 2011). Fang et al. (2011) used 95% confidence intervals regarding coefficients of variation to compare the diversity of morphological dimensions (i.e., one aspect of the face versus another), with attributes considered as being of unequal variability when there was no overlap of their confidence intervals. Coefficients of variation could be bootstrapped for measurements within races, and confidence intervals created to give an idea of the usefulness of attributes for individuating the faces of a race group. This would indicate whether the dimensions which are most useful for telling apart faces in one race group are similarly applicable in another race.

6.2.2.3 Objective variability: Behavioural

The decline in morphological variability and the existence of some ethnic and racial differences in structural variability do not mean that variability differences are actually a contributor to the other-ethnicity/race effect. Furthermore, there are factors outside of morphology which can be reflected on in terms of racial/ethnic variability differences; the colour of the iris is considered to be relatively heterogeneous amongst Eastern and Northern Europeans (Frost, 2006), and, pertaining to the skin colour of the arm, Africans are more variable than Europeans (Relethford, 2000) which could be speculated to extend to the face.

If variability is a moderator of the other-race effect (Rossion & Michel, 2011), one would expect the magnitude of the other-ethnicity/race effect to be larger for the population which are theorised to be physically heterogeneous. Nonetheless, just as motivation can overshadow the relationship between experience and the other-race effect (Young & Hugenberg, 2012), variability differences could be eclipsed by motivation and also by experience. Therefore, the possible moderating role of variability could be tested in experiments where the magnitude of the other-race effect is the dependent variable, and the independent variables are observer race and the other-race experience disadvantage; motivation to individuate other-race faces could be manipulated (e.g., the individuation encouragement method used in Young & Hugenberg, 2012, Experiment 1), such that the moderating effect of variability (controlling for the other-race experience disadvantage) could be tested under conditions of relatively low and high motivation for individuating other-race faces. In Young and Hugenberg (2012), other-race experience related to the other-race effect when motivation was heightened, but not at baseline levels. Consequently, it is speculated that the moderating effect of variability would be apparent (or more so) when motivation is high.

6.2.2.4 Expertise

Given the analysis of the Bate, Bennetts et al. (2018b) data in Chapter 3, it seems somewhat perplexing that associations between other-race expertise disadvantages and the other-race effect have not been forthcoming. A suggestion may be to include participant race as a moderator (Hancock & Rhodes, 2008, did so concerning the relationship between the other-race disadvantage in the face inversion effect and the other-race effect), or to analyse relationships in participant race groups separately; when exploring links between expertise disadvantages and the other-race effect, it is not uncommon for race groups to be pooled together without any differentiation (e.g., Zhao et al., 2014b, Experiment 1), yet there are instances of Caucasian groups being analysed singularly wherein relationships have not been apparent (Michel, Rossion, et al., 2006; Zhao et al., 2014b Experiment 2).

Still, as the other-race effect may not be equally perceptually-driven across race groups, it would seem reasonable to distinguish between race groups. Indeed, Zhao et al. (2014b) Experiment 1, in which Caucasian and East Asian observers were combined, remains the seldom study to test whether the other-race configural disadvantage relates to the other-race effect. It could be that this particular expertise disadvantage is critical for the other-race effect amongst Caucasians, but not East Asians, and the absence of a relationship was due to pooling. Yet the small correlation coefficients (still combined) presented in that study (*r*s < .10) may tentatively indicate that a relationship would not have been apparent even with a separated Caucasian sample; as stated in Section 1.2.6, it could be some combination of expertise disadvantages which matters, or interactions between types of expertise disadvantages. Future research could focus on such combinations and interactions in a controlled manner (rather than using the face inversion effect).[43] Indeed, the raw data for such an analysis would exist from Zhao et al. (2014b), wherein participants completed various expertise tasks and versions of the CFMT (Caucasian, and East Asian).

6.3 Gender categorisation and race

6.3.1 Recap and implications

Experiment 3 concerned the other-race gender effect and possible routes which it may arise from. This experiment used two questionnaires (PI20 and IRCQ), and two gender categorisation tasks. Participants and stimulus faces were Caucasian, East Asian, and Black. In line with previous research (O'Toole et al., 1996; Zhao & Bentin, 2008), Caucasian faces were more perceptually distinctive in terms of the difference concerning male and female than East Asian faces, regardless of observer race. Similarly, males and females were more subjectively

[43] A concern might be multicollinearity, however, expertise disadvantages have generally been uncorrelated in previous research (Rhodes, Ewing et al., 2009; Zhao et al., 2014b), which may signal that multicollinearity would not be present between the types of expertise.

dissimilar in Caucasian faces than Black faces. Results suggested that these patterns extended to aperture-view trials (i.e., expertise for local face areas was greater for Caucasian faces). As in prior research (Zhao & Hayward, 2010), an other-race holistic disadvantage for gender was not present.

Theory-wise, the outcomes of Experiment 3 suggest that the underpinnings of the other-race gender effect and the other-race effect have differences and similarities concerning experience, expertise, and dissimilarity. Regarding experience, it is uncertain over whether childhood contact has an association with the adulthood other-race effect (Davis et al., 2016; Wan et al., 2013; Wan et al., n.d.) even when that experience is quantified as an other-race experience disadvantage (Wan et al., 2013; see Section 2.3); in Experiment 3, there was no relationship between the experience disadvantage and the other-race gender effect. It is unclear whether there are relationships between other-race experience and other-race disadvantages in the expert processing of identity (Bukach et al., 2012; Rhodes, Ewing, et al., 2009; Zhao et al., 2014b); Experiment 3 used the other-race experience disadvantage in place of other-race experience. Nonetheless, relationships between experience and other-race disadvantages in expert processing were not evident. Overall, results do not favour the other-race gender effect being influenced by experience (although see Section 4.4.3).

Whilst expertise disadvantages do not seem to relate to the other-race effect (e.g., Horry et al., 2015; Michel, Caldara, & Rossion, 2006; Zhao et al., 2014b), a relationship between the disadvantage in local processing (yet not holistic processing) and the other-race gender effect was clear.

Pertaining to dissimilarity, Caucasians, East Asians, and Blacks would seem to be of equal morphological variability in terms of identity (Chapter 3).

The results of Chapter 4 may indicate that there is larger dimorphism in a local facial region (or regions) of Caucasian faces than East Asian and Black faces, which translates into greater gender discriminability for Caucasian faces.

6.3.2 Limitations and future directions

6.3.2.1 Non-engagement

In Section 4.4.2, it was suggested that the use of an oval may have removed facial details which may ordinarily be useful for gender discrimination, such as in the jaw (e.g., Brown & Perrett, 1993), and that these may be particularly useful for categorising the gender of Black faces (oval use may have caused the female bias in the gender categorisation of Black faces in Black participants). Of the three previous studies on the other-race gender/sex effect (none of which presented Black faces), two used ovals: part of the jaw area was occluded in one study (Zhao & Hayward, 2010), but example stimuli were not presented in the other (O'Toole et al., 1996) so the situation regarding the jaw in that study is unclear; Zhao and Bentin (2008) certainly did not obscure the jaw area in their stimuli, and it is encouraged that future research follow suit.

Of the 16 non-engaged participants in the categorisation task, six were in regard to Caucasian faces, eight to East Asian faces, and 13 to Black faces. It should be noted that a participant was removed even if they were non-engaged in response to faces of one race. The use of an oval could have led to the relatively high non-engagement rate. Indeed, removals seem to be greater for Black faces than faces of the other two races; this could be attributed once more to oval usage being particularly disruptive pertaining to Black faces.

Prior to Experiment 3, a pilot perceptual adaptation experiment was completed by six Caucasian participants (i.e., about one third the target sample size for a race in Experiment 3), and it featured the Caucasian and East Asian gender-morphed faces used in Experiment 3 (with similar ovals to the ones of Experiment 3) for which gender categorisation decisions were made, and the same stimulus presentation duration was used as employed in Experiment 3 (500 ms); visual inspection of the data suggested that engagement difficulties were not apparent, whether in baseline or adaptation conditions (blocks were multiracial). In the categorisation task of Experiment 3, however, of the non-engaged participants three were Caucasian (13.64% of 22 Caucasians), five were East Asian (20.83% of 24 East Asians), and eight were Black (33.33% of 24 Blacks). It is not unheard of for there to be a relatively high removal of participants due to performance issues, for example, 16.67% of participants (10 out of 60) in Experiment 1 of Curby, Goldstein, and Blacker (2013) which featured face and car stimuli, and 23.08% (15 out of 65) in Experiment 4 of Susilo, Rezlescu, and Duchaine (2013) who used face stimuli. However, it is unknown why non-engagement would seem to have been greater for non-Caucasian participants than Caucasians, in particular Black ones. It could be speculated that this was due to oval usage and observer racial differences in how weightings are used for gender categorisation.

6.3.2.2 Expertise

Performance on aperture-view trials has been construed as reflecting local processing (Murphy & Cook, 2017); as some configural information (e.g., the distance between the eyes) and featural details would have been preserved by the aperture, it is unclear which type of other-race expertise disadvantage (configural or featural) related to the other-race gender effect. Future research could test whether an other-race featural disadvantage or other-race configural disadvantage for gender mediate a (possible) relationship between the other-race experience disadvantage and the other-race gender effect. As in Zhao and Hayward (2010), the scrambling paradigm could be used to measure the featural processing of gender; blurring (Hayward et al., 2008) could be employed regarding configural gender processing.

As could be undertaken regarding expertise and the other-race effect (Section 6.2.2.4), one could find whether other-race expertise disadvantages (and their interactions) in gender categorisation predict the other-race gender effect in a multiple regression analysis. Bearing in mind that other-race subjective homogeneity predicts the other-race effect for Caucasians and not East Asians (Section 3.2.2), future research may test (potential) associations between the other-race gender effect and potential contributors (such as expertise) for each observer race separately, or/and test if they are similar across each race of observer.

6.3.2.3 Bias

Experiment 3 demonstrated an association between the other-race gender categorisation bias and the other-race gender effect. However, it is uncertain where such a bias arises from. For insight into whether gender categorisation biases stem from the extent of similarity between other-race faces and own-race gender norms, similarity ratings between own- and other-race faces could be subject to multidimensional scaling to provide distances between categories (i.e., as in Byatt & Rhodes, 2004, but using female as well as male faces). If other-race faces hold a greater resemblance to an own-race gender norm (e.g., male more so than female), one could expect a categorisation bias; future research could examine whether the relative closeness of other-race faces to own-race gender norms predicts the other-race PSE.

6.4 Race categorisation

6.4.1 Recap and implications

Experiments 4, 5, and 6 collectively concerned whether race categorisation is affected by the gender of faces and facial orientation. Experiment 4 aimed to discover if facial gender affects how upright faces are racially categorised in terms of race category boundaries and the precision with which race categorisation occurs. Experiment 4 directly demonstrated gender effects on race category boundaries, with Caucasian/Black and Caucasian/East-Asian boundaries being a phenotype which was less prototypically Caucasian (for the gender) for female faces than male faces. Additionally, race categorisation for faces of Caucasian and East Asian ancestry was more proficient (lower decision noise) for male faces than female faces, whilst proficiency for male faces and female faces was alike for faces of Caucasian and Black ancestry. Gender categorisation ability was predictive of stimulus gender affecting Caucasian/Black categorisation proficiency, unlike Caucasian/East-Asian decision noise, and both Caucasian/Black and Caucasian/East-Asian decision noise.

Experiments 5 and 6 were extensions of Experiment 4. The sixth experiment presented faces upright and upside down in order to find whether facial orientation affects the influence of gender on the Caucasian/Black category boundary, and race categorisation proficiency. For female faces, inversion caused the race category boundary to become a less prototypically Caucasian face, thereby increasing the size of the gender effect on the race categorisation boundary; as for male faces, the race category boundary was unaffected by inversion. This tentatively suggests that, when racially categorising

(Caucasian/Black), observers try to adjust for the relatively lighter skin tone of females and not the relatively darker skin tone of males, hence the skin tone of monoracial males may be perceived as more prototypical for races than the skin tone of monoracial females (Section 5.4.2.1). Inversion made race categorisation more difficult, thereby suggesting that an increased reliance on skin tone does not wholly compensate for the weakened perception of facial structure under inversion, although it should be noted that faces were shown solely in greyscale (Section 5.4.2.2). Experiment 6 manipulated the luminance of faces in order to see if the gender difference concerning the Caucasian/Black boundary arose from females having a higher facial luminance than males. No influence of luminance was demonstrated, which indicated that there is another reason for females and males having different race category boundaries; morphology is possibly that reason (Section 5.4.3).

6.4.2 Limitations and future directions

6.4.2.1 Participant race and experience

Not accounting for the race and experience of observers may have minimised the effect of orientation on race categorisation in Experiment 5, such that there was no effect on the race category boundary of male faces. Face inversion decreases the distinctiveness of faces (Leder et al., 2017), i.e., it increases the perceived similarity to a norm. This overall norm would be determined by experience, and oftentimes be more representative of the own-race category than an other-race category (Valentine, 1991). Therefore, by inversion disrupting the perception of morphological cues (Cloutier & Macrae, 2007; Yovel & Kanwisher, 2004), faces may be perceived as more similar to the overall face norm, thereby encouraging own-race categorisations if an observer has predominantly own-race experience; inversion may shift own/other-race category boundaries away from the own-race norm (on the racial morph continuum). As mentioned in Section 5.2.1, inversion may have encouraged White categorisations in Willenbockel et al. (2011) for faces which were morphed midway between Caucasian and Black faces; participants in that study were Caucasians with “little or no contact with Black people” (Willenbockel et al., 2011, p. 623), and assumedly in North America, therefore such results would tentatively align with inversion having caused faces to seem more own-race.

6.4.2.2 Racial prototypicality

In Section 5.4.1.1, it was theorised that gender differences in racial prototypicality were the cause of the gender difference in race category boundaries. This idea could be explored by research on whether the gender effects on race category boundaries are predicted by differences in race categorisation response latencies. For instance, pertaining to faces of Black and Caucasian ancestry, the male-minus-female race category boundary difference could be predicted by response latencies (e.g., the average latencies of female-minus-male faces concerning Black faces and male-minus-female faces regarding Caucasian faces).

6.4.2.3 Bias

The effect of inversion on the race category boundary for female, but not male, faces (Experiment 5) may suggest that *male* is generally used as the default gender in Caucasian/Black race categorisation, and, due to inversion increasing the difficulty with which gender/sex is categorised (Cloutier & Macrae, 2007; Zhao & Hayward, 2010), inversion increases the reliance on a default gender, i.e., inverted faces are perceived as more male than upright faces. Miscategorising female faces as male could explain why facial orientation affected the Caucasian/Black race category boundaries for females and not males; there may have been a diminished adjustment for the lighter skin tone of females (whilst there was no adjustment regarding male skin tone whether faces were upright or inverted).

Overall, it would seem inconclusive whether there is a gender categorisation bias for Caucasian or Black faces. A bias (response criterion) for categorising upright Caucasian faces as male has been found to be present for upright Caucasian faces in previous research (DeGutis et al., 2012). Yet, in Experiment 3 regarding upright Caucasian faces, no bias (gender PSE) was present for Caucasian participants, (means and *SD*s for East Asian and Black participants would be indicative of a bias not having occurred in their responses to Caucasian faces) (Section 4.3.3). As for Black faces, Experiment 3 found a bias (PSE) for faces to be categorised female amongst Black participants, and means would suggest that this bias occurred for Caucasian participants, but perhaps not for East Asian participants (Section 4.3.3).

As for whether inversion introduces a male categorisation bias, the Caucasian race norm is possibly more female than male (Section 1.4.2), and there is evidence favouring inverted faces seeming more average than upright faces (Leder, Goller, Forster, Schlageter, & Paul, 2017), therefore one could expect that inversion would cause a female categorisation bias (shifting the boundary towards the Caucasian male prototype past the 50% point on the continuum). For Caucasian faces, the male response criterion bias (which is present at an upright orientation) does not occur under inversion (DeGutis, Chatterjee, et al., 2012). This implies that inversion causes Caucasian faces to be perceived as less male (more female) compared to an upright orientation, i.e., inversion does not cause a female bias, but reduces the male categorisation bias. Hence, although not causing a bias, inversion would still cause the gender category boundary to be moved to a more prototypically male Caucasian face. The Black norm may be weighted more towards Black males than Black females (Section 1.4.2); inversion may promote male categorisations amongst Black faces.

Therefore, whilst inversion could have different effects on the gender category boundaries of Caucasian faces and Black faces, the results of Experiment 5 may reflect that an overall male categorisation bias was introduced by inversion (for Caucasian-to-Black faces), and that this introduction affected the race category boundary for female faces. Future research could i) explore whether inversion causes a bias in gender categorisation (or otherwise affects the boundary) and not only using Caucasian faces, but also faces of Blacks and other races, and ii) find if the influence of inversion on gender categorisation PSEs (perhaps as an average of Caucasian and Black gender PSEs) would predict the effect of inversion on the Caucasian/Black PSE of female faces.

6.4.2.4 Morphology and skin tone

Future research could gain further insight into whether facial morphology or skin tone drive gender differences in race category boundaries and proficiency. Experiments 4, 5, and 6 did not display faces in colour; depictions in greyscale may be somewhat removed from the everyday (i.e., in colour) (Anzures et al., 2011), therefore future research could use colour. At each level of a racial morph continuum, the average skin tone within a gender at a morph level could be manipulated to have the skin tone of another gender at that morph level (e.g., a 15% Black and 85% Caucasian male face can have the skin tone of a 15% Black and 85% Caucasian female face). For a type of ancestry (e.g., Black/Caucasian), there could be four racial morph continua: two with unaltered skin tones (one male continuum, one female continuum), a male continuum with skin tone adjusted to the skin tone of a typical female face at each morph level, and a female continuum with typical male skin tones per morph level. An ANOVA testing for main effects (gender, skin tone) and interactions would indicate whether morphology or skin tone define the influence which gender has on race categorisation.

6.4.2.5 Multiracial categorisation

The occurrence of gender affecting self-categorisation as multiracial (compared to Asian, Black, or Latino categorisations in the USA) (Davenport, 2016) was crucial in motivating Chapter 5 regarding race category boundaries. However, Experiments 4-6 only explored monoracial race category boundaries rather than monoracial/multiracial boundaries. Given the use of multiracial categorisations (Davenport, 2016; Ho et al., 2011), solely using monoracial categories would leave the race categorisation tasks of Experiments 4, 5 and 6 limited when compared to real life situations. An effect of gender on monoracial/multiracial race category boundaries of faces could be explored with a paradigm from Ho et al. (2011) (wherein participants adjusted phenotypes of male faces until they considered the face to be a member of a particular category) but with female faces in addition to male faces.

6.4.2.6 Oval

It could very well be that, as may potentially have occurred with stimuli in Experiment 3 (Section 4.4.2), presenting the face in an oval in Experiment 5 may have removed facial qualities which are typically used in gender categorisation.

Therefore, the oval may have affected gender categorisation in Chapter 5 compared to if faces were presented in full. In Experiment 3, for Black observers, there was a bias for categorising Black faces as female (Section 4.3.3), which could possibly have been due to the oval being used (Section 4.4.2). This, in particular, raises the possibility that effects of gender on the Caucasian/Black race category in Chapter 5 could have been underestimated compared to if faces were presented in full. Regarding the future research ideas suggested earlier in Section 6.4, presenting faces in full rather than using an oval could be considered.

6.5 Conclusion

The present thesis approached *the face and race* from a diverse perspective. It explored two *other*- effects pertaining to race: facial identity recognition (the other-race effect, Chapters 2-3) and gender categorisation (the other-race gender effect, Chapter 4). Additionally, as somewhat of a converse to race influencing gender categorisation, the thesis also sought insight regarding effects of gender on race categorisation (Chapter 5). Whilst contributions were made pertaining to the recognition and categorisation of faces in the context of race, future research is certainly encouraged perhaps along the lines of the ideas described in this chapter.

REFERENCES

Abdi, H. (2010). Holm's sequential Bonferroni procedure. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 573–577). Thousand Oaks, CA: Sage.

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., . . . Schaller, M. (2006). They all look the same to me (unless they're angry): From out-group homogeneity to out-group heterogeneity. *Psychological Science*, *17*(10), 836–840.

Adams, R. B., Jr., Pauker, K., & Weisbuch, M. (2010). Looking the other way: The role of gaze direction in the cross-race memory effect. *Journal of Experimental Social Psychology*, *46*(2), 478–481.

Adams, W. J., Gray, K. L. H., Garner, M., & Graf, E. W. (2010). High-level face adaptation without awareness. *Psychological Science*, *21*(20), 205–210.

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

Anzures, G., Pascalis, O., Quinn, P. C., Slater, A. M., & Lee, K. (2011). Minimizing skin color differences does not eliminate the own-race recognition advantage in infants. *Infancy*, *16*(6), 640–654.

Arifin, W. N. (2015). The graphical assessment of normality using SPSS. *Educational Resource*, *7*(2), e71–e75.

Ashworth, A. R. S., III, Vuong, Q. C., Rossion, B., & Tarr, M. J. (2008). Recognizing rotated faces and Greebles: What properties drive the face inversion effect? *Visual Cognition*, *16*(6), 754–784.

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

Baldwin, M., Keefer, L. A., Gravelin, C. R., & Biernat, M. (2013). Perceived importance of cross-race targets facilitates recall: Support for a motivated account of face memory. *Group Processes and Intergroup Relations*, *16*(4), 505–515.

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R. M., Harwood, J., . . . Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629–1643.

Basto, M., & Pereira, J. M. (2012). An SPSS R-menu for ordinal factor analysis. *Journal of Statistical Software*, *46*(4), 1–29. doi:10.18637/jss.v046.i04

Basu, A., Sarkar-Roy, N., & Majumder, P. P. (2016). Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(6), 1594–1599.

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2018a). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/xhp0000607

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2018b). *Super-recognisers and controls – own-race and other-ethnicity memory and matching dataset* [Dataset]. Retrieved from https://brunel.figshare.com/articles/Super-recognisers_and_controls_-_own-race_and_cross-race_memory_and_matching_dataset/6752825

Bate, S., Bennetts, R., Murray, E., & Portch, E. (2020). Enhanced matching of children's faces in "super-recognisers" but not high-contact controls. *i-Perception*, *11*(4), 1–12.

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., Wills, H., & Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*(22), 1–19.

Baudouin, J.-Y., & Gallay, M. (2006). Is face distinctiveness gender based? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 789–798.

Baudouin, J.-Y., & Humphreys, G. W. (2006). Configural information in gender categorisation. *Perception*, *35*(4), 531–540.

Benn-Torres, J., Bonilla, C., Robbins, C. M., Waterman, L., Moses, T. Y., Hernandez, W., . . . Kittles, R. A. (2008). Admixture and population stratification in African Caribbean populations. *Annals of Human Genetics*, *72*(1), 90–98.

Benton, C. P., & Skinner, A. L. (2015). Deciding on race: A diffusion model analysis of race-categorisation. *Cognition*, *139*, 18–27.

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

Betti, L., Balloux, F., Amos, W., Hanihara, T., & Manica, A. (2009). Distance from Africa, not climate, explains within-population phenotypic diversity in humans. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1658), 809–814.

Betti, L., von Cramon-Taubadel, N., Manica, A., & Lycett, S. J. (2013). Global geometric morphometric analyses of the human pelvis reveal substantial neutral population history effects, even across sexes. *PLoS ONE*, *8*(2), 1–10. doi:10.1371/journal.pone.0055909

Bhopal, R. (2004). Glossary of terms relating to ethnicity and race: For reflection and debate. *Journal of Epidemiology & Community Health*, *58*(6), 441–445.

Bickel, P. J., Götze, F., & van Zwet, W. R. (1997). Resampling fewer than *n* observations: Gains, losses, and remedies for losses. *Statistica Sinica*, *7*, 1–31.

Biotti, F., Cook, R. (2016). Impaired perception of facial emotion in developmental prosopagnosia. *Cortex*, *81*, 126-136.

Biotti, F., Gray, K. L. H., & Cook, R. (2017). Impaired body perception in developmental prosopagnosia. *Cortex*, *93*, 41–49.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin*, *15*(1), 19–25.

Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Memory & Cognition*, *34*(4), 854–864.

Brewer, R., Biotti, F., Bird, G., & Cook, R. (2017). Typical integration of emotion cues from bodies and faces in Autism Spectrum Disorder. *Cognition*, *165*, 82–87.

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In R. Lindsay, D. Ross, J. Read, & M. Toglia (Eds.), *Handbook of eyewitness psychology* (pp. 257–281). Mahwah, NJ: Lawrence Erlbaum & Associates.

Brigham, J. C., & Ready, D. J. (1985). Own-race bias in lineup construction. *Law and Human Behavior*, *9*(4), 415–424.

Brittingham, A., & de la Cruz, G. P. (2004). *Ancestry: 2000.* Retrieved from https://www.census.gov/prod/2004pubs/c2kbr-35.pdf

Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22*(7), 829–840.

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

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *The American Journal of Human Genetics*, *96*(1), 37–53.

Bukach, C. M., Cottle, J., Ubiwa, J., & Miller, J. (2012). Individuation experience predicts other-race effects in holistic processing for both Caucasian and Black participants. *Cognition*, *123*(2), 319–324.

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

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

Byatt, G., & Rhodes, G. (2004). Identification of own-race and other-race faces: Implications for the representation of race in face space. *Psychonomic Bulletin & Review*, *11*(4), 735–741.

Carbon, C.-C., Grüter, M., & Grüter, T. (2013). Age-dependent face detection and face categorization performance. *PLoS ONE*, *8*(10), 1–7. doi:10.1371/journal.pone.0079164

Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin*, *41*(3), 405–419.

Cassidy, K. D. (2011). *The other-race effect in face perception and recognition: Contributions of social categorisation and processing strategy* (Unpublished doctoral thesis). University of Birmingham, UK.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Cellerino, A., Borghetti, D., & Sartucci, F. (2004). Sex differences in face gender recognition in humans. *Brain Research Bulletin*, *63*(6), 443–449.

Chang, A., Murray, E., & Yassa, M. A. (2015). Mnemonic discrimination of similar face stimuli and a potential mechanism for the "other race" effect. *Behavioral Neuroscience*, *129*(5), 666–672.

Chatterjee, G., & Nakayama, K. (2012). Normal facial age and gender perception in developmental prosopagnosia. *Cognitive Neuropsychology*, *29*(5–6), 482–502.

Cheng, C.-Y., Reich, D., Haiman, C. A., Tandon, A., Patterson, N., Selvin, E., . . . Kao, W. H. L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: A genetic admixture analysis in three U.S. population cohorts. *PLoS ONE*, *7*(3), 1–9. doi:10.371/journal.pone.0032840

Chiroro, P. M., Tredoux, C. G., Radaelli, S., & Meissner, C. A. (2008). Recognizing faces across continents: The effect of within-race variations on the own-race bias in face recognition. *Psychonomic Bulletin & Review*, *15*(6), 1089–1092.

Chiroro, P., & Valentine, V. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology*, *48A*(4), 879–894.

Cloutier, J., Li, T., & Correll, J. (2014). The impact of childhood experience on amygdala response to perceptually familiar Black and White faces. *Journal of Cognitive Neuroscience*, *26*(9), 1992–2004.

Cloutier, J., & Macrae, C. N. (2007). Who or what are you?: Facial orientation and person construal. *European Journal of Social Psychology*, *37*(6), 1298–1309.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Colombatto, C., & McCarthy, G. (2017). The effects of face inversion and face race on the P100 ERP. *Journal of Cognitive Neuroscience*, *29*(4), 664–676.

Corneille, O., Hugenberg, K., & Potter, T. (2007). Applying the attractor field model to social cognition: Perceptual discrimination is facilitated, but memory is impaired for faces displaying evaluatively congruent expressions. *Journal of Personality and Social Psychology*, *93*(3), 335–352.

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research*, *3*(1), 58–67.

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, *102*(4), 874–888.

Crookes, K., Ewing, L., Gildenhuys, J.-D., Kloth, N., Hayward, W. G., Oxner, M., . . . Rhodes, G. (2015). How well do computer-generated faces tap face expertise? *PloS ONE*, *10*(11), 1–18. doi:10.1371/journal.pone.0141353

Crookes, K., Favelle, S., & Hayward, W. G. (2013). Holistic processing for other-race faces in Chinese participants occurs for upright but not inverted faces. *Frontiers in Psychology, 4,* 1–9. doi:10.3389/fpsyg.2013.00029

Crookes, K., & Rhodes, G. (2017) Poor recognition of other-race faces cannot always be explained by a lack of effort. *Visual Cognition*, *25*(4–6), 430–441

Cumming, G. (2008). Replication and *p* intervals: *p* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*(4), 286–300.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170–180.

Curby, K. M., Goldstein, R. R., & Blacker, K. (2013). Disrupting perceptual grouping of face parts impairs holistic face processing. *Attention, Perception, & Psychophysics*, *75*(1), 83–91.

Curby, K. M., Johnson, K. J., Tyson, A. (2012). Face to face with emotion: Holistic face processing is modulated by emotional state. *Cognition and Emotion*, *26*(1), 93–102.

Davenport, L. D. (2016). The role of gender, class, and religion in biracial Americans' racial labeling decisions. *American Sociological Review*, *81*(1), 57–84.

Davis, M. M., Hudson, S. M., Ma, D. S., & Correll, J. (2016). Childhood contact predicts hemispheric asymmetry in cross-race face processing.

Psychonomic Bulletin & Review, *23*(3), 824–830.

De Rigal, J., Des Mazis, I., Diridollou, S., Querleux, B., Yang, G., Leroy, F., & Barbosa, V. H. (2010). The effect of age on skin color and color heterogeneity in four ethnic groups. *Skin Research and Technology*, *16*(2), 168–178.

DeGutis. J., Chatterjee, G., Mercado, R. J., & Nakayama, K. (2012). Face gender recognition in developmental prosopagnosia: Evidence for holistic processing and use of configural information. *Visual Cognition*, *20*(10), 1242–1253.

DeGutis, J., DeNicola, C., Zink, T., McGlinchey, R., & Milberg, W. (2011). Training with own-race faces can improve processing of other-race faces: Evidence from developmental prosopagnosia. *Neuropsychologia*, *49*(9), 2505–2513.

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), 1–13. doi:10.1371/journal.pone.0058253

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

Ding, M., Bressler, S. L., Yang, W., & Liang, H. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: Data preprocessing, model validation, and variability assessment. *Biological Cybernetics*, *83*(1), 35–45.

Dobel, C., Bölte, J., Aicher, M., & Schweinberger, S. R. (2007). Prosopagnosia without apparent cause: Overview and diagnosis of six cases. *Cortex*, *43*(6), 718–733.

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.

Dunham, Y., Dotsch, R., Clark, A. R., & Stepanova, E. V. (2016). The development of White-Asian categorization: Contributions from skin color and other physiognomic cues. *PLoS ONE*, *11*(6), 1–20. doi:10.1371/journal.pone.0158211

Dunham, Y., Stepanova, E. V., Dotsch, R, & Todorov, A. (2015). The development of race-based perceptual categorization: Skin color dominates early category judgements. *Developmental Science*, *18*(3), 469–483.

Edler, R., Rahim, M. A., Wertheim, D., & Greenhill, D. (2010). The use of facial anthropometrics in aesthetic assessment. *The Cleft Palate-Craniofacial Journal*, *47*(1), 48–57.

Ellis, H. D. (1975). Recognizing faces. *British Journal of Psychology*, *66*(4), 409–426.

Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., & Bülthoff, I. (2016). Face perception and test reliabilities in congenital prosopagnosia in seven tests. *i-Perception*, *7*(1), 1–37. doi:10.1177/2041669515625797

Fang, F., Clapham, P. J., & Chung, K. C. (2011). A systematic review of inter-ethnic variability in facial dimensions. *Plastic and Reconstructive Surgery*, *127*(2), 874–881.

Farkas, L. G., Katic, M. J., & Forrest, C. R. (2005). International anthropometric study of facial morphology in various ethnic groups/races. *The Journal of*

Craniofacial Surgery, *16*(4), 615–646.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fehr, J., & Sassenberg, K. (2010). Willing and able: How internal motivation and failure help to overcome prejudice. *Group Processes & Intergroup Relations*, *13*(2), 167–181.

Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology*, *2*, 1–4. doi:10.1186/1471-2288-2-8

Ferguson, D. P., Rhodes, G., Lee, K., & Sriram, N. (2001). 'They all look alike to me': Prejudice and cross-race face recognition. *British Journal of Psychology*, *92*(4), 567–577.

Ferrario, V. F., Sforza, C., Pizzini, G., Vogel, G., & Miani, A. (1993). Sexual dimorphism in the human face assessed by euclidean distance matrix analysis. *Journal of Anatomy*, *183*(3), 593–600.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics (4th ed.)*. London: Sage Publications.

Fish, J. M. (2009, December 23). *How come other folks all look alike? Do whites look more different from one another than blacks?* Retrieved from https://www.psychologytoday.com/blog/looking-in-the-cultural-mirror/200912/how-come-other-folks-all-look-alike

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS ONE*, *6*(9), 1-10. doi:10.1371/journal.pone.0025107

Friedlaender, J. S. (1975). *Patterns of human variation: The demography, genetics and phenetics of Bougainville Islanders*. Cambridge, MA: Harvard University Press.

Friedlaender, J. S., Friedlaender, F. R., Reed, F. A., Kidd, K. K., Kidd, J. R., Chambers, G. K., . . . Weber, J. L. (2008). The genetic structure of Pacific Islanders. *PLoS Genetics*, *4*(1), 0173–0190. doi:10.1371/journal.pgen.0040019

Friedlaender, J. S., Sgaramella-Zonta, L. A., Kidd, K. K., Lai, L. Y. C., Clark, P., & Walsh, R. J. (1971). Biological divergences in south-central Bougainville: An analysis of blood polymorphism gene frequencies and anthropometric measurements utilizing tree models, and a comparison of these variables with linguistic, geographic, and migrational "distances". *The American Journal of Human Genetics*, *23*(3), 253–270.

Frost, P. (2006). European hair and eye color: A case of frequency-dependent sexual selection? *Evolution and Human Behavior*, *27*(2), 85–103.

Gaetano, J. (2013). Holm-Bonferroni sequential correction: An EXCEL calculator. Retrieved from https://www.researchgate.net/publication/236969037_Holm-Bonferroni_Sequential_Correction_An_EXCEL_Calculator

Gaither, S. E., Chen, J. M., Pauker, K., & Sommers, S. R. (2019). At face value: Psychological outcomes differ for real vs. computer-generated multiracial faces. *The Journal of Social Psychology*, *159*(5), 592–610.

Garn, S. M., Lavelle, M., & Smith, B. H. (1985). Quantification of dysmorphogenesis: pattern variability index, σz. *American Journal of Roentgenology*, *144*(2), 365–369.

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Cambridge, MA: Harvard University Press.

Gates, A. (2008). *Face recognition: Establishing the effect of positive emotion on eliminating the cross-race deficit* (Unpublished master's thesis). University College London, UK.

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., . . . Lee, K. (2009). Two faces of the other-race effect: Recognition and categorisation of Caucasian and Chinese faces. *Perception*, *38*(8), 1199–1210.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, *118*, 201–210.

Goff, P. A., Thomas, M. A., & Jackson, M. C. (2008). “Ain’t I a Woman?”: Towards an intersectional approach to person perception and group-based harms. *Sex Roles*, *59*(5–6), 392–403.

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 995–1001.

Goldstein, A. G. (1979a). Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society*, *13*(3), 187–190.

Goldstein, A. G. (1979b). Facial feature variation: Anthropometric data II. *Bulletin of the Psychonomic Society*, *13*(3), 191–193.

Goldstein, A. G., & Chance, J. (1976). Measuring psychological similarity of faces. *Bulletin of the Psychonomic Society*, *7*(4), 407–408.

Goldstein, A. G., & Chance, J. (1978). Judging face similarity in own and other races. *The Journal of Psychology*, *98*(2), 185–193.

Goldstein, A. G., & Chance, J. (1979). Do "foreign" faces really look alike? *Bulletin of the Psychonomic Society*, *13*(2), 111–113.

Goldstein, A. G., & Chance, J. E. (1980). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, *23*(3), 211–214.

Gordon, C. C., Blackwell, C. L., Bradtmiller, B., Parham, J. L., Barrientos, P., Paquette, S. P., . . . Kristensen, S. (2014). *2012 Anthropometric Survey of U.S. Army Personnel: Methods and summary statistics*. Natick, MA: U.S. Army Natick Soldier Research, Development and Engineering Center.

Gordon, C. C. & Bradtmiller, B. (2012). Anthropometric change: Implications for office ergonomics. *Work*, *41*(Supplement 1), 4606–4611.

Gordon C. C., Churchill, T., Clauser, C. E., Bradtmiller, B., McConville, J. T., Tebbetts, I., & Walker, R. A. (1989). *1988 Anthropometric Survey of U.S. Army Personnel: Methods and summary statistics*. Natick, MA: United States Army Natick Research, Development and Engineering Center.

Gwinn, J. D., Barden, J., & Judd, C. M. (2015). Face recognition in the presence of angry expressions: A target-race effect rather than a cross-race effect. *Journal of Experimental Social Psychology*, *58*, 1–10.

Gwinn, O. S., & Brooks, K. R. (2013). Race-contingent face aftereffects: A result of perceived racial typicality, not categorization. *Journal of Vision*, *13*(10), 1–11. doi:10.1167/13.10.13

Gwinn, O. S., & Brooks, K. R. (2015). No role for lightness in the encoding of Black and White: Race-contingent face aftereffects depend on facial morphology, not facial luminance. *Visual Cognition*, *23*(5), 597–611.

Halberstadt, J., Sherman, S. J., & Sherman, J. W. (2011). Why Barack Obama is Black: A cognitive account of hypodescent. *Psychological Science*, *22*(1), 29–33.

Halder, I., Kip, K. E., Mulukutla, S. R., Alyer, A. N., Marroquin, O. C., Huggins, G. S., & Reis, S. E. (2012). Biographic ancestry, self-identified race, and admixture-phenotype associations in the Heart SCORE study. *American Journal of Epidemiology*, *176*(2), 146–155.

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.

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

Harris, P. (2008). *Designing and reporting experiments in psychology (3rd ed.)*. Maidenhead: Open University Press.

Harrison, S. A., Gauthier, I., Hayward, W. G., & Richler, J. J. (2014). Other-race effects manifest in overall performance, not qualitative processing style. *Visual Cognition*, *22*(6), 843–864.

Hayward, W. G., Crookes, K., Chu, M. H., & Favelle, S. K. (2016). Holistic processing of face configurations and components. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1482–1489.

Hayward, W. G., Crookes, K., & Rhodes, G. (2013). The other-race effect: Holistic coding differences and beyond. *Visual Cognition*, *21*(9–10), 1224–1247.

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.

Hedrick, P. (2008). East Asian and Melanesian ancestry in Polynesians. *The American Journal of Human Genetics*, *83*(1), 139–140.

Heron-Delaney, M., Anzures, G., Herbert, J. S., Quinn, P. C., Slater, A. M., Tanaka, J. W., . . . Pascalis, O. (2011). Perceptual training prevents the emergence of the other race effect during infancy. *PLoS ONE*, *6*(5), 1–5. doi:10.1371/journal.pone.0019858

Herzmann, G., Minor, G., & Curran, T. (2018). Neural evidence for the contribution of holistic processing but not attention allocation to the other-race effect on memory. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(5), 1015–1033.

Hesterberg, T., Monaghan, S., Moore, D. S., Clipson, A., & Epstein, R. (2003). *Bootstrap methods and permutation tests: Companion chapter to the practice of business statistics*. New York, NY: W. H. Freeman and Company.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: An eye-tracking study. *Visual Cognition*, *21*(7), 876–898.

Hills, P. J., & Lewis, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *The Quarterly Journal of Experimental Psychology*, *59*(6), 996–1002.

Hills, P. J., & Lewis, M. B. (2011). Reducing the own-race bias in face recognition by attentional shift using fixation crosses preceding the lower half of a face. *Visual Cognition*, *19*(3), 313–339.

Hills, P. J., & Lewis, M. B. (2018). The development of face expertise: Evidence for a qualitative change in processing. *Cognitive Development*, *48*, 1–18.

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition*, *129*(3), 586–597.

Hills, P. J., Ross, D. A., & Lewis, M. B. (2011). Attention misplaced: The role of diagnostic features in the face-inversion effect. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1396–1406.

Ho, A. K., Sidanius, J., Levin, D. T., & Banaji, M. J. (2011). Evidence for hypodescent and racial hierarchy in the categorization and perception of biracial individuals. *Journal of Personality and Social Psychology*, *100*(3), 492–506.

Ho, M. R., & Pezdek, K. (2016). Postencoding cognitive processes in the cross-race effect: Categorization and individuation during face recognition. *Psychonomic Bulletin & Review*, *23*(3), 771–780.

Holland, E. (2009). Limitations of traditional morphometrics in research on the attractiveness of faces. *Psychonomic Bulletin & Review*, *16*(3), 613–615.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hopper, W. J., Finklea, K. M., Winkielman, P., & Huber, D. E. (2014). Measuring sexual dimorphism with a race-gender face space. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1779–1788.

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.

Horry, R., Wright, D. B., & Tredoux, C. G. (2010). Recognition and context memory for faces from own and other ethnic groups: A remember-know investigation. *Memory & Cognition*, *38*(2), 134–141.

Howells, W. W. (1973). *Cranial variation in man: A study by multivariate analysis of patterns of difference among recent human populations*. *Papers of the Peabody Museum of Archaeology and Ethnology Volume 67.* Cambridge, MA: Harvard University.

Howells, W. W. (1989). *Skull shapes and the map: Craniometric analyses in the dispersion of Modern Homo. Papers of the Peabody Museum of Archaeology and Ethnology Volume 79*. Cambridge, MA: Harvard University.

Howells, W. W. (1995). *Who's who in skulls: Ethnic identification of crania from measurements. Papers of the Peabody Museum of Archaeology and Ethnology Volume 82*. Cambridge, MA: Harvard University.

Howells, W. W. (1996). Howells' craniometric data on the internet. *American Journal of Physical Anthropology*, *101*(3), 441–442.

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

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science*, *33*(6), 1173–1181.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334–340.

Hugenberg, K., Wilson, J. P., See, P. E., & Young, S. G. (2013). Towards a synthetic model of own group biases in face memory. *Visual Cognition*, *21*(9–10), 1392–1417.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other race recognition deficit. *Psychological Review*, *117*(4), 1168–1187.

Humes, K. R., Jones, N. A., & Ramirez, R. R. (2011). *Overview of race and Hispanic origin: 2010*. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2011/dec/c2010br-02.pdf

Innocence Project (n.d.). *DNA exonerations in the United States*. Retrieved from https://www.innocenceproject.org/dna-exonerations-in-the-united-states/

Innocence Project (2017). *The future*. Retrieved from https://25years.innocenceproject.org/future/

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, *39*(1), 57–106.

Jaquet, E., Rhodes, G., & Hayward, W. G. (2008). Race-contingent aftereffects suggest distinct perceptual norms for different race faces. *Visual Cognition*, *16*(6), 734–753.

Jerovich, A. (2017). *Implicit attitudes and the other race effect* (Unpublished master's thesis). Western Sydney University, Australia.

Jin, L., & Su, B. (2000). Natives or immigrants: Modern human origin in East Asia. *Nature Reviews Genetics*, *1*(2), 126–133.

Johnson, K. J., & Fredrickson, B. L. (2005). “We all look the same to me”: Positive emotions eliminate the own-race bias in face recognition. *Psychological Science*, *16*(11), 875–881.

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102*(1), 116–131.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36.

Kariks, J., Kooptzoff, O., & Walsh, R. J. (1957). Blood groups of the native inhabitants of Bougainville, New Guinea. *Oceania*, *28*(2), 146–158.

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, *104*(1), 105–114.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089.

Khan, R. (2008, May 21). *Fear of a white planet*. Retrieved from: https://blogs.discovermagazine.com/gnxp/2008/05/fear-of-a-white-planet/

Kibria, N. (1998). The contested meanings of 'Asian American': racial dilemmas in the contemporary US. *Ethnic and Racial Studies*, *21*(5), 939–958.

Kikutani, M. (2018). Influence of social anxiety on recognition memory for happy and angry faces: Comparison between own- and other-race faces.

The Quarterly Journal of Experimental Psychology, *71*(4), 870–878.

Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674.

Kinzler, K. D., & Shutts, K. (2008). Memory for "mean" over "nice": The influence of threat on children's face memory. *Cognition*, *107*(2), 775–783.

Kittoe, A. G. (2013). *On human biological diversity: Variation in sexual dimorphism of the skull between African-Americans and European-Americans* (Unpublished master's thesis). Louisiana State University, LA, United States.

Klimentidis, Y. C., Miller, G. F., & Shriver, M. D. (2009). Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *American Journal of Physical Anthropology*, *138*(4), 375–383.

Klimentidis, Y. C., & Shriver, M. D. (2009). Estimating genetic ancestry proportions from faces. *PLoS ONE*, *4*(2), 1–8. doi:10.1371/journal.pone.0004460

Kloth, N., Shields, S. E., & Rhodes, G. (2014). On the other side of the fence: Effects of social categorization and spatial grouping on memory and attention for own-race and other-race faces. *PLoS ONE*, *9*(9), 1–10. doi:10.1371/journal.pone.0105979

Knuycky, L. R., Kleider, H. M., & Cavrak, S. E. (2014). Line-up misidentifications: When being ‘prototypically Black’ is perceived as criminal. *Applied Cognitive Psychology*, *28*(1), 39–46.

Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(25), 9079–9084.

Krzywinski, M., & Altman, N. (2013). Points of significance: Error bars. *Nature Methods*, *10*(10), 921–922.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1–12. doi:10.3389/fpsyg.2013.00863

Lavrakas, P. J., Burl, J. R., & Mayzner, M. S. (1976). A perspective on the recognition of other-race faces. *Perception & Psychophysics*, *20*(6), 475–481.

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

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature*, *410*(6831), 890.

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PLoS ONE*, *4*(1), 1–7. doi:10.1371/journal.pone.0004215

Leder, H., Goller, J., Forster, M., Schlageter, L., & Paul, M. A. (2017). Face inversion increases attractiveness. *Acta Psychologica*, *178*, 25–31.

Levin, D.T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General*, *129*(4), 559–574.

Li, Y., & Tse, C.-S. (2016). Interference among the processing of facial emotion, face race, and face gender. *Frontiers in Psychology*, *7*, 1–14. doi:10.3389/fpsyg.2016.01700

Li, Y., Tse, C.-S., & Sun, J. (2018). Other-race categorisation advantage in a binary- vs. ternary-response race categorisation task. *Journal of Cognitive Psychology*, *30*(2), 242–256.

Lindsay, D. S., Jack, P. C., Jr., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology*, *76*(4), 587–589.

Lorenzino, M., Caminati, M., & Caudek, C. (2018). Tolerance to spatial-relational transformations in unfamiliar faces: A further challenge to a configural processing account of identity recognition. *Acta Psychologica*, *188*, 25–38.

Lucas, T., & Henneberg, M. (2016). Comparing the face to the body, which is better for identification? *International Journal of Legal Medicine*, *130*(2), 533–540.

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.

MacLin, O. H., Van Sickler, B. R., MacLin, K. M., & 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.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334.

Manica, A., Amos, W., Balloux, F., & Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. *Nature*, *448(7151)*, 346–348.

Marcon, J. L., Meissner, C. A., Frueh, M., Susa, K. J., & MacLin, O. H. (2010). Perceptual identification and the cross-race effect. *Visual Cognition*, *18*(5), 767–779.

Margalit, E., Biederman, I., Herald, S. B., Yue, X., & von der Malsburg, C. (2016). An applet for the Gabor similarity scaling of the differences between complex stimuli. *Attention, Perception, & Psychophysics*, *78*(8), 2298–2306.

Matheson, H. E., & McMullen, P. A. (2011). A computer-generated face database with ratings on realism, masculinity, race, and stereotypy. *Behavior Research Methods*, *43*(1), 224–228.

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

McDonnell, G. P., Bornstein, B. H., Laub, C. E., Mills, M., & Dodd, M. D. (2014). Perceptual processes in the cross-race effect: Evidence from eyetracking. *Basic and Applied Social Psychology*, *36*(6), 478–493.

McEvoy, B. P., Lind, J. M., Wang, E. T., Moyzis, R. K., Visscher, P. M., van Holst Pellekaan, S. M., & Wilton, A. N. (2010). Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. *The American Journal of Human Genetics*, *87*(2), 297–305.

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, 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), 1–6. doi:10.1371/journal.pone.0047956

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*(1), 5–22.

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

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35.

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*(1), 55–73.

Michel, C., Rossion, B., Bülthoff, I., Hayward, W. G., & Vuong, Q. C. (2013). The contribution of shape and surface information in the other-race face effect. *Visual Cognition*, *21*(9–10), 1202–1223.

Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one’s own race. *Psychological Science*, *17*(7), 608–615.

Miller, D. A., Smith, E. R., & Mackie, D. M. (2004). Effects of intergroup contact and political predispositions on prejudice: Role of intergroup emotions. *Group Processes & Intergroup Relations*, *7*(3), 221–237.

Mondloch, C. J., Elms, N., Maurer, D., Rhodes, G., Hayward, W. G., Tanaka, J. W., & Zhou, G. (2010). Processes underlying the cross-race effect: An investigation of holistic, featural, and relational processing of own-race versus other-race faces. *Perception*, *39*(8), 1065–1085.

Montalan, B., Veujoz, M., Boitout, A., Leleu, A., Camus, O., Lalonde, R., &

Rebaï, M. (2013). Investigation of effects of face rotation on race processing: An ERPs study. *Brain and Cognition*, *81*(3), 360–369.

Murphy, J., & Cook, R. (2017). Revealing the mechanisms of human face perception using dynamic apertures. *Cognition, 169,* 25–35.

Mydlová, M., Dupej, J., Koudelová, J., & Velemínská, J. (2015). Sexual dimorphism of facial appearance in ageing human adults: A cross-sectional study. *Forensic Science International*, *257*, 519.e1–519.e9.

Ng, W.-J., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology*, *25*(2), 217–232.

Nguyen, T. B., & Pezdek, K. (2017). Memory for disguised same- and cross-race faces: The eyes have it. *Visual Cognition*, *25*(7–8), 762–769.

Nystrom, K. C., & Malcom, C. M. (2010). Sex-specific phenotypic variability and social organization in the Chiribaya of southern Peru. *Latin American Antiquity*, *21*(4), 375–397.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*(3)*,* 396–402.

O'Toole, A., Deffenbacher, K., Abdi, H., & Bartlett, J. (1991). Simulating the "other-race effect" as a problem in perceptual learning. *Connection Science*, *3*(2), 163–178.

O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1996). An 'other-race effect' for categorizing faces by sex. *Perception*, *25*(6), 669–676.

Office for National Statistics (2011). *Household questionnaire: England*. Retrieved from https://census.ukdataservice.ac.uk/media/50966/2011_england_household.pdf

Office for National Statistics (n.d.). *Ethnic group, national identity and religion. Measuring equality: A guide for the collection and classification of ethnic group, national identity and religion data in the UK.* Retrieved from https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion

Oppenheimer, S. (2012). Out-of-Africa, the peopling of continents and islands: Tracing uniparental gene trees across the map. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 770–784.

Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition*, *116*(2), 283–288.

Park, J. Z. (2008). Second-generation Asian American pan-ethnic identity: Pluralized meanings of a racial label. *Sociological Perspectives*, *51*(3), 541–561.

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., Jr., & Ivcevic, Z. (2009). Not so Black and White: Memory for ambiguous group members. *Journal of Personality and Social Psychology, 96*(4), 795–810.

Petersen, H. C. (2000). On statistical methods for comparison of intrasample morphometric variability: Zalavár revisited. *American Journal of Physical Anthropology*, *113*(1), 79–84.

Phipps, G. S., German, R. Z., & Smith, R. J. (1988). Comparative craniofacial variation in Navajo Indians and North American Caucasians. *American Journal of Physical Anthropology*, *76*(2), 145–154.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811–832.

Pilucik, D. L., & Madsen, J. K. (2017). The effect of economic scarcity priming on perception of race. *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2902–2906). Austin, TX: Cognitive Science Society.

Porter, J. P. (2004). The average African American male face: An anthropometric analysis. *Archives of Facial Plastic Surgery*, *6*(2), 78–81.

Porter, J. P., & Olson, K. L. (2001). Anthropometric facial analysis of the African American woman. *Archives of Facial Plastic Surgery*, *3*(3), 191–197.

Prete, G., Fabri, M., Foschi, N., & Tommasi, L. (2016). Facial gender categorization and hemispheric asymmetries: Contrasting evidence from connected and disconnected brains. *Neuroscience*, *339*, 210–218.

Proietti, V., Laurence, S., Matthews, C. M., Zhou, X., & Mondloch, C. J. (2019). Attending to identity cues reduces the own-age but not the own-race recognition advantage. *Vision Research*, 157, 184–191.

R Core Team (2016). R: A Language and Environment for Statistical Computing (Version 3.3.0) [Computer software]. Retrieved from https://cran.r-project.org/bin/windows/base/old/3.3.0/

R Core Team (2019). R: A Language and Environment for Statistical Computing (Version 3.5.3) [Computer software]. Retrieved from https://www.R-project.org

Reh, W., & Scheffler, B. (1996). Significance tests and confidence intervals for coefficients of variation. *Computational Statistics & Data Analysis*, *22*(4), 449–452.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489–494.

Relethford, J. H. (2000). Human skin color diversity is highest in Sub-Saharan African populations. *Human Biology, 72*(5), 773–780.

Relethford, J. H. (2001). Global analysis of regional differences in craniometric diversity and population substructure. *Human Biology*, *73*(5), 629–636.

Relethford, J. H. (2009). Race and global patters of phenotypic variation. *American Journal of Physical Anthropology*, *139*(1), 16–22.

Relethford, J. H., & Blangero, J. (1990). Detection of differential gene flow from patterns of quantitative variation. *Human Biology*, *62*(1), 5–25.

Reyes-Centeno, H., Ghirotto, S., Détroit, F., Grimaud-Hervé, D., Barbujani, G., & Harvati, K. (2014). Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proceedings of the National Academy of Sciences* of the United States *of America*, *111*(20), 7248–7253.

Rhee, S. C. (2018). Differences between Caucasian and Asian attractive faces. *Skin Research & Technology*, *24*(1), 73–79.

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80*(3), 313–331.

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.

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review, 13*(3), 499–505.

Rhodes, G., Locke, V., Ewing, L., & Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception*, *38*(2), 232–241.

Rhodes, G., Watson, T. L., Jeffery, L., & Clifford, C. W. G. (2010). Perceptual adaptation helps us identify faces. *Vision Research*, *50*(10), 963–968.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*(3), 287-290.

Richeson, J. A., & Shelton, J. N. (2007). Negotiating interracial interactions. *Current Directions in Psychological Science*, *16*(6), 316-320.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011a). Holistic processing predicts face recognition. *Psychological Science*, *22*(4), 464–471.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011b). Beliefs alter holistic face processing ... if response bias is not taken into account. *Journal of Vision*, *11*(13), 1–13. doi:10.1167/11.13.17

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*(5), 1281–1302.

Risch, N., Burchard, E., Ziv, E., & Tang, H. (2002). Categorization of humans in biomedical research: Genes, race and disease. *Genome Biology*, *3*(7), 1–12. doi:10.1186/gb-2002-3-7-comment2007

Roelfsema, N. M., Hop, W. C. J., van Adrichem, L. N. A., & Wladimiroff, J. W. (2007). Craniofacial variability index determined by three-dimensional ultrasound in isolated vs. syndromal fetal cleft lip/palate. *Ultrasound in Obstetrics & Gynecology*, *29*(3), 265–270.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., Feldman, M. W. (2002). Genetic structure of human populations. *Science*, *298*(5602), 2381–2385.

Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision*, *8*(3), 1–13. doi:10.1167/8.4.3

Rossion, B., & Michel, C. (2011). An experienced-based holistic account of the other-race face effect. In: A. J. Calder, G. Rhodes, M. H. Johnson, & J.V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 215–244). Oxford, UK: Oxford University Press.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, *1*(1), 43–46.

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

Samal, A., Subramani, V., & Marx, D. (2007). Analysis of sexual dimorphism in human face. *Journal of Visual Communication and Image Representation*, *18*(6), 453–463.

Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*(6), 440–444.

Saperstein, A., & Penner, A. M. (2012). Racial fluidity and inequality in the United States. *American Journal of Sociology*, *118*(3), 676–727.

Sassenberg, K., Moskowitz, G. B., Jacoby, J., & Hansen, N. (2007). The carry-over effect of competition: The impact of competition on prejudice towards uninvolved outgroups. *Journal of Experimental Social Psychology*, *43*(4), 529–538.

Scherer, A. K., & Wright, L. E. (2015). Dental morphometric and strontium isotope evidence for population history at Tikal, Guatemala. In A. Cucina (Ed.), *Archaeology and bioarchaeology of population movement among Prehispanic Maya* (pp. 109–118). New York, NY: Springer International Publishing.

Scott, L. S., & Monesson, A. (2009). The origin biases in face perception. *Psychological Science*, *20*(6), 676–680.

Segovia, K. Y., Bailenson, J. N., & Leonetti, C. (2012). Virtual human identification line-ups. In C. Wilkinson & C. Rynn (Eds.), *Craniofacial Identification* (pp. 101–114). New York, NY: Cambridge University Press.

SenGupta, A. (1987). Tests for standardized generalized variances of multivariate normal populations of possibly different dimensions. *Journal of Multivariate Analysis*, *23*(2), 209–219.

Shah, P., Bird, G., & Cook, R. (2015). Face processing in autism: Reduced integration of cross-feature dynamics. *Cortex*, *75*, 113–119.

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): A self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, *2*(6), 1–11. doi:10.1098/rsos.140343

Shaw, R. B., Jr., Katzel, E. B., Koltz, P. F., Yaremchuk, M. J., Girotto, J. A., Kahn, D. M., & Langstein, H. N. (2011). Aging of the facial skeleton: Aesthetic implications and rejuvenation strategies. *Plastic and Reconstructive Surgery*, *127*(1), 374–383.

Sheehan, M. J., & Nachman, M. W. (2014). Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nature Communications*, *5*, 1–10. doi:10.1038/ncomms5800.

Shriner, D., Tekola-Ayele, F., Adeyemo, A., & Rotimi, C. N. (2014). Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific Reports*, *4*, 1–9. doi:10.1038/srep06055

Shriver, E. R., & Hugenberg, K. (2010). Power, individuation, and the cross-race recognition deficit. *Journal of Experimental Social Psychology*, *46*(5), 767–774.

Shriver, M. D., Parra, E. J., Dios, S., Bonilla, C., Norton, H., Jovel, C., . . . Kittles, R. A. (2003). Skin pigmentation, biogeographical ancestry and admixture mapping. *Human Genetics*, *112*(4), 387–399.

Slice, D. E. (2007). Geometric morphometrics. *Annual Review of Anthropology*, *36*, 261–281.

Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic and Applied Social Psychology*, *22*(2), 71–84.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups. *Psychology, Public Policy, and Law*, *7*(1), 36–97.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stefan, V. H. (1999). Craniometric variation and homogeneity in prehistoric/protohistoric Rapa Nui (Easter Island) regional populations. *American Journal of Physical Anthropology*, *110*(4), 407–419.

Stepanova, E. V., & Strube, M. J. (2012). The role of skin color and facial physiognomy in racial categorization: Moderation by implicit racial attitudes. *Journal of Experimental Social Psychology*, *48*(4), 867–878.

Steyn, M., & İşcan, M. Y. (1998). Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Science International*, *98*(1–2), 9–16.

Stiff, C., & Bowen, T. (2016). Two-player game: Playing casual video games with outgroup members reduces levels of prejudice toward that outgroup. *International Journal of Human–Computer Interaction*, *32*(12), 912–920.

Strom, M. A., Zebrowitz, L. A., Zhang, S., Bronstad, P. M., & Lee, H. K. (2012). Skin and bones: The contribution of skin tone and facial structure to racial prototypicality ratings. *PLoS ONE*, *7*(7), 1–8. doi:10.1371/journal.pone.0041193

Suhrke, J., Freitag, C., Lamm, B., Teiser, J., Fassbender, I., Poloczek, S., . . . Schwarzer, G. (2014). The other-race effect in 3-year-old German and Cameroonian children. *Frontiers in Psychology*, *5*, 1–7. doi:10.3389/fpsyg.2014.00198

Susilo, T., Rezlescu, C., & Duchaine, B. (2013). The composite effect for inverted faces is reliable at large sample sizes and requires basic face configuration. *Journal of Vision*, *13*(13), 1–9. doi:10.1167/13.13.14

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

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.

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 122–131.

Tanikawa, C., Zere, E., & Takada, K. (2016). Sexual dimorphism in the facial morphology of adult humans: A three-dimensional analysis. *HOMO - Journal of Comparative Human Biology*, *67*(1), 23–49.

Thomas, E. L., Dovidio, J. F., & West, T. V. (2014). Lost in the categorical shuffle: Evidence for the social non-prototypicality of Black women. *Cultural Diversity and Ethnic Minority Psychology*, *20*(3), 370–376.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., … Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science*, *324*(5930), 1035–1044.

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied*, *8*(3), 180–193.

United States Census Bureau (2017, January 12). *About*. Retrieved from https://www.census.gov/topics/population/race/about.html

United States Census Bureau (n.d.). *Quick facts: United States*. Retrieved from https://www.census.gov/quickfacts/fact/table/US/PST045217

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

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

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

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology*, *44A*(4), 671–703.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology, 69*(10), 1996–2019.

von Cramon-Taubadel, N., & Lycett, S. J. (2008). Brief communication: Human cranial variation fits iterative founder effect model with African origin. *American Journal of Physical Anthropology*, *136*(1), 108–113.

Waldman, C. (2006). *Encyclopedia of Native American tribes (3rd ed.)*. New York, NY: Infobase Publishing.

Walker, D. A. (2003). JMASM9: Converting Kendall's tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, *2*(2), 525–530.

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.

Walker, P. M., & Hewstone, M. (2008). The influence of social factors and implicit racial bias on a generalized own-race effect. *Applied Cognitive Psychology*, *22*(4), 441–453.

Wallace, C. (2016, January 28). *Daughters of interracial couples are more likely* *to say they are multiracial*. Retrieved from: http://www.time.com/4195141/daughters-of-interracial-couples-are-more-likely-to-say-they-are-multiracial/

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.

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition*, *144*, 91–115.

Wan, L., McKone, E., Irons, J. L., & Crookes, K. (2013). Reducing the other-race effect requires childhood visual experience, not increased social motivation. *Journal of Vision*, *13*(9), 862. doi:10.1167/13.9.862

Wan, L., Pidcock, M., Crookes, K., Fiorentini, C., Reynolds, K., & McKone, E. (n.d.). *Greater plasticity of face recognition in childhood than in adulthood*. Unpublished manuscript. Retrieved from https://openresearch-repository.anu.edu.au/handle/1885/112506

Wang, S., Lewis, C. M., Jr., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., . . . Ruiz-Linares, A. (2007). Genetic variation and population structure in Native Americans. *PLoS Genetics*, *3*(11), 2049–2067. doi:10.1371/journal.pgen.0030185

Wang, Y., & Zhou, L. (2016). Long-term exposure to American and European movies and television series facilitates Caucasian face perception in young Chinese watchers. *Perception*, *45*(10), 1151–1165.

Wang, Z., Quinn, P. C., Tanaka, J. W., Yu, X., Sun, Y. H. P., Liu, J., . . . Lee, K. (2015). An other-race effect for configural and featural processing of faces: Upper and lower face regions play different roles. *Frontiers in Psychology*, *6*, 1–8. doi:10.3389/fpsyg.2015.00559

Ward, R. E., Jamison, P. L., & Farkas, L. G. (1998). Craniofacial variability index: A simple measure of normal and abnormal variation in the head and face. *American Journal of Medical Genetics*, *80*(3), 232–240.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*(6982), 557–561.

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1571), 1702–1725.

Weisensee, K. E. (2001). *Pecos revisited: A modern analysis of Earnest Hooton's, the Indians of Pecos Pueblo* (Unpublished master's thesis). University of Tennessee, Knoxville, TN.

Wells, G. L., & Olson, E. A. (2001). The other-race effect in eyewitness identification: What do we do about it? *Psychology, Public Policy, and Law*, *7*(1), 230–246.

Willenbockel, V., Fiset, D., & Tanaka, J. W. (2011). Relative influences of lightness and facial morphology on perceived race. *Perception*, *40*(5), 621–624.

Williams, E. (2011). *Aviation formulary V1.46*. Retrieved from http://edwilliams.org/avform.htm

Wilson, J. P. (2010). *Motivation to respond without prejudice and the cross race effect* (Unpublished master's thesis). Miami University, OH.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1–3), 37–52.

Yaeger, R., Avila-Bront, A., Abdul, K., Nolan, P. C., Grann, V. R., Birchette, M. G., . . . Joe A. K. (2008). Comparing genetic ancestry and self-described race in African Americans born in the United States and in Africa. *Cancer Epidemiology, Biomarkers & Prevention*, *17*(6), 1329–1338.

Yamaguchi, M. K., Hirukawa, T., & Kanazawa, S. (1995). Judgment of gender through facial parts. *Perception*, *24*(5), 563–575.

Yao, W. (2014). *Spatially dense 3D morphometric study of human craniofacial sexual dimorphism with respect to ancestry* (Unpublished master's thesis). The Pennsylvania State University, PA, United States.

Yarrow, K. (2018). Collecting and interpreting judgements about perceived similarity: A model-fitting tutorial. In A. Vatakis, F. Balcı, M. Di Luca, & A. Correa (Eds.), *Timing and Time Perception: Procedures, Methods, and* Applications (pp. 295–325). Leiden, Netherlands: Brill.

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

Young, N. M., Sherathiya, K., Gutierrez, L., Nguyen, E., Bekmezian, S., Huang, J. C., . . . Marcucio, R. S. (2016). Facial surface morphology predicts variation in internal skeletal shape. *American Journal of Orthodontics and Dentofacial Orthopedics*, *149*(4), 501–508.

Young, S. G., & Hugenberg, K. (2012). Individuation motivation and face experience can operate jointly to produce the own-race bias. *Social Psychological and Personality Science*, *3*(1), 80–87.

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

Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review*, *15*(6), 1093–1099.

Zhao, L., & Bentin, S. (2011). The role of features and configural processing in face-race classification. *Vision Research*, *51*(23–24), 2462–2470.

Zhao, M., & Hayward, W. G. (2010). Holistic processing underlies gender judgments of faces. *Attention, Perception, & Psychophysics, 72*(3), 591–596.

Zhao, M., & Hayward, W. G. (2013). Integrative processing of invariant aspects of faces: Effect of gender and race processing on identity analysis. *Journal of Vision, 13*(1), 1–18. doi:10.1167/13.1.15

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014a). Face format at encoding affects the other-race effect in face memory. *Journal of Vision*, *14*(9), 1–13. doi:10.1167/14.9.6

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014b). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research*, *105*, 61–69.

Zhuang, Z., & Bradtmiller, B. (2005). Head-and-face anthropometric survey of U.S. respiratory users. *Journal of Occupational and Environmental Hygiene*, *2*(11), 567–576.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21–29.

APPENDICES

Appendix A: Facial variabilities and interdimensional correlations 252
Appendix B: Cranial sexual size dimorphism and the distances migrated by anatomically modern humans 254
Appendix C: Howells cranial data sample sizes, and statistics used in the analysis in Appendix B 295

Appendix A

Facial Skeletal Diversities and Mean Interdimensional Correlations

	Males				Females			
Population	MV	MPV	MIC	SGV	MV	MPV	MIC	SGV
Anyang	0.64	0.72	0.21	0.83				
Greenlandic Inuit	0.66	0.75	0.19	0.84	0.65	0.73	0.18	0.84
Buryat	0.78	0.77	0.23	0.87	0.67	0.72	0.22	0.86
Ainu	0.72	0.72	0.27	0.85	0.67	0.72	0.23	0.85
Andaman	0.56	0.66	0.23	0.81	0.57	0.66	0.23	0.81
San	0.82	0.74	0.31	0.85	0.75	0.75	0.25	0.84
Egypt	0.65	0.73	0.19	0.85	0.63	0.71	0.20	0.84
Guam	0.60	0.66	0.26	0.82	0.53	0.65	0.24	0.82
Philippines	0.65	0.71	0.23	0.85				
Atayal	0.61	0.65	0.29	0.82	0.66	0.68	0.28	0.82
Hainan Island	0.63	0.72	0.20	0.85	0.70	0.76	0.22	0.88
South Japan	0.64	0.69	0.24	0.83	0.62	0.68	0.25	0.84
North Japan	0.72	0.75	0.22	0.87	0.76	0.75	0.25	0.87
Peru	0.55	0.64	0.24	0.81	0.58	0.67	0.23	0.84
Santa Cruz	0.58	0.66	0.26	0.81	0.50	0.63	0.22	0.80
Arikara	0.62	0.68	0.25	0.82	0.65	0.71	0.24	0.85
Moriori	0.63	0.70	0.22	0.83	0.60	0.69	0.23	0.83
Easter Island	0.61	0.68	0.24	0.83	0.65	0.69	0.24	0.83
Mokapu	0.60	068	0.25	0.81	0.59	0.67	0.23	0.82
Tolai	0.62	0.72	0.20	0.83	0.65	0.68	0.28	0.83
Tasmania	0.75	0.78	0.23	0.87	0.74	0.75	0.23	0.86
Australia	0.62	0.69	0.23	0.83	0.66	0.71	0.23	0.83
Zulu	0.71	0.74	0.22	0.86	0.75	0.77	0.22	0.89
Dogon	0.72	0.76	0.25	0.85	0.66	0.73	0.22	0.85
Teita	0.73	0.79	0.21	0.84	0.72	0.77	0.19	0.88
Berg	0.71	0.73	0.26	0.85	0.69	0.74	0.20	0.86
Zalavár	0.61	0.71	0.20	0.84	0.69	0.72	0.24	0.86
Norse	0.64	0.73	0.19	0.85	0.65	0.73	0.19	0.87

Note. MV = mean variance; MIC = mean interdimensional correlation; MPV = mean pattern variability; SGV = standardised generalised variance.

Appendix B

CRANIAL SEXUAL SIZE DIMORPHISM AND THE DISTANCES MIGRATED BY ANATOMICALLY MODERN HUMANS

Abstract

Regarding *Homo sapiens*, studies have found cranial dimensions to often be sexually dimorphic, and that a serial founder effect occurred with migration from the African geographic origin of anatomically modern humans. Extrapolations from the results of some prior studies might indicate that cranial size dimorphism increased as migratory distance grew. The current research established whether cranial size dimorphism relates to migratory distance. Additionally, it attempted to make inroads regarding why an association possibly happens by finding whether crania enlarge as distance furthers and (given Rensch's rule) if cranial dimorphism increases with cranial size. To represent cranial size, geometric means were calculated from 26 Holocene modern human populations from the Howells dataset. Analyses adjusted cranial dimorphism and size for absolute latitude. Holm-Bonferroni corrections were employed on *p*-values. Dimorphism positively correlated with distance from Africa, and results suggested a stronger relationship between dimorphism and distance when using African, rather than non-African, origins. However, size had no association with distance from Africa, and dimorphism was not indicated to vary with size. The relationship between cranial size dimorphism and distance may be a new pointer towards the location from whence modern humans originated. As for a potential reason for this association, the speculated route of size getting larger with distance, and dimorphism increasing with size, seems unlikely. Possible reasons

are presented in the context of the serial founder effect and archaic human ancestry. It is considered whether migrating larger distances lessened a temporal decrease in cranial size dimorphism.

Introduction

Typically, measurements of the modern human cranium are sexually dimorphic (i.e., different for males than for females) in the direction of males having a larger size than females (e.g., Maina, Mahdi, & Kalayi, 2011; Steyn & İşcan, 1998). Research has occasionally concerned whether the sexual dimorphism of skull measurements is of the same magnitude across different groups (e.g., Spradley & Jantz, 2011). Referring to previous studies on the skull, sex, and race (presumably up to the late 1800s), Stepan (1986) noted that “[o]ne novel conclusion to result from scientists' investigations … was that the gap in head size between men and women had apparently widened over historic time, being largest in the "civilized" races such as the European” (p. 270). More recent research has found sexual dimorphism in cranial capacity to be equivalent across three races (measurements of Africans, East Asians, and [European] Caucasians were tested) (Rushton, 1994) and cranial dimorphism to be the same for Blacks and Whites (Spradley & Jantz, 2011).[44] Although, regarding cranial capacity, a sex × population interaction was present in Rushton (1992), a population category of Asian and Pacific persons (Gordon et al., 1989) was not a monoracial category (see Risch, Burchard, Ziv, & Tang, 2002) (the other populations used in

[44] When previous studies (e.g., Rushton, 1992; Spradley & Jantz, 2011) referred to in the current research have used the terms *Black* or *White*, bearing in mind descriptions regarding populations in those studies (locations/sources) and the use of race terms (e.g., Bhopal, 2004; Risch et al., 2002; United States Census Bureau, 2017), it is presumed that Black and White represent African and Caucasian respectively.

Rushton, 1992, were Black, and White). The current study considered if the distance to which modern human populations migrated from their geographic start in Africa (Prugnolle, Manica, & Balloux, 2005) relates to cranial sexual size dimorphism (when controlling dimorphism for absolute latitude).

Migration

Congruent with there being a serial founder effect (Ramachandran et al., 2005), as anatomically modern humans migrated from their onset in Africa (e.g., Manica, Amos, Balloux, & Hanihara, 2007), there was a decline in genetic diversity (Balloux, Lawson Handley, Jombart, Liu, & Manica, 2009; Li et al., 2008; Prugnolle et al., 2005; Ramachandran et al., 2005) which is exhibited in the morphological variability of some aspects of the skeleton (e.g., Betti, von Cramon-Taubadel, Manica, & Lycett, 2013; Betti, von Cramon-Taubadel, & Lycett, 2012; Hanihara, 2008) yet not in others (Betti et al., 2012). Regarding morphology, the fall in diversity has been evident in the cranium generally (Betti, Balloux, Amos, Hanihara, & Manica, 2009; Manica et al., 2007; von Cramon-Taubadel & Lycett, 2008) and within a number of individual cranial dimensions (Manica et al., 2007).

Indeed, Manica et al. (2007), regarding the link between migratory distance and cranial variability, sought to determine which locale for an origin led to the most potent distance-variability relationship (controlling for variability being explained by climate). The area was on the African continent, hence the origin of modern humans was indicated to be situated in Africa (Manica et al., 2007). Additionally, utilising the same male crania from 105 populations as Manica et al. (2007), results in Betti et al. (2009) suggested that parts of Africa gave the greatest relationship between distance and cranial variability. When using female crania, however, they found that areas of not only Africa but also some of Asia (yet still relatively near to Africa) were indicated, however, a smaller number of populations (39) was used in comparison to the 105 (Betti et al., 2009).

The outcomes of previous research may hint at cranial sexual dimorphism having become higher as migratory distance from the origin increased; if one were to average data from the fifth and sixth tables of L'Abbé, Kenyhercz, Stull, Keough, and Nawrocki (2013) (which presented the sexual dimorphism of Blacks and Whites in millimetres for 24 cranial dimensions), data indicates that cranial sexual dimorphism is numerically (although not necessarily significantly) lower for Blacks than Whites in South Africa ($M_{\text{Black}} = 2.91$, $M_{\text{White}} = 3.17$) and North America ($M_{\text{Black}} = 3.84$, $M_{\text{White}} = 4.56$). Furthermore, outcomes in Kimmerle, Ross, and Slice (2008) (presented in their first figure) would suggest a larger (numerical) magnitude of cranial sexual dimorphism amongst Whites than Blacks. A calculation of dimorphism from figures in Rushton (1992) favours cranial capacity dimorphism (in cubic centimetres) existing in the direction of being lower for Blacks ($M = 193.00$) than Whites ($M = 207.00$). Nevertheless, when using summary statistics presented in Rushton (1994), the direction was for dimorphism in cranial capacity to be higher for Africans ($M = 256.00$) than Caucasians ($M = 222.56$) and also East Asians ($M = 190.00$). Considering migratory distances (e.g., Manica et al., 2007), except for in regard to the data in Rushton (1994), these numerical differences would seem congruent with the idea of sexual dimorphism becoming higher as migratory distance lengthens; a significant difference in cranial sexual dimorphism between populations could appear when migratory distances differ beyond a certain extent. Therefore, the similar dimorphism of (certain) race groups in previous research (Rushton, 1994;

Spradley & Jantz, 2011) might not be surprising given the migratory distances (e.g., Betti et al., 2009) of those groups, which would indicate that Africans have small distances, whilst those of Caucasians and East Asians are medium. Still, whilst a previous study had examined if the distance between populations relates to sexual dimorphism in the pelvis (Betti, 2014), no research had tested whether migratory distance from the African origin has an association with dimorphism.

Cranial size and climate

If sexual dimorphism increased with migratory distance, this increase could have been due to cranial size changes for males or/and females. Rensch's rule (Rensch, 1959) signifies sexual dimorphism as elevating with body size when the larger sex is male rather than female, yet, when females are bigger than males, dimorphism lessening with increased size (see Abouheif & Fairbairn, 1997). This pattern can be considered to be a product of sexual selection (e.g., Dale et al., 2007; Székely, Freckleton, & Reynolds, 2004). Rensch (1959) stated that it "applies only to subspecies of a species, to related species of a genus, or to related genera of a family" (p. 159). The rule has been studied, for instance, with respect to the skull length of bat species (Stevens & Platt, 2015), although, regarding dimorphism of body size, "[p]rimarily, studies measure body size with univariate proxies such as skull length or directly" (Schutz, Polly, Krieger, & Guralnick, 2009, p. 339).

Interestingly, Gustafson and Lindenfors (2004) note that "Rensch claimed that the rule also should apply to "subspecies of a species" (Rensch, 1959, p. 159), thus implying that it ought to be possible to also trace effects of Rensch's rule in comparisons between populations" (p. 254). They tested Rensch's rule regarding human populations in terms of stature (Gustafsson & Lindenfors, 2004); studies had not occurred on Rensch's rule concerning the cranial size of *Homo sapiens*. This is not to imply that modern human population groups (e.g., of the Howells, 1996, dataset) are different subspecies from one another, just that the rule could extend to crania within the modern human species even without there being subspecies partitions.

Research had not explored whether cranial size became bigger with migratory distance. Nevertheless, Howells (1989) did make interesting comments regarding the skull sizes of the populations featured in the Howells (1996) cranial dataset. Howells (1989) described San as having "very small" skulls (p. 13); bearing in mind an African origin for modern humans (e.g., Manica et al., 2007), of the 28 main populations in the Howells dataset, San would have relatively short migratory distances. Still, Howells (1989) also referred to Andamanese as being "very small-skulled" (p. 14), whilst, given potential migratory routes (Oppenheimer, 2012; Reyes-Centeno et al., 2014), their migratory distances would be medium. Howells (1989) noted Polynesians as being "large-skulled" (p. 15); possible paths of migration (Jin & Su, 2000; Tassi et al., 2015) would suggest that Polynesians are of long migratory distances, although such a description from Howells was not limited to persons of a lengthy migratory distance, as Howells (1989) described Buryats as having "large skulls" (p. 15), yet in prior research (Betti et al., 2009) the Buryat migratory distance was middling.

As Caucasians and East Asians would have bigger migratory distances than Africans (e.g., looking at migratory distances presented in Manica et al., 2007), results in previous research which were consistent with the cranial capacity of Blacks being below that of Whites (Rushton, 1992, 1994) and East Asians

(Rushton, 1994), may seem to hint at cranial size having increased with migratory distance from the origin. However, in regard to cranial capacities presented within the second table of Beals, Smith, and Dodd (1984), although means for the locations of Africa and Europe would appear supportive of capacity having risen with distance, considering those summary statistics alongside the means for other presented locations (Asia, North America, Oceania, and South America) would not seem to give an impression of cranial capacity being larger as distance accrues. Indeed, traits of the cranium can exhibit adaptivity to environments, such that a cooler climate may promote size to minimise heat loss (Hubbe, Hanihara, & Harvati, 2009), with cranial capacity being linked to climate such that cranial capacity is larger in cooler, than in warmer, climates (Beals et al., 1984); the fourth figure of Beals et al. (1984) indicates the relative warmness of Africa.

Given migratory distances (e.g., Prugnolle et al., 2005), if cranial size became bigger with distance, non-Africans might be expected to have a larger cranial size compared to Africans, climate aside. In conjunction with findings from Relethford and Smith (2018), some supplementary calculations using the same data which Relethford and Smith employed perhaps suggests that non-Africans have a larger cranium than Africans, even when allowing for climate. Relethford and Smith (2018) used the means of cranial dimensions of Neanderthals presented in Weaver, Roseman, and Stringer (2007) and the raw cranial data of 30 populations of modern humans from the Howells (1996) dataset. The modern human data which Relethford and Smith made use of included the main 28 populations featured in the Howells data, which can be considered to be of modern humans from the Holocene based on descriptions from Howells (1989, 1995), and two small-sample-sized (10 crania each) Māori populations. Considering the history of migration into Polynesia and New Zealand in particular (e.g., Wilmshurst, Anderson, Higham, & Worthy, 2008; Wilmshurst, Hunt, Lipo, & Anderson, 2011), the Māori data would also be of Holocene modern humans.

Having derived geometric means from 37 shared dimensions regarding the Neanderthals and modern human data, Relethford and Smith (2018) found non-Africans to be more similar than Africans to Neanderthals in terms of geometric means, i.e., cranial size. This outcome also occurred when the analysis of non-Africans was limited to Oceanic populations, who would be populations from areas where the climate is relatively warm (Relethford & Smith, 2018). The research of Relethford and Smith was in the context of the idea that the Neanderthal ancestry within modern humans (via non-African ancestry) would be reflected in Neanderthal crania having a higher similarity to non-Africans than to Africans (Relethford & Smith, 2018). Whilst prior studies are indicative of Neanderthals having held a larger cranial size (Howells, 1989) or capacity than modern humans (e.g., Wood & Collard, 1999, who used descriptive statistics from Kappelman, 1996, who, in turn, described the sampled humans as having being alive in the 20th century, i.e., being of the Holocene period), research suggests that cranial capacity lowered during the Holocene (Henneberg, 1998; Henneberg & Steyn, 1993) (although see Jantz & Jantz, 2016, for evidence of a resurgence between the late 1800s and early 1900s, in terms of decade of birth, amongst their sampled group who were White Americans).

Nonetheless, also utilising the cranial data and dimensions which Relethford and Smith (2018) used, the Neanderthal cranial geometric mean was found to be above the 99% confidence interval of each of the Howells populations (with separate calculations occurring for males and females of the Howells data), which would indicate more sizeable crania for Neanderthals over the Holocene modern humans represented in the Howells data. Given Relethford and Smith (2018), and even with a climatic influence (Beals et al., 1984), this could point towards Africans having a smaller cranial size than non-Africans.

Absolute latitude has been used as a stand-in for temperature (Gustafsson & Lindenfors, 2009), and, indeed, cranial capacity increases with absolute latitude (Beals et al., 1984) as does cranial module size (Short, 2016), and previous research indicates that this pattern applies to the facial skeleton (Newman, 1953). This suggests that any exploration of whether migratory distance relates to cranial size would be more focussed if one were to control for climate. Regarding stature, sexual dimorphism increases with absolute latitude when common ancestry is not accounted for, yet not when such commonalities are taken into bearing (Gustafsson & Lindenfors, 2009); although research had not analysed whether cranial size dimorphism has an association with absolute latitude, it would seem prudent to adjust cranial size dimorphism for climate in case there is an association, particularly if cranial size affects cranial dimorphism.

To summarise, cranial sexual size dimorphism could be expected to have increased as migratory distance from an African commencement grew if there was an expansion in cranial size over migratory distances, and cranial dimorphism increased with the size of the cranium. With absolute latitude controlled for in each analysis to various extents, the current study sought to discover if cranial size dimorphism correlates with migratory distance from a location within Africa, and took some steps to begin to find why the distance-dimorphism relationship may be apparent by testing whether the cranial size of males and females correlates (positively) with migratory distance and if dimorphism rises with cranial size. Additionally, it sought an indication of whether the possible relationship between distance and dimorphism is stronger when distance is measured from within, compared to outside of, Africa.

Method

The present analysis used cranial measurements from 26 populations of Holocene *Homo sapiens* (Howells, 1989, 1995) of the Howells dataset (Howells, 1996: http://web.utk.edu/~auerbach/HOWL.htm). The geometric mean of cranial measurements has been used to denote cranial size in previous research (e.g., Brewster, Meiklejohn, von Cramon-Taubadel, & Pinhasi, 2014), and it was used for this purpose in the current study. For males and females, geometric means were calculated for individuals in the main populations from the Howells data which feature both males and females. Fifty-six linear measurement dimensions (listed in Table A.1) were made use of; glabella projection (Howells, 1989, 1996) was eliminated due to instances of zeros in its raw measurements, as inclusion of that dimension would have resulted in the geometric means of some individuals being zero. Two mean geometric means were determined for each of the 26 populations, with one being calculated from the male crania, and the other from the crania of females.

Sexual dimorphism was calculated as the natural log of the male/female ratio (see Smith, 1999), i.e., ln[(male geometric mean)/(female geometric mean)]. The “size index 1” of Schutz et al. (2009, p. 342), which was used in Betti (2014), was considered as another measure of cranial dimorphism. However, its derivation is the addition of the male variance and female variance as the denominator, with the squared difference between male and female means being the numerator (Schutz et al., 2009). Therefore, given the fall in cranial variance with furthering migratory distance (e.g., Betti et al., 2009), finding a distance-dimorphism association when using this dimorphism measure could have arguably been attributable to a relationship between distance and variance.

Approximated (historic) migratory distances (Section 3.1.2.1) were utilised, i.e., distances had been calculated using a manner from Williams (2011) as great circle distances between latitude/longitude estimates of a southern African start of humanity (Botswana) and population locations (including intercontinental crossings) (coordinates were found in von Cramon-Taubadel & Lycett, 2008) given probable migratory routes. Estimated latitudes of the Howells populations (von Cramon-Taubadel & Lycett, 2008) were obtained and used as absolute values. Absolute latitude has been employed as a representative of temperature (Gustafsson & Lindenfors, 2009) and was utilised as such in the present analysis.

Tests have been conducted regarding if the variability of individual cranial dimensions declined with lengthening migratory distance (Manica et al., 2007). As for the current research, the large number of dimensions relative to the amount of populations effectively eliminated any thorough exploration concerning which specific dimensions increased/decreased in sexual dimorphism (controlling for latitude) as migratory distance broadened, because adjustments to *p*-values to control the family-wise error rate (e.g., Holm-Bonferroni, Holm, 1979) would have resulted in very conservative thresholds for assigning significance. Nonetheless, for completeness, correlation coefficients (semi-partial) regarding individual dimensions were calculated.

Across previous research, different methods have been used when calculating migratory distance (e.g., Prugnolle et al., 2005; von Cramon-Taubadel & Lycett, 2008). For instance, Betti et al. (2009) used a manner which permits one "to model movement over land while avoiding major mountain ranges (more than 2000 m altitude)" (p. 810). Compared to such a method, the estimates employed in the present study could arguably be poorer approximations of actual migratory distances as the estimates did not take altitude into account. Nevertheless, those estimates negatively correlate with types of cranial variability, including (mean) variance, when variabilities are calculated via 32 cranial dimensions of the Howells (1996) data (male crania from 28 populations, and female crania from 26) (Section 3.1.3.1) which appears to support real migratory distances having being satisfactorily approximated.

Still, given the coupling between estimated migratory distance and cranial diversity (e.g., Manica et al., 2007), an occurrence of dimorphism and size relating to cranial diversity and not migratory distance could indicate that the estimates of migratory distances were not sufficiently reflective of reality. From the Howells data, mean variances were calculated following von Cramon-Taubadel and Lycett (2008), i.e., the cranial measurements for a person were divided by the geometric mean of their cranium therefore becoming shape measures (see also Jungers, Falsetti, & Wall, 1995; Betti et al., 2012), with mean variances (of *z*-score-standardised measurements) then generated in RMET 5.0 (Relethford & Blangero, 1990). Regarding hominoids, as previous research suggested a greater correspondence between genetics and cranial morphology amongst males than females (Zichello, Baab, McNulty, Raxworthy, & Steiper, 2018), the mean shape variances of males alone were used rather than averaging between males and females, or solely using female shape variances. Von

Cramon-Taubadel and Lycett (2008) divided using the geometric mean "[t]o remove the potentially confounding influence of climate on cranial size" (p. 109). As certain relationships between climatic measures and types of cranial shape have been found (Harvati & Weaver, 2006), mean shape variance was adjusted for absolute latitude in the analysis of the present study. Therefore, correlational analysis with respect to mean variances was partial (controlling for absolute latitude). Mean shape variances are available in Appendix C alongside the migratory distances from calculations using Botswana as a starting point (see Section 3.1.2.1), population sample sizes, mean geometric means, and sexual size dimorphisms regarding the Howells cranial dataset.

A way of testing Rensch's rule consists of finding whether the *b* regarding the relationship between the log of female size (*x*-axis) and the log of male size (*y*-axis) is above one, with a value exceeding one being congruent with Rensch's rule, unlike a *b* which is equal to or under one (Fairbairn, 1997). In the current analyses, this method was used, after absolute latitude was adjusted for.

A previous study had explored the relationship between migratory distance and mean cranial shape variance using three non-African (Beijing, Delhi, Tel Aviv) and three African geographic locations from which distances were calculated (von Cramon-Taubadel & Lycett, 2008); regarding the present study, distance-dimorphism analyses (geometric means) were additionally undertaken using those three non-African geographic points substituted in for the southern African origin. Distances were calculated in the same manner as they were for when the southern African location was used.

Research had explored whether cranial diversity (controlling for climate) is predicted by migratory distance, and involved finding whether several climatic measures (maximum and minimum temperatures, and precipitation) and their interactions predicted diversity (Manica et al., 2007); the current analysis did not survey potential climatic contributions to cranial size dimorphism as the number of populations, being 26, would have been small for using the desired number of climatic variables, and, as it was envisioned that migratory distance and absolute latitude ought not to be related, semi-partial correlations were used rather than partial correlations regarding distance-dimorphism and distance-size tests (e.g., the relationship between migratory distance and the cranial size of males was explored after having controlled for absolute latitude determining male cranial size).

Semi-partial and partial correlations were tested for in R Versions 3.5.2 and 3.5.3 (R Core Team, 2018, 2019) by way of the ppcor package (Kim, 2015). Linear regressions (ordinary least squares) were conducted in SPSS Version 25.0. Holm-Bonferroni corrections (Gaetano, 2013; Holm, 1979) were applied across correlational/dimorphism analyses (except for the correlation tests with respect to distance and dimorphism within each of the 56 dimensions, and the rankings of 26 coefficients using African and non-African origins, see Tables A.1 and A.2).

Results

A semi-partial correlation was evident between the magnitude of cranial size dimorphism (having adjusted for absolute latitude) and migratory distance from southern Africa, $sr_s(23) = .58$, $p = .022$, with the amount of dimorphism increasing with distance. The correlation coefficients regarding correlation tests between dimorphism and distance from southern Africa within specific dimensions (Table A.1) lean toward the overall correlation between distance and cranial dimorphism being driven by a sizeable subset of dimensions. Yardsticks of effect sizes for *r* (.50 representing large, .30 for medium, and .10 being small, Cohen, 1992) and the link between *r* and r_s (Strahan, 1982), would indicate that, in the positive direction, there was a strong effect size in seven dimensions, and a medium effect size in 22. With absolute latitude controlled for regarding cranial size, there was no semi-partial correlation between migratory distance (southern Africa origin) and the cranial mean geometric means of either males, $sr_s(23)$ = .44, p = .18, or females, $sr_s(23)$ = .23, p = .80. In the partial correlational analyses, controlling for absolute latitude, as mean shape variance decreased, cranial size increased for males, $pr_s(23)$ = -.65, p = .004, and females, $pr_s(23)$ = -.55, p = .041, and the extent of cranial dimorphism had no association with mean shape variance, $pr_s(23)$ = -.45, p = .18.

Table A.1

Semi-Partial Spearman's Correlation Coefficients Regarding Dimensional Sexual Size Dimorphism (When Absolute Latitude was Controlled for) and Migratory Distance from Southern Africa

Dimension	$sr_s(23)$
Bizygomatic breadth (ZYB)	0.64
Bijugal breadth (JUB)	0.63
Palate breadth, external (MAB)	0.59
Nasion-subtense fraction (FRF)	0.59
Bimaxillary breadth (ZMB)	0.59
Vertex radius (VRR)	0.51
Nasal breadth (NLB)	0.51
Nasion-bregma chord (FRC)	0.45
Biauricular breadth (AUB)	0.44
Minimum cranial breadth (WCB)	0.44
Ectoconchion radius (EKR)	0.43
Biorbital breadth (EKB)	0.43

Nasal height (NLH)	0.42
Basion-bregma height (BBH)	0.41
Naso-dacryal subtense (NDS)	0.40
Nasion radius (NAR)	0.39
Simotic subtense (SIS)	0.39
Bifrontal breadth (FMB)	0.38
Interorbital breadth (DKB)	0.36
Basion-nasion length (BNL)	0.35
Mastoid height (MDH)	0.33
Simotic chord (WNB)	0.33
Dacryon radius (DKR)	0.33
Biasterionic breadth (ASB)	0.33
Lambda-opisthion chord (OCC)	0.32
Molar alveolus radius (AVR)	0.32
Frontomalare radius (FMR)	0.31
Cheek height (WMH)	0.31
Subspinale radius (SSR)	0.31
Glabello-occipital length (GOL)	0.27
Prosthion radius (PRR)	0.26
Zygomaxillary subtense (SSS)	0.25
Nasio-occipital length (NOL)	0.24
Supraorbital projection (SOS)	0.23
Lambda-opisthion subtense (OCS)	0.19
Mastoid breadth (MDB)	0.18
Lambda-subtense fraction (OCF)	0.17
Bregma-lambda chord (PAC)	0.17
Basion-prosthion length (BPL)	0.16
Nasion-prosthion height (NPH)	0.16
Orbit breadth, left (OBB)	0.15
Zygomaxillare radius (ZMR)	0.14
Maximum frontal breadth (XFB)	0.14
Dacryon subtense (DKS)	0.11
Bregma-subtense fraction (PAF)	0.09

Malar length, maximum (XML)	0.09
Nasio-frontal subtense (NAS)	0.08
Zygoorbitale radius (ZOR)	0.08
Malar length, inferior (IML)	0.05
Maximum cranial breadth (XCB)	0.04
Bregma-lambda subtense (PAS)	0.03
Foramen magnum length (FOL)	-0.001
Bistephanic breadth (STB)	-0.05
Orbit height, left (OBH)	-0.07
Malar subtense (MLS)	-0.17
Nasion-bregma subtense (FRS)	-0.22

Note. Full dimension names and abbreviations were from Howells (1989).

With respect to the assessment of Rensch's rule (Fairbairn, 1997), the gradient of the slope for predicting log-transformed male size from log-transformed female size (after having removed any explanatory ability of absolute latitude) was not greater than, or indeed less than, one, $t(23) = 2.27$, $p = .18$, $b = 1.15$ ($SE\ b = .07$, 95% CI [1.01, 1.29]), which suggests that cranial sexual dimorphism did not increase with cranial size (Figure A.1).

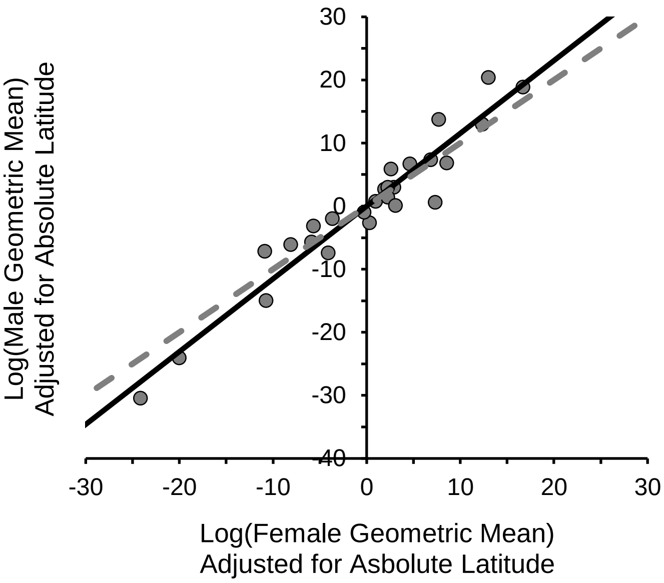


Figure A.1. The relationship between female and male log-transformed geometric means of cranial measurements (multiplied by 1,000 for visual convenience) controlling for absolute latitude. The solid black linear trend line refers to the data analysed in the present study, and the dashed grey line (having a gradient of one) is in regard to there being no dimorphism.

When distances were calculated from each of the three non-African geographic points, dimorphism magnitudes (given absolute latitude) did not correlate with migratory distance, and each of the sr_s was numerically lower than when the African origin was used, Beijing: $sr_s(23) = .05$, $p = .89$; Delhi: $sr_s(23) = .33$, $p = .45$; Tel Aviv: $sr_s(23) = .51$, $p = .076$. For a further indication as to whether the association between distance and dimorphism was stronger with an African (than non-African) origin, using the locations corresponding to the 26 Howells dataset populations as starting locations (i.e., coordinates from von Cramon-Taubadel & Lycett, 2008), 26 post-hoc semi-partial correlation analyses were run between migratory distances and absolute-latitude-adjusted dimorphism. The determination of distances and correlational analyses were undertaken as described in the Method section. When sr_s values were ranked in descending absolute order, coefficients using African origins were each above non-African ones (Table A.2). Therefore, results imply that the correlation of cranial sexual size dimorphism (accounting for absolute latitude) with distance is indicative of the geographic beginning of modern humans.

Table A.2

Ranked Correlation Coefficients (Absolutely Descending) Regarding Dimorphism, Amended for Absolute Latitude, and Distance from Populations

Area	Population	$sr_s(23)$
Africa	Zulu	0.58
Africa	San	0.58
Africa	Dogon	0.56
Africa	Teita	0.56
Africa	Egypt	0.53
Europe	Berg	0.41
Europe	Zalavár	0.41
Europe	Norse	0.29
South America	Peru	-0.25
North America	Santa Cruz	-0.25
North America	Arikara	-0.24
North America	Greenlandic Inuit	-0.24
Asia	Andaman	0.20
Oceania	Mokapu	-0.18
Oceania	Easter Island	-0.15
Asia	Hainan	0.13
Oceania	Moriori	-0.07
Asia	Atayal	0.07
Asia	Ainu	-0.05
Oceania	Tolai	-0.04
Asia	Buryat	0.04
Asia	North Japan	-0.04

Oceania	Guam	-0.03
Oceania	Tasmania	-0.02
Oceania	Australia	0.01
Asia	South Japan	0.002

In the present analyses, the amount of size dimorphism was defined as a ratio variable. Dimorphism variables have indeed been used in correlational/regression analyses in previous research (Betti, 2014; Gustafsson & Lindenfors, 2009; Kurki, 2011; Madrigal & Kelly, 2007; Wells, 2012). Still, whether dimorphism is equivalent across populations has been examined via the presence of a population × sex interaction (Betti, 2014); post-hoc, when the distance-dimorphism relationship was tested by comparing distance-size gradients between males and females (using a formula from Paternoster, Brame, Mazerolle, & Piquero, 1998, and controlling for absolute latitude explaining cranial size), dimorphism was not found to differ as distance from southern Africa broadened, $z = .76$, $p = .89$. However, this outcome should be treated with restraint as the residuals presented in Figure A.2 were contrary to the homogeneity of variances assumption.

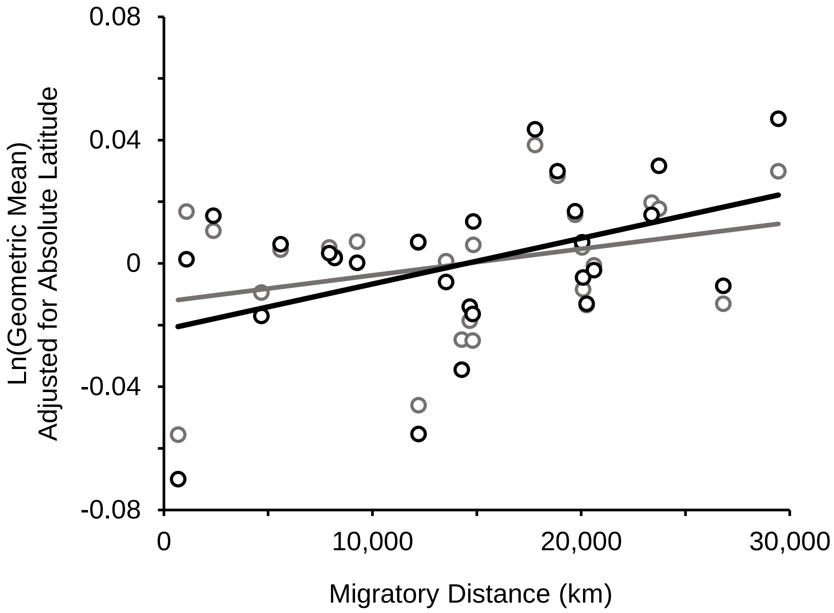


Figure A.2. Residuals from parametric regressions for males and females separately with absolute latitude as the independent variable and ln(mean geometric mean) as the dependent variable (the natural log transformation was used for synergy with the dimorphism variable) (*y*-axis), plotted against migratory distance from southern Africa (*x*-axis). Black-outlined circles and black lines (linear trend lines) are with respect to distances/residuals for males, whilst grey ones are for females.

It is not intended for results to suggest that there would be no cranial size dimorphism at shorter migratory distances; within each of the 26 populations, 99% confidence intervals of geometrics means did not overlap between males and females, and, when those confidence intervals were graphed against migratory distances (origin within southern Africa) for males and females (zero-order), the linear trends lines of the lower limit and upper limit of the confidence intervals for males and females respectively did not overlap within the approximate range of migratory distances used in the current study (circa 0-30,000 km), thereby suggesting that cranial size dimorphism (males bigger than females) occurred across all migratory distances.

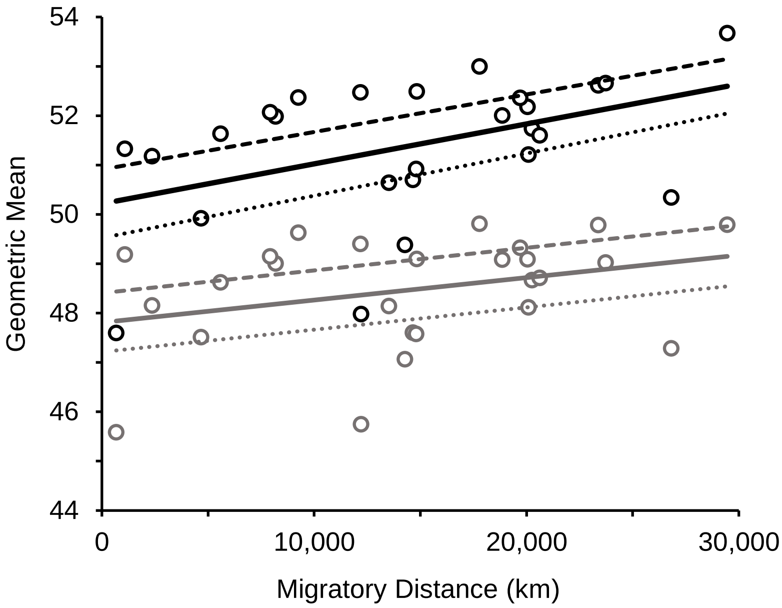


Figure A.3. For males (black circles and lines) and females (grey), the unbroken line, which is a linear trend line, is with respect to mean geometric means. Within each sex, the line above this one is a linear trend line plotted regarding the upper limit of the 99% confidence intervals in each population, whilst the one below concerns the lower limit of that interval.

Discussion

A semi-partial correlation between migratory distance and sexual size dimorphism was evident using a southern African origin, but not when non-African starting locations were employed, and distance-dimorphism coefficients with starts in Africa ranked over those which were non-African. Therefore, if Africa is indeed the origin of modern humans (see Relethford, 2008), as with the relationship between migratory distance and types of variability such as cranial (Betti et al., 2009; Manica et al., 2007; von Cramon-Taubadel & Lycett, 2008), the association between distance and the size dimorphism of the cranium appears to indicate the geographic dawn of *Homo sapiens*.

Whilst migratory distance correlated with dimorphism and not either male or female cranial size, mean shape variance had no association with dimorphism but it did correlate with the sizes. Migratory distance does not by any means wholly account for cranial diversity (von Cramon-Taubadel & Lycett, 2008). For instance, in von Cramon-Taubadel and Lycett (2008), distance explained 31% of the variance in mean cranial shape variance when utilising Botswana as the origin and employing 28 populations from the Howells data (male crania). Therefore, regarding the current study, the relationship between distance and dimorphism was not countered by the occurrence of shape variance not correlating with dimorphism. Perhaps any link between climate and cranial characteristics (Beals et al.,1984; Smith, Terhune, & Lockwood, 2007) was not greatly accounted for in the current study, which resulted in the correlations between cranial shape variance and cranial size.

Migratory distance and cranial size dimorphism: Why the correlation?

Clarity is lacking on what underlies the correlation of distance and dimorphism; with absolute latitude controlled for regarding cranial size, results found that size did not relate to migratory distance from southern Africa (for males or females), and outcomes aligned with cranial size dimorphism not increasing with cranial size. Still, Figure A.2 is suggestive of the correlation between the magnitude of size dimorphism and distance from Africa being a product of the distance-size gradient being greater (numerically) for males than females.

The fall in variability with furthering distance from Africa can be explained as having resulted from a serial founder effect (e.g., Atkinson, 2011; Manica et al., 2007; Ramachandran et al., 2005; von Cramon-Taubadel & Lycett, 2008); given that the results of the present research suggested that the association between dimorphism and distance is more apt with an African (than a non-African) beginning, it might appear reasonable to speculate that a serial founder effect caused cranial size dimorphism to have increased with migratory distance. Furthermore, Table A.2 would seem to follow patterns shown in previous research regarding distance and diversity in terms of the suitability of potential origin locations being greatest in Africa (Manica et al., 2007; Ramachandran et al., 2005), decreasing from Europe and the western half of Asia into the eastern half and Oceania (Betti et al., 2009; Manica et al., 2007; Ramachandran et al., 2005), and increasing into the Americas from north-eastern Asia (Ramachandran et al., 2005). However, it is not forthcoming why an influence of a serial founder effect on dimorphism would occur.

Whether the extent of Neanderthal ancestry may have been a factor in the distance from southern Africa relating to dimorphism could be reflected upon. Results in previous research have been in the direction of Neanderthals having a higher cranial sexual dimorphism than (Upper Paleolithic, and onwards) modern humans (Smith, 1980), including when measurements for modern humans were composed of populations pooled together from a version of the Howells data (Wolpoff, 1980). For example, in Wolpoff (1980), Neanderthal (male divided by female) ratios (two) numerically, although not necessarily significantly, exceeded the average of modern human population ratios on eight out of 10 measurement dimensions. Those results would indicate that Neanderthals were at least on the upper end of Holocene anatomically modern human cranial sexual dimorphism (and, bearing in mind Rensch's rule, Rensch, 1959, perhaps a larger Neanderthal cranial dimorphism over Holocene modern humans could have arisen from Neanderthals having a bigger cranial size, see Introduction).

Populations of modern humans can have extents of archaic human ancestry (e.g., Qin & Stoneking, 2015). Regarding Neanderthal ancestry, previous research suggests that modern humans interbred with Neanderthals outside of Africa, with such ancestry accordingly being present in modern humans of (at least some) non-African genetic heritage (e.g., Wall et al., 2013). Neanderthal ancestry is evident in the cranial shape morphology of modern humans (Gregory et al., 2017), and, compared to Africans, non-Africans do have a greater resemblance to Neanderthals regarding the size of the cranium, with the data of modern humans being from the Howells dataset (Relethford & Smith, 2018); crania could be predicted to generally be more sexually dimorphic amongst Africans than non-Africans, although results in the current study may be suggestive of otherwise when controlling for a latitude-dimorphism relationship (see Figure A.2). In Sankararaman, Mallick, Patterson, and Reich (2016), persons of the Oceanic region were found to have more Neanderthal genetic ancestry than West Eurasians, but not East Asians (although Oceanic persons were numerically higher than East Asians with respect to Neanderthal ancestry). Regarding the populations featured in Sankararaman et al. (2016), and judging by theorised migratory routes (Oppenheimer, 2012), the disparity in migratory distance (calculated from within Africa) between Oceanic and West Eurasian groups would be larger than between Oceanic populations and East Asians. Therefore, it could seem that Neanderthal ancestry may have been a potential factor responsible for the magnitude of dimorphism increasing with distance.

However, preceding research is suggestive of the cranial size dimorphism of modern humans being lower in the Holocene than the Upper Palaeolithic period (Frayer, 1980; Smith, 1980); Neanderthals are regarded as being of the

Pleistocene (e.g., Quinney & Collard, 1997) and having become extinct prior to the Holocene (Higham et al., 2014; Zilhão et al., 2017). Therefore, a fall in cranial size dimorphism as time progressed (Frayer, 1980; Smith, 1980) would not seem supportive of a speculation that the amount of Neanderthal ancestry within Holocene modern humans affected cranial size dimorphism. Furthermore, Neanderthals are not the sole archaic human group to have an ancestral presence in the genetic ancestry of modern humans (e.g., Reich et al., 2010; Qin & Stoneking, 2015), however, the extent of cranial sexual dimorphism amongst those other archaic humans appears to be unknown. Still, results of the present analyses raise the idea that the chronological decline in cranial size dimorphism (i.e., Frayer, 1980; Smith, 1980) was buffered against by whatever the process was which caused dimorphism to correlate with migratory distance.

It has been speculated as to whether the size of a population is a potential contributor to sexual dimorphism regarding postcranial dimorphism being higher in a rural setting (i.e., a small population) compared to an urban one (Charisi, Laffranchi, & Jiménez-Brobeil, 2016). Interestingly, “[t]he worldwide expansion of humans out of Africa probably happened in many small steps” with "each step involving a small sample of founders from the population at the front of expansion” (Deshpande, Batzoglou, Feldman, & Cavalli-Sforza, 2009, p. 291); was the reduction of cranial size dimorphism with the advancement of time (see Frayer, 1980; Smith, 1980) minimised by the successive featuring of small founding populations as migration furthered? There is no relationship between population size and stature dimorphism (Gray & Wolfe, 1980). Whilst there are positive associations between stature and head measurements (Krishan & Kumar, 2007) such as cranial capacity (Acer, Usanmaz, & Erteki’n, 2007), considering that femoral and pelvic size dimorphisms are largely unrelated (Kurki, 2011), it should not be assumed that what does or does not relate to one type of dimorphism also holds for another type; whether the size of populations relates to cranial size dimorphism is unknown.

Given the idea of there being a diminishing in reactive aggression[45] during the history of humans (Wrangham, 2018, 2019), as supported by a fall in dimorphism, and if "where food resources are constrained, women should prefer more masculine and aggressive partners who are more capable of securing needed resources and might pass this capability to offspring" (Gleeson & Kushnick, 2018, p. 459), perhaps the uncertainty over food, which may accompany the founding of a new population, would have promoted females wanting male partners higher in masculinity/aggressiveness, thereby cushioning the aforementioned fall across time of cranial size dimorphism. However, on the masculinity of faces, the attractiveness preferences of heterosexual and bisexual women are lower (in terms of facial masculinity) when envisioning a harsher life situation compared to a less harsh one (Little, Cohen, Jones, & Belsky, 2007), and, moreover, the extent to which heterosexual women are concerned about their future personal finances (as an indicator of worries pertaining to resources) does not interact with the facial masculinity of face stimuli regarding how attractive faces are rated to be (Holzleitner & Perrett, 2017).

Limitations

A decline in cranial capacity within the Holocene (Henneberg & Steyn, 1993) could signal that changes in crania over time may have dulled the ability of

[45] Reactive aggression has been described as "a response to a threat or frustrating event, with the goal being only to remove the provoking stimulus" (Wrangham, 2018, p. 246).

the present analyses to find associations between distance and cranial size. The number of populations (26) may have been on the small side for the present analyses, and, in the concert with applying Holm-Bonferroni corrections, have led to Type II errors. Indeed, respectable effect sizes (Cohen, 1992; Strahan, 1982) were found in the correlation analyses between i) migratory distance and male cranial size, ii) mean shape variance and dimorphism, and iii) dimorphism and migratory distance from Tel Aviv, yet these tests had non-significant *p*-values following the Holm-Bonferroni adjustments. Moreover, concerning the gradient with respect to male and female cranial size (Figure A.1), the number one was outside of the 95% confidence interval.[46]

Further directions

This present study was envisioned as a first venture regarding a possible connection between the migratory distance of modern humans and sexual dimorphism. Subsequent research is encouraged, ideally with data which have a greater number of population groups (males, and females). The Hanihara dataset (e.g., Hanihara, 1997; Hanihara & Ishida, 2005; Hubbe et al., 2009) contains more populations of males and females than the Howells dataset. It has been analysed regarding whether cranial variability diminished as migratory distance became longer (Betti et al., 2009; Manica et al., 2007). Both datasets could be combined regarding common dimensions (to maximise the number of populations, although not all dimensions are mutual) (Algee-Hewitt, 2011). With

[46] Unadjusted *p*-values were lower than .05 regarding the correlation tests between: distance from southern Africa and male cranial size (p = .030) (semi-partial), variance and dimorphism (p = .025) (partial), distance from Tel Aviv and dimorphism (p =.010) (semi-partial). This also occurred regarding testing whether the slope between log-transformed female and male cranial size was different from the figure one (p = .033), having taken into bearing absolute latitude.

or without this unification of datasets, research could explore the ancillary finding of cranial size correlating with shape variance amongst males and females. It could be determined whether those associations were due to a particular group of dimensions. It may be desirable to find whether archaic ancestry levels relate to cranial size dimorphism and cranial size. Some populations of the Howells cranial data have previously been paired with genetic data (Roseman, 2004); perhaps pairings could occur between cranial data and the archaic ancestry levels presented in Sankararaman et al. (2016).

Studies on whether morphological diversity fell as migratory distance extended have not been limited to cranial measurements, but have also used postcranial ones (e.g., Betti et al., 2012). Postcranial dimensions could also be made use of regarding whether dimorphism relates to migratory distance. Interestingly, mandibular sexual size dimorphism has been found to be lower in Holocene than Pleistocene modern humans, with the Holocene sample being from Europe whilst the Pleistocene group was more global (Quinney & Collard, 1997). If the size dimorphism of the mandible increased with migratory distance, this would raise the possibility of the difference between Pleistocene and Holocene humans potentially having arisen from geographic sampling.

Whilst research on the relationship concerning distance and cranial diversity has explored which geographic starting area leads to an optimal relationship (suggesting the origin of *Homo sapiens*) (Betti et al., 2009; Manica et al., 2007), the current study did not do so in as great a depth regarding the distance-dimorphism association; it is hoped that future studies will consider this. Therefore, whilst the present research may have found a new indicator of the place at which modern humans had their first onset, the usefulness of the apparent link between distance and dimorphism in this regard seems quite a distance from being certain.

References

Abouheif, E., & Fairbairn, D. J. (1997). A comparative analysis of allometry for sexual size dimorphism: Assessing Rensch's rule. *The American Naturalist*, *149*(3), 540–562.

Acer, N., Usanmaz M., & Erteki'n, T. (2007). Estimation of cranial capacity in 17-26 year old university students. *International Journal of Morphology*, *25*(1), 65–70.

Algee-Hewitt, B. F. B. (2011). *If and how many 'races': The application of mixture modeling to world-wide human craniometric variation* (Unpublished doctoral thesis). University of Tennessee, Knoxville, TN.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, *332*(6027), 346–349.

Balloux, F., Lawson Handley, L.-J., Jombart, T., Liu, H., & Manica, A. (2009). Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1672), 3447–3455.

Beals, K. L., Smith, C. L., & Dodd, S. M. (1984). Brain size, cranial morphology, climate, and time machines. *Current Anthropology*, *25*(3), 301–330.

Betti, L. (2014). Sexual dimorphism in the size and shape of the os coxae and the effects of microevolutionary processes. *American Journal of Physical Anthropology*, *153*(2), 167–177.

Betti, L., Balloux, F., Amos, W., Hanihara, T., & Manica, A. (2009). Distance from Africa, not climate, explains within-population phenotypic diversity in humans. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1658), 809–814.

Betti, L., von Cramon-Taubadel, N., Manica, A., & Lycett, S. J. (2013). Global geometric morphometric analyses of the human pelvis reveal substantial neutral population history effects, even across sexes. *PLoS ONE*, *8*(2), 1–10. doi:10.1371/journal.pone.0055909

Betti, L., von Cramon-Taubadel, N., & Lycett, S. J. (2012). Human pelvis and long bones reveal differential preservation of ancient population history and migration out of Africa. *Human Biology*, *84*(2), 139–152.

Bhopal, R. (2004). Glossary of terms relating to ethnicity and race: For reflection and debate. *Journal of Epidemiology & Community Health*, *58*(6), 441–445.

Brewster, C., Meiklejohn, C., von Cramon-Taubadel, N. & Pinhasi, R. (2014). Craniometric analysis of European Upper Palaeolithic and Mesolithic samples supports discontinuity at the Last Glacial Maximum. *Nature Communications*, *5*, 1–10. doi:10.1038/ncomms5094

Charisi, D., Laffranchi, Z., & Jiménez-Brobeil, S. A. (2016). Sexual dimorphism in two mediaeval Muslim populations from Spain. *Journal of Comparative Human Biology*, *67*(5), 397–408.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Dale, J., Dunn, P. O., Figuerola, J., Lislevand, T., Székely, T., & Whittingham, L. A. (2007). Sexual selection explains Rensch's rule of allometry for sexual size dimorphism. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1628), 2971–2979.

Deshpande, O., Batzoglou, S., Feldman, M. W., & Cavalli-Sforza, L. L. (2009). A serial founder effect model for human settlement out of Africa. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1655), 291–300.

Fairbairn, D. J. (1997). Allometry for sexual size dimorphism: Pattern and process in the coevolution of body size in males and females. *Annual Review of Ecology and Systematics*, *28*, 659–687.

Frayer, D. W. (1980). Sexual dimorphism and cultural evolution in the Late Pleistocene and Holocene of Europe. *Journal of Human Evolution*, *9*(5), 399–415.

Gaetano, J. (2013). Holm-Bonferroni sequential correction: An EXCEL calculator. Retrieved from https://www.researchgate.net/publication/236969037_Holm-Bonferroni_Sequential_Correction_An_EXCEL_Calculator

Gleeson, B. T., & Kushnick, G. (2018). Female status, food security, and stature sexual dimorphism: Testing mate choice as a mechanism in human self-domestication. *American Journal of Physical Anthropology*, *167*(3), 458–469.

Gordon C. C., Churchill, T., Clauser, C. E., Bradtmiller, B., McConville, J. T., Tebbetts, I., & Walker, R. A. (1989). *1988 Anthropometric Survey of U.S. Army Personnel: Methods and summary statistics*. Natick, MA: United States Army Natick Research, Development and Engineering Center.

Gray, J. P., & Wolfe, L. D. (1980). Height and sexual dimorphism of stature among human societies. *American Journal of Physical Anthropology*, *53*(3), 441–456.

Gregory, M. D., Kippenhan, J. S., Eisenberg, D. P., Kohn, P. D., Dickinson, D., Mattay, V. S., … Berman, K. F. (2017). Neanderthal-derived genetic variation shapes modern human cranium and brain. *Scientific Reports*, *7*, 1–11. doi:10.1038/s41598-017-06587-0

Gustafsson, A., & Lindenfors, P. (2004). Human size evolution: No evolutionary allometric relationship between male and female stature. *Journal of Human Evolution*, *47*(4), 253–266.

Gustafsson, A., & Lindenfors, P. (2009). Latitudinal patterns in human stature and sexual stature dimorphism. *Annals of Human Biology*, *36*(1), 74–87.

Hanihara T. (1997). Craniofacial affinities of Mariana Islanders and Circum-Pacific peoples. *American Journal of Physical Anthropology*, *104*(3), 411–425.

Hanihara, T. (2008). Morphological variation of major human populations based on nonmetric dental traits. *American Journal of Physical Anthropology*, *136*(2), 169–182.

Hanihara T., & Ishida, H. (2005). Metric dental variation of major human populations. *American Journal of Physical Anthropology*, *128*(2), 287–298.

Harvati, K., & Weaver, T. D. (2006). Human cranial anatomy and the differential preservation of population history and climate signatures. *The Anatomical Record Part A*, *288A*(12), 1225–1233.

Henneberg, M. (1988). Decrease of human skull size in the Holocene. *Human Biology*, *60*(3), 395–405.

Henneberg, M., & Steyn, M. (1993). Trends in cranial capacity and cranial index in Subsaharan Africa during the Holocene. *American Journal of Human Biology*, *5*(4), 473–479.

Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., … Jacobi, R. (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*, *512*(7514), 306–309.

Holzleitner, I. J., & Perrett, D. I. (2017). Women's preferences for men's facial masculinity: Trade-off accounts revisited. *Adaptive Human Behavior and Physiology*, *3*(4), 304–320.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Howells, W. W. (1989). *Skull shapes and the map: Craniometric analyses in the dispersion of Modern Homo. Papers of the Peabody Museum of Archaeology and Ethnology Volume 79*. Cambridge, MA: Harvard University.

Howells, W. W. (1995). *Who's who in skulls: Ethnic identification of crania from measurements. Papers of the Peabody Museum of Archaeology and Ethnology Volume 82*. Cambridge, MA: Harvard University.

Howells, W. W. (1996). Howells' craniometric data on the internet. *American Journal of Physical Anthropology*, *101*(3), 441–442.

Hubbe, M., Hanihara, T., & Harvati, K. (2009). Climate signatures in the morphological differentiation of worldwide modern human populations. *The Anatomical Record*, *292*(11), 1720–1733.

Jantz, R. L., & Jantz, L. M. (2016). The remarkable change in Euro-American cranial shape and size. *Human Biology*, *88*(1), 56–64.

Jin, L., & Su, B. (2000). Natives or immigrants: Modern human origin in East Asia. *Nature Reviews Genetics*, *1*(2), 126–133.

Jungers, W. L., Falsetti, A. B., & Wall, C. E. (1995). Shape, relative size, and size-adjustments in morphometrics. *Yearbook of Physical Anthropology*, *38*(S21), 137–161.

Kappelman, J. (1996). The evolution of body mass and relative brain size in fossil hominids. *Journal of Human Evolution*, *30*(3), 243–276.

Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674.

Kimmerle, E. H., Ross, A., & Slice, D. (2008). Sexual dimorphism in America: Geometric morphometric analysis of the craniofacial region. *Journal of Forensic Sciences*, *53*(1), 54–57.

Krishan, K., & Kumar, R. (2007). Determination of stature from cephalo-facial dimensions in a North Indian population. *Legal Medicine*, *9*(3), 128–133.

Kurki, H. K. (2011). Pelvic dimorphism in relation to body size and body size dimorphism in humans. *Journal of Human Evolution*, *61*(6), 631–643.

L'Abbé, E. N., Kenyhercz, M., Stull, K. E., Keough, N., & Nawrocki, S. (2013). Application of Fordisc 3.0 to explore differences among crania of North American and South African Blacks and Whites. *Journal of Forensic Sciences*, *58*(6), 1579–1583.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., … Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, *319*(5866), 1100–1104.

Little, A. C., Cohen, D. L., Jones, B. C., & Belsky, J. (2007). Human preferences for facial masculinity change with relationship type and environmental harshness. *Behavioral Ecology and Sociobiology*, *61*(6), 967–973.

Madrigal, L., & Kelly, W. (2007). Human skin-color sexual dimorphism: A test of the sexual selection hypothesis. *American Journal of Physical Anthropology*, *132*(3), 470–482.

Maina, M. B., Mahdi, O., & Kalayi, G. D. (2011). Sexual dimorphism in cranial dimensions among three ethnic groups of North-Eastern Nigeria. *American Journal of Scientific and Industrial Research*, *2*(6), 871–876.

Manica, A., Amos, W., Balloux, F., & Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. *Nature*, *448*(7151), 346–348.

Newman, M. T. (1953). The application of ecological rules to the racial anthropology of the Aboriginal new world. *American Anthropologist*, *55*(3), 311–327.

Oppenheimer, S. (2012). Out-of-Africa, the peopling of continents and islands: Tracing uniparental gene trees across the map. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 770–784.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859–866.

Prugnolle, F., Manica, A., & Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Current Biology*, *15*(5), R159–R160.

Qin, P., & Stoneking, M. (2015). Denisovan ancestry in East Eurasian and Native American populations. *Molecular Biology and Evolution*, *32*(10), 2665–2674.

Quinney, P. S., & Collard, M. (1997). Sexual dimorphism in the mandible of

Homo Neanderthalensis and *Homo Sapiens*: Morphological patterns and behavioural implications. In A. Sinclair, E. Slater, & J. Gowlett (Eds.), *Archaeological sciences 1995* (pp. 420–425). Oxford, UK: Oxbow Books.

R Core Team (2018). R: A Language and Environment for Statistical Computing (Version 3.5.2) [Computer software]. Retrieved from https://www.R-project.org

R Core Team (2019). R: A Language and Environment for Statistical Computing (Version 3.5.3) [Computer software]. Retrieved from https://www.R-project.org

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M.W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(44), 15942–15947.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., … Svante Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*, 1053–1060.

Relethford, J. H. (2008). Genetic evidence and the modern human origins debate. *Heredity*, *100*, 555–563.

Relethford, J. H., & Blangero, J. (1990). Detection of differential gene flow from patterns of quantitative variation. *Human Biology*, *62*(1), 5–25.

Relethford, J. H., & Smith, F. H. (2018). Cranial measures and ancient DNA both show greater similarity of Neandertals to recent modern Eurasians than to recent modern sub-Saharan Africans. *American Journal of Physical Anthropology*, *166*(1), 170–178.

Rensch, B. (1959). *Evolution above the species level*. London, UK: Methuen and Co.

Reyes-Centeno, H., Ghirotto, S., Détroit, F., Grimaud-Hervé, D., Barbujani, G., & Harvati, K. (2014). Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), 7248–7253.

Risch, N., Burchard, E., Ziv, E., & Tang, H. (2002). Categorization of humans in biomedical research: Genes, race and disease. *Genome Biology*, *3*(7), 1–12. doi:10.1186/gb-2002-3-7-comment2007

Roseman, C. C. (2004). Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 12824–12829.

Rushton, J. P. (1992). Cranial capacity related to sex, rank, and race in a stratified random sample of 6,325 U.S. military personnel. *Intelligence*, *16*(3–4), 401–413.

Rushton, J. P. (1994). Sex and race differences in cranial capacity from international labour office data. *Intelligence*, *19*(3), 281–294.

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Current Biology*, *26*(9), 1241–1247.

Schutz, H., Polly, P. D., Krieger, J. D., & Guralnick, R. P. (2009). Differential sexual dimorphism: Size and shape in the cranium and pelvis of grey foxes (*Urocyon*). *Biological Journal of the Linnean Society*, *96*(2), 339–353.

Short, A. (2016). The relationship between latitudinal light variation and orbit and cranial size in humans. *Anthropological Review*, *79*(2), 143–149.

Smith, F. H. (1980). Sexual differences in European Neanderthal crania with special reference to the Krapina remains. *Journal of Human Evolution*, *9*(5), 359–375.

Smith, H. F., Terhune, C. E., & Lockwood, C. A. (2007). Genetic, geographic, and environmental correlates of human temporal bone variation. *American Journal of Physical Anthropology*, *134*(3), 312–322.

Smith, R. J. (1999). Statistics of sexual size dimorphism. *Journal of Human Evolution*, *36*(4), 423–459.

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences, 56*(2), 289–296.

Stepan, N. L. (1986). Race and gender: The role of analogy in science. *Isis*, *77*(2), 261–277.

Stevens, R. D., & Platt, R. N. (2015). Patterns of secondary sexual size dimorphism in New World Myotis and a test of Rensch’s rule. *Journal of Mammalogy*, *96*(6), 1128–1134.

Steyn, M., & İşcan, M. Y. (1998). Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Science International*, *98*(1–2), 9–16.

Strahan, R. F. (1982). Assessing magnitude of effect from rank-order correlation coefficients. *Educational and Psychological Measurement*, *42*(3), 763–765.

Székely, T., Freckleton, R. P., & Reynolds, J. D. (2004). Sexual selection explains Rensch’s rule of size dimorphism in shorebirds. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(33), 12224–12227.

Tassi, F., Ghirotto, S., Mezzavilla, M., Vilaça, S. T., De Santi, L., & Barbujani, G. (2015). Early modern human dispersal from Africa: Genomic evidence for multiple waves of migration. *Investigative Genetics*, *6*(13), 1–16. doi:10.1186/s13323-015-0030-2

United States Census Bureau (2017, January 12). *About*. Retrieved from https://www.census.gov/topics/population/race/about.html

von Cramon-Taubadel, N., & Lycett, S. J. (2008). Brief communication: Human cranial variation fits iterative founder effect model with African origin. *American Journal of Physical Anthropology*, *136*(1), 108–113.

Wall, J. D., Yang, M. A., Jay, F., Kim, S. K., Durand, E. Y., Stevison, L. S., … Slatkin, M. (2013). Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics*, *194*(1), 199–209.

Wells, J. C. K. (2012). Sexual dimorphism in body composition across human populations: Associations with climate and proxies for short- and long-term energy supply. *American Journal of Human Biology*, *24*(4), 411–419.

Williams, E. (2011). *Aviation formulary V1.46*. Retrieved from http://edwilliams.org/avform.htm

Wilmshurst, J. M., Anderson, A. J., Higham, T. F. G., & Worthy, T. H. (2008). Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(22), 7676 –7680.

Wilmshurst, J. M., Hunt, T. L., Lipo, C. P., & Anderson, A. J. (2011). High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proceedings of the National Academy of*

Sciences of the United States of America, *108*(5), 1815–1820.

Wolpoff, M. H. (1980). Cranial remains of Middle Pleistocene European hominids. *Journal of Human Evolution*, *9*(5), 339–358.

Wood, B., & Collard, M. (1999). The human genus. *Science*, *284*(5411), 65–71.

Weaver, T. D., Roseman, C. C., & Stringer, C. B. (2007). Were neandertal and modern human cranial differences produced by natural selection or genetic drift? *Journal of Human Evolution*, *53*(2), 135–145.

Wrangham, R. W. (2018). Two types of aggression in human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(2), 245–253.

Wrangham, R. W. (2019). Hypotheses for the evolution of reduced reactive aggression in the context of human self-domestication. *Frontiers in Psychology*, *10*, 1–11. doi:10.3389/fpsyg.2019.01914

Zichello, J. M., Baab, K. L., McNulty, K. P., Raxworthy, C. J., & Steiper, M. E. (2018). Hominoid intraspecific cranial variation mirrors neutral genetic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11501–11506.

Zilhão, J., Anesin, D., Aubry, T., Badal, E., Cabanes, D., Kehl, M., … Zapata, J. (2017). Precise dating of the Middle-to-Upper Paleolithic transition in Murcia (Spain) supports late Neandertal persistence in Iberia. *Heliyon*, *3*(11), 1–51. doi:10.1016/j.heliyon.2017.e00435

Appendix C

Summary Statistics for the Howells Populations

Population	Mig. dist.	Sample size M	Sample size F	Geo. mean M	Geo. mean F	MV	Dim.
Anyang	13,142.05	42					
Greenlandic Inuit	20,261.24	53	55	51.74	48.67	0.76	0.061
Buryat	12,187.82	55	54	52.47	49.40	0.75	0.060
Ainu	14,833.98	48	38	52.49	49.10	0.62	0.067
Andaman	12,208.12	35	35	47.98	45.75	0.80	0.048
San	676.30	41	49	47.60	45.59	1.00	0.043
Egypt	5,599.35	58	53	51.63	48.63	0.63	0.060
Guam	17,792.82	30	27	53.00	49.81	0.57	0.062
Philippines	15,242.62	50					
Atayal	14,283.30	29	18	49.38	47.07	0.69	0.048
Hainan Island	13,525.69	45	38	50.64	48.14	0.68	0.051
South Japan	14,659.62	50	41	50.70	47.60	0.72	0.063
North Japan	14,799.86	55	32	50.92	47.58	0.79	0.068
Peru	26,819.65	55	55	50.34	47.28	0.62	0.063
Santa Cruz	20,090.41	51	51	51.22	48.12	0.59	0.062
Arikara	20,053.93	42	27	52.18	49.09	0.58	0.061
Moriori	23,380.77	57	51	52.62	49.78	0.64	0.055
Easter Island	29,454.05	49	37	53.67	49.79	0.59	0.075
Mokapu	23,733.36	51	49	52.66	49.03	0.60	0.072
Tolai	18,866.29	56	54	52.01	49.08	0.60	0.058
Tasmania	20,621.33	45	42	51.61	48.72	0.69	0.058
Australia	19,706.54	52	49	52.37	49.33	0.55	0.060
Zulu	1,081.12	55	46	51.33	49.19	0.77	0.043
Dogon	4,675.90	47	52	49.92	47.51	0.82	0.049
Teita	2,369.89	33	50	51.18	48.16	0.84	0.061
Berg	8,190.67	56	53	51.99	49.00	0.74	0.059
Zalavár	7,928.09	53	45	52.07	49.15	0.63	0.058
Norse	9,255.68	55	55	52.37	49.63	0.65	0.054

South Māori	10
North Māori	10

Note. Mig. dist. = migratory distance; M = male; F = female; Geo. = geometric; MV = mean variance of shape; Dim. = dimorphism. Regarding Appendix B, all 30 populations were utilised in the brief analysis written of in the Introduction (construction of 99% confidence intervals), and the 26 populations featuring males and females were employed in the main analysis.