Cultural differences in the development of face perception

Cultural Differences in the Development of Face Perception

Jennifer X. Haensel

Department of Psychological Sciences

Birkbeck, University of London

A thesis submitted for the degree of

Doctor of Philosophy

Originality Statement

I, Jennifer Xiu-Fang Haensel, declare that the work submitted in this thesis is my own. Where information has been derived from other sources or where significant contribution to the work involved other people, I confirm that this has been indicated in the thesis.

Signed: ______________________________________

14 December 2018

Acknowledgements

The completion of this thesis would not have been possible without the support and guidance of many people. First and foremost, I would like to thank my supervisors. Atsushi Senju, whose invaluable insights and theoretical rigour allowed me to grow as a researcher, and Tim Smith, whose methodological expertise and continued support for my often challenging ideas helped me find my own path in research. Thanks to both for taking me on as a PhD student when I knew absolutely nothing about babies, for labelling my project "very high-risk" but letting me pursue it nevertheless, and, most of all, for their academic and personal support throughout the years.

I was incredibly fortunate to collaborate with numerous people. Most of all, I would like to express my gratitude to Shoji Itakura, who generously allowed me to work in his friendly lab in Kyoto, and who introduced me to Japanese culture. I would also like to thank Mitsu Ishikawa, who worked tirelessly to collect data in Kyoto and who provided stimulating discussions. To Nadia Neesgaard for her help with data collection despite four exhausting months of failed piloting, to Matt Danvers for remaining pro-active despite the tight timeline for data collection, and to Sarah Kuhn for her hard work throughout her placement. Thanks also to Raffaele Tucciarelli, whose expert knowledge came at the right time to bring the analysis to the finish line.

Thanks also to Scott Hodgins, Luke Mason, and Nanami Toya for their help with the studies. To Ana, Angélina, Harish, Irati, Marian, Nanami, and Nataşa, who I could always turn to for help and advice. This work would not have been possible without the participating families, and the MRC and JSPS who generously funded this work.

Thank you to those who made my time at Birkbeck enjoyable. To Jono for putting up with "April" and cheering me on when work became difficult, to Manu for fun chats and gigs (and support and all), and to Tobi – dangermouse sagt liebsten Dank.

Beyond Birkbeck, I would like to thank Paul Overton for invaluable advice back in 2015 that put things back into perspective. Cimer, Max, for helping me settle in Kyoto and making my time an enjoyable one. Thanks to Péter for long-distance battleships and knowing when it's time for Tetris (and for all the support during the past year). To Maria and Vee for checking in and managing to switch off my 'thesis mode'. To Lisa and Emily for always having my back, and to Nina for a rather memorable time in Japan. To Georgie who always had the right words ready to tackle thesis-induced stress, who was there for me no matter where I was in the world, and who made the last five years so much more fun! Thanks to my mother for her unconditional support and for patiently putting up with my work ethic all these years without a single complaint. To Asa who supported me more than anyone, in every possible way.

My final acknowledgement is to my friend Andrew Ruddiman, whose words kept me going throughout the PhD, and to my father Peter Haensel, who would have been the happiest and proudest of all. This thesis is dedicated to them.

Abstract

The development of specialised face processing is shaped by postnatal social experience. Previous literature indicates cultural influences on face scanning, but when and how culture modulates the development of expert face processing remains unclear. Current interpretations are additionally restricted to highly controlled screen-based paradigms that lack the social presence and visual complexities common to social interactions. This thesis explores cultural differences in infants' and adults' face scanning during naturalistic dyadic interactions and within screen-based paradigms to cast light on possible mechanisms that can explain how the postnatal environment shapes face perception. Chapter 2 discusses the significant methodological challenges associated with the analysis of head-mounted eye tracking data and presents a semi-automatic computational solution as well as a novel data-driven method based on permutation testing. Chapter 3 adopts dual eye tracking techniques in Western Caucasian and East Asian adults to explore face scanning during dyadic interactions. Chapter 4 presents a methodologically refined follow-up study and reveals greater eye scanning in Japanese adults and more mouth looking in British/Irish individuals. Chapter 5 employs a cross-sectional screen-based paradigm to examine face scanning in British and Japanese infants (aged 10 and 16 months) and adults. Independent effects of culture and age are revealed, suggesting that cultural differences largely manifest by 10 months of age. Chapter 6 examines whether scanning strategies of British and Japanese 10-month-olds extend to dyadic interactions but finds that both groups predominantly scan the lower face region. Altogether, the thesis findings suggest that the manifestation of cultural differences in face scanning and the degree to which they can be observed depends on various factors, e.g., age, social presence, or the dynamic complexity of faces. Overall, this points to a highly adaptive face processing system that is shaped by early postnatal social experience and modulated by contextual factors.

Table of Contents

Originality Statement **2**

Acknowledgements **3**

Abstract **4**

List of Figures **11**

List of Tables **16**

List of Abbreviations **24**

Chapter 1: General Introduction **25**

1.1. Chapter Overview 26

1.2. Introduction 26

1.2.1. Cultural differences in face scanning 29

1.2.2. Face perception in the 'real world' and the relevance for cross-cultural comparisons 30

1.2.3. Explaining cultural differences in face scanning 33

1.2.4. Cultural differences in the development of face scanning 38

1.2.5. Outstanding questions and thesis aims 40

Chapter 2: Methods Development for 'Real-World' Eye Tracking **42**

2.1. Chapter Overview 43

2.2. Introduction 43

2.2.1. General principles of eye tracking 43

2.2.2. Differences in eye tracking systems and model selections for the present thesis 45

2.2.3. Calibration 49

2.2.4. Pre-processing eye tracking data 51

2.3. Data Analysis 54

2.3.1. Background: Regions-of-interest analysis 54

2.3.2. Semi-automatic regions-of-interest coding and analysis 57

2.3.3. Background: Data-driven analysis 73

2.3.4. Data-driven analysis of 'real-world' eye tracking data 77

2.4. Conclusion 86

Chapter 3: Cultural Differences in Face Scanning During Live Dyadic Interactions I 87

3.1. Chapter Overview 88

3.2. Introduction 88

3.2.1. Cultural differences in face scanning 88

3.2.2. Limitations of screen-based paradigms 89

3.2.3. The current study 91

3.3. Methods 93

3.3.1. Participants 93

3.3.2. Apparatus 94

3.3.3. Procedure 95

3.3.4. Data pre-processing 97

3.4. Results 99

3.4.1. Regions-of-interest analysis 99

3.4.2. Monte Carlo permutation test 103

3.4.3. Mutual gaze 107

3.4.4. Autism Quotient (AQ) 108

3.5. Discussion 109

Chapter 4: Cultural Differences in Face Scanning During Live Dyadic Interactions II 117

4.1. Chapter Overview..118

4.2. Introduction 118

4.2.1. Equipment changes.. 120

4.2.2. Paradigm changes 121

4.2.3. Participant changes.. 122

4.2.4. Measurement of social anxiety traits 123

4.2.5. The current study 124

4.3. Methods 125

4.3.1. Participants 125

4.3.2. Apparatus... 126

4.3.3. Procedure 127

4.3.4. Data pre-processing 130

4.4. Results..131

4.4.1. Regions-of-interest analysis...131

4.4.2. Monte Carlo permutation test.. 139

4.4.3. Autism Quotient (AQ) 142

4.4.4. Liebowitz Social Anxiety Scale (LSAS) 143

4.4.5. Control analysis for age... 144

4.5. Discussion 145

Chapter 5: Cultural Differences in the Development of Face Scanning and the Role of Executive Function 153

5.1. Chapter Overview.. 154

5.2. Introduction 154

5.2.1. Cross-cultural developmental studies on face scanning 154

5.2.2. The role of executive function 156

5.2.3. The current study 158

5.3. Methods 160

5.3.1. Participants 160

5.3.2. Apparatus 163

5.3.3. Design 164

5.3.4. Procedure 173

5.3.5. Data pre-processing 175

5.4. Results 177

5.4.1. Regions-of-interest (ROI) analysis 177

5.4.2. Gap-overlap task 194

5.4.3. Cognitive control task 196

5.4.4. Relationship between executive function and face scanning 197

5.5. Discussion 199

5.5.1. Cultural differences in face scanning 200

5.5.2. Age differences in face scanning 202

5.5.3. Age-related changes for cultural differences in face scanning 203

5.5.4. Role of face ethnicity and stimulus condition 204

5.5.5. Executive function and its role in the development of face scanning 206

5.5.6. Conclusion 208

Chapter 6: Cultural Differences in Naturalistic Face Scanning of 10-Month-Old Infants 210

6.1. Chapter Overview..211

6.2. Introduction 211

6.2.1. The current study 212

6.3. Methods 213

6.3.1. Participants 213

6.3.2. Apparatus 214

6.3.3. Procedure 216

6.3.4. Data pre-processing 217

6.4. Results 219

6.4.1. Regions-of-interest analysis... 219

6.4.2. Monte Carlo permutation test..222

6.5. Discussion 224

Chapter 7: General Discussion... 228

7.1. Thesis Overview 229

7.2. Critical Discussion 231

7.2.1. Cultural differences in adults' face scanning 231

7.2.2. Mapping the developmental trajectory... 235

7.2.3. Developmental mechanisms for the expression of cultural differences. 237

7.2.4. Eye tracking methodology 240

7.3. Limitations... 241

7.3.1. What is 'naturalistic'?... 241

7.3.2. Systematic differences in social interaction partner.............................244

7.3.3. Parameter setting for analysis 245

7.3.4. Cultural differences and the role of genetics 248

7.4. Future Directions 249

7.4.1. Improving the regions-of-interest (face) tracker 249

7.4.2. The temporal dynamics of face scanning 250

7.4.3. What about covert attention? 251

7.4.4. The neural basis of cultural differences in face processing 251

7.5. Conclusion 253

References 255

Appendix A 284

Appendix B 288

Appendix C 289

Appendix D 290

Appendix E 294

Appendix F 298

Appendix G 343

List of Figures

Figure 1.1. Western Caucasians showed greater triangular scanning of the eyes and mouth (red), whereas East Asians fixated more on the face centre (blue). Source: Blais et al. (2008). 29

Figure 2.1. Participant wearing the Positive Science head-mounted eye tracker. 47

Figure 2.2. Adult wearing the SMI eye tracking glasses....47

Figure 2.3. Infant wearing the Positive Science eye tracker. Source: Franchak, Kretch, Soska, & Adolph (2011)....48

Figure 2.4. Post-hoc calibration procedure using the Tobii TX300 in standalone mode to study infants' naturalistic face scanning patterns. 49

Figure 2.5. Screenshot of the video used to maintain infants' visual attention on-screen while the eye tracker is being set up. 50

Figure 2.6. Screenshots of a calibration point at the start (left) and at the end of the presentation after spiralling inward (right).... 51

Figure 2.7. Examples of Haar-like features....58

Figure 2.8. The upper face region represented as a Haar-like feature. Source: Viola & Jones (2001)....59

Figure 2.9. Screenshot of face detection process. The automatic algorithm attempts to fit a rectangular bounding box around the face, and the user is given the option to accept or reject the outcome.... 61

Figure 2.10. Flowchart visualising the semi-automatic face detection and tracking process. Green and red lines indicate 'yes' and 'no' responses, respectively....62

Figure 2.11. Close up of a randomly selected frame showing the manually coded face region based on pre-defined guidelines....63

Figure 2.12. Screenshot of the video player that visualised the face tracking process...66

Figure 2.13. Screenshot of pushbutton that can be activated at any time during the face tracking period to return to the face detection stage....67

Figure 2.14. Close-up screenshot showing the midline of the bounding box that divides the face area into an upper and a lower region. 69

Figure 2.15. Snippet of coding timeline. 70

Figure 2.16. Data-driven visualisation of face scanning. Source: Caldara & Miellet (2011). 74

Figure 2.17. Interpretation of scanning behaviour can depend on the analysis method. The data-driven method highlights a single gaze cluster between the nose and mouth (in red), whereas the regions-of-interest approach cannot capture the intermediary area between the nose and mouth. Source: Caldara & Miellet (2011). 75

Figure 2.18. A rectangle (in blue) is fitted around the four vertices (in red) of a polygon. 78

Figure 2.19. Rotating the tilted bounding face box by angle α such that the bottom edge (with endpoints (v_{x1}, v_{y1}) and (v_{x2}, v_{y2}) is parallel to the *x*-axis of the scene frame. 79

Figure 2.20. An illustration of the bounding box (in red) within a unified coordinate system, with the origin corresponding to the nose tip. 80

Figure 3.1. Positive Science headset for adults (left) and adapted 'infant' headgear (middle, right). 95

Figure 3.2. Correct automatic feature tracking (left), inaccurate pupil tracking (middle), and pupil centre position after manual correction (right). 98

Figure 3.3. Proportional face dwell time for Western Caucasian and East Asian participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points). 101

Figure 3.4. Proportional upper face dwell time for Western Caucasian and East Asian participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points). 103

Figure 3.5. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for the introduction task, with red and blue colours depicting regions that East Asians and Western Caucasians scanned more, respectively. 104

Figure 3.6. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for the story-telling game, with red and blue colours depicting regions that East Asians and Western Caucasians scanned more, respectively.......... 105

Figure 3.7. Gaze clusters for the introduction task during listening (left) and speaking periods (right), illustrating regions that were scanned significantly more by Western Caucasian (blue) and East Asian participants (red). 106

Figure 3.8. Gaze clusters for the story-telling game during listening (left) and speaking periods (right), illustrating regions that were scanned significantly more by Western Caucasian (blue) and East Asian participants (red). 106

Figure 4.1. Small calibration object. Participants were asked to fixate the image centre. 127

Figure 4.2. Participants' view of the local research assistant in the UK (left) and in Japan (right). 128

Figure 4.3. Participant's view during the three-point calibration procedure. 128

Figure 4.4. Proportional face dwell time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points). 133

Figure 4.5. Proportional face fixation time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points). 135

Figure 4.6. Proportional upper face dwell time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points). 137

Figure 4.7. Proportional upper face fixation time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points). 138

Figure 4.8. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for dwell measures, with red and blue

colours depicting regions that East Asians and Western Caucasians scanned more, respectively........ 140

Figure 4.9. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for fixation measures, with red and blue colours depicting regions that East Asians and Western Caucasians scanned more, respectively........ 140

Figure 4.10. Gaze clusters for listening (left) and speaking periods (right) using dwell data, illustrating regions that were scanned significantly more by British/Irish (blue) or Japanese participants (red). 141

Figure 4.11. Gaze clusters for listening (left) and speaking periods (right) using fixation data, illustrating regions that were scanned significantly more by British/Irish (blue) or Japanese participants (red). 142

Figure 5.1. Hypothesised relationship between culture, executive function, and social development based on previous empirical findings. 157

Figure 5.2. Example stimuli used in the static condition. 165

Figure 5.3. Example stimuli used in the dynamic-neutral condition........ 166

Figure 5.4. Example stimuli used in the dynamic-social condition. 168

Figure 5.5. Screenshots from the gap-overlap paradigm displaying the central stimulus (left) and also the peripheral stimulus (right). 169

Figure 5.6. Screenshots from the cognitive control task displaying the central fixation point (top), the two rectangles during the anticipatory period (bottom left), and the rewarding clip (bottom right). 171

Figure 5.7. Screenshots of a spiral at the start (left) and at the end of the presentation (right). 173

Figure 5.8. Outlines of regions-of-interest superimposed onto a face. 177

Figure 5.9. Proportional dwell times for the eye and bridge region, separately for stimulus condition, face ethnicity, and cultural group. Error bars represent +/- 1 SD. 182

Figure 5.10. Proportional dwell times for the nose and mouth region, separately for stimulus condition, face ethnicity, and cultural group. Error bars represent +/- 1 SD. 183

Figure 5.11. Proportional fixation times for each region-of-interest, stimulus condition, and cultural group. Error bars represent +/- 1 SD. 184

Figure 5.12. Proportional dwell times for the eyes and bridge region, separately for stimulus condition, face ethnicity, and age group. Error bars represent +/- 1 SD....... 188

Figure 5.13. Proportional dwell times for the nose and mouth region, separately for stimulus condition, face ethnicity, and age group. Error bars represent +/- 1 SD....... 189

Figure 5.14. Proportional fixation times for each region-of-interest, stimulus condition, and age group. Error bars represent +/- 1 SD. 191

Figure 6.1. Eye tracking an infant participant for post-hoc data quality checks.......... 215

Figure 6.2. Experimental set-up. The foam boards were removed after successful calibration. 216

Figure 6.3. The infant's point of view: research assistants interacting with infants in Japan (left) and in the UK (right). 217

Figure 6.4. Proportional dwell time (left) and fixation time (right) spent looking at the face. Error bars represent +/-1 SD. 221

Figure 6.5. Proportional dwell time (left) and fixation time (right) spent looking at the upper face. Error bars represent +/-1 SD. 222

Figure 6.6. Heat maps of group differences in gaze density for dwell data (left) and fixation data (right), illustrating regions that were scanned significantly more by British (blue) and Japanese infants (red). 223

Figure 6.7. Gaze cluster map illustrating the findings from the Monte Carlo permutation test. No significant gaze clusters could be found for dwell (left) or fixation (right) measures. 224

List of Tables

Table 3.1. *Medians and interquartile ranges for face dwell time (in %).* 101

Table 3.2. *Medians and interquartile ranges for upper face dwell time (in %).* 102

Table 3.3. *Time spent on mutual gaze proportional to valid recording time (in %).* . .108

Table 3.4. *Means and standard errors for the mean durations of dyad's mutual gaze occurrence (in milliseconds).* 108

Table 3.5. *Pearson's correlation coefficients for the relationship between AQ scores and (upper) face dwell time during speaking and listening periods.* 109

Table 4.1. *Medians and interquartile ranges for face dwell time (in %).* 132

Table 4.2. *Medians and interquartile ranges for face fixation time (in %).* 134

Table 4.3. *Medians and interquartile ranges for upper face dwell time (in %).* 136

Table 4.4. *Medians and interquartile ranges for upper face fixation time (in %).* 138

Table 4.5. *Pearson's correlation coefficients for the relationship between AQ scores and (upper) face dwell time during speaking and listening periods.* 143

Table 4.6. *Pearson's correlation coefficients for the relationship between LSAS scores and (upper) face dwell time during speaking and listening periods.* 144

Table 4.7. *Pearson's correlation coefficients for the relationship between age and (upper) face dwell time during speaking and listening periods.* 145

Table 5.1. *Means and standard deviations of ratings for amount of contact with same-ethnicity and other-ethnicity individuals.* 162

Table 5.2. *Mean spatial offsets and standard deviations (in degrees visual angle) for each experimental group.* 173

Table 5.3. *Parameter settings for automatic processing of gaze data in GraFIX.* 175

Table 5.4. *Means and standard deviations (in brackets) of saccadic latencies (in milliseconds) for the baseline, overlap, and gap conditions.* 195

Table 5.5. *Medians and interquartile ranges (in brackets) of disengagement latencies (in milliseconds) by group and age.* 195

Table 5.6. *Means and standard deviations (in brackets) of proportion of correct anticipatory looks.* 196

Table 10.1. *Pearson's correlation coefficients for the relationship between AQ scores and (upper) face fixation time during speaking and listening periods.* 289

Table 10.2. *Pearson's correlation coefficients for the relationship between LSAS scores and (upper) face fixation time during speaking and listening periods.* 289

Table 10.3. *Pearson's correlation coefficients for the relationship between age and (upper) face fixation time during speaking and listening periods.* 289

Table 13.1. *Means and standard deviations of eye dwell time proportional to inner face looking time of static faces.* 298

Table 13.2. *Means and standard deviations of eye dwell time proportional to inner face looking time of dynamic-neutral faces.* 298

Table 13.3. *Means and standard deviations of eye dwell time proportional to inner face looking time of dynamic-social faces.* 299

Table 13.4. *Means and standard deviations of bridge dwell time proportional to inner face looking time of static faces.* 299

Table 13.5. *Means and standard deviations of bridge dwell time proportional to inner face looking time of dynamic-neutral faces.* 300

Table 13.6. *Means and standard deviations of bridge dwell time proportional to inner face looking time of dynamic-social faces.* 300

Table 13.7. *Means and standard deviations of nose dwell time proportional to inner face looking time of static faces.* 301

Table 13.8. *Means and standard deviations of nose dwell time proportional to inner face looking time of dynamic-neutral faces.* 301

Table 13.9. *Means and standard deviations of nose dwell time proportional to inner face looking time of dynamic-social faces.* 302

Table 13.10. *Means and standard deviations of mouth dwell time proportional to inner face looking time of static faces.* 302

Table 13.11. *Means and standard deviations of mouth dwell time proportional to inner face looking time of dynamic-neutral faces.*..303

Table 13.12. *Means and standard deviations of mouth dwell time proportional to inner face looking time of dynamic-social faces.* 303

Table 13.13. *Summary table of the five-factor ANOVA ROI x Group x Age x Stimulus Condition x Face Ethnicity on dwell time data.*...304

Table 13.14. *Summary table of the four-factor ANOVA ROI x Group x Age x Stimulus Condition for face stimuli of Japanese ethnicity.* 306

Table 13.15. *Summary table of the four-factor ANOVA ROI x Group x Age x Stimulus Condition for face stimuli of White-British ethnicity.* 307

Table 13.16. *Means and standard deviations of eye fixation time proportional to inner face looking time of static faces.*.. 308

Table 13.17. *Means and standard deviations of eye fixation time proportional to inner face looking time of dynamic-neutral faces.*.. 308

Table 13.18. *Means and standard deviations of eye fixation time proportional to inner face looking time of dynamic-social faces.* 309

Table 13.19. *Means and standard deviations of bridge fixation time proportional to inner face looking time of static faces.*...309

Table 13.20. *Means and standard deviations of bridge fixation time proportional to inner face looking time of dynamic-neutral faces.*... 310

Table 13.21. *Means and standard deviations of bridge fixation time proportional to inner face looking time of dynamic-social faces.*.. 310

Table 13.22. *Means and standard deviations of nose fixation time proportional to inner face looking time of static faces.*..311

Table 13.23. *Means and standard deviations of nose fixation time proportional to inner face looking time of dynamic-neutral faces.*...311

Table 13.24. *Means and standard deviations of nose fixation time proportional to inner face looking time of dynamic-social faces.* 312

Table 13.25. *Means and standard deviations of mouth fixation time proportional to inner face looking time of static faces.* 312

Table 13.26. *Means and standard deviations of mouth fixation time proportional to inner face looking time of dynamic-neutral faces.* 313

Table 13.27. *Means and standard deviations of mouth fixation time proportional to inner face looking time of dynamic-social faces.* 313

Table 13.28. *Summary table of the five-factor ANOVA ROI x Group x Age x Stimulus Condition x Face Ethnicity on fixation data.* 314

Table 13.29. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of White-British ethnicity in the static condition.* 316

Table 13.30. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of White-British ethnicity in the dynamic-neutral condition.* 316

Table 13.31. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of White-British ethnicity in the dynamic-social condition.* 316

Table 13.32. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of Japanese ethnicity in the static condition.* 317

Table 13.33. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of Japanese ethnicity in the dynamic-neutral condition.* 317

Table 13.34. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of Japanese ethnicity in the dynamic-social condition.* 317

Table 13.35. *Summary table of independent t-tests contrasting British > Japanese participants in the dwell time measure.* 318

Table 13.36. *Summary table of the two-factor ANOVA ROI x Group for static faces.* 319

Table 13.37. *Summary table of the two-factor ANOVA ROI x Group for dynamic-neutral faces.* 319

Table 13.38. *Summary table of the two-factor ANOVA ROI x Group for dynamic-social faces.* 319

Table 13.39. *Summary table of independent t-tests contrasting British > Japanese participants in the fixation time measure.* 320

Table 13.40. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of White-British ethnicity in the static condition.*..320

Table 13.41. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of White-British ethnicity in the dynamic-neutral condition.*...................................... 321

Table 13.42. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of White-British ethnicity in the dynamic-social condition*... 321

Table 13.43. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of Japanese ethnicity in the static condition.*... 321

Table 13.44. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of Japanese ethnicity in the dynamic-neutral condition.*...322

Table 13.45. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of Japanese ethnicity in the dynamic-social condition.* 322

Table 13.46. *Summary table of the one-way ANOVA with factor Age using dwell time data.* 323

Table 13.47. *Summary table of independent t-tests contrasting each age group for Japanese face stimuli using dwell time data.* 324

Table 13.48. *Summary table of independent t-tests contrasting each age group for White-British face stimuli using dwell time data.* 325

Table 13.49. *Summary table of the two-factor ANOVA ROI x Age for static faces.* ...326

Table 13.50. *Summary table of the two-factor ANOVA ROI x Age for dynamic-neutral faces.* 326

Table 13.51. *Summary table of the two-factor ANOVA ROI x Age for dynamic-social faces.* 326

Table 13.52. *Summary table of the one-way ANOVA with factor Age using fixation time data.* 327

Table 13.53. *Summary table of independent t-tests contrasting each age group*.......328

Table 13.54. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI eyes.* 329

Table 13.55. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI bridge.* 329

Table 13.56. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI nose.*..329

Table 13.57. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI mouth*...330

Table 13.58. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for Japanese participants viewing face stimuli of Japanese ethnicity*............................330

Table 13.59. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for British participants viewing face stimuli of Japanese ethnicity.*...............................330

Table 13.60. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for Japanese participants viewing face stimuli of White-British ethnicity.*..................... 331

Table 13.61. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for British participants viewing face stimuli of White-British ethnicity.*......................... 331

Table 13.62. *Summary table of the one-way ANOVA with factor Stimulus Condition using dwell time data.* 332

Table 13.63. *Summary table of paired t-tests contrasting each stimulus condition: British participants.* 333

Table 13.64. *Summary table of paired t-tests contrasting each stimulus condition: Japanese participants.* 333

Table 13.65. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 10-month-olds viewing face stimuli of Japanese ethnicity.* 334

Table 13.66. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 15- to 17-month-olds viewing face stimuli of Japanese ethnicity*....................................334

Table 13.67. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for adults viewing face stimuli of Japanese ethnicity.* 334

Table 13.68. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 10-month-olds viewing face stimuli of White-British ethnicity*......................................335

Table 13.69. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 15- to 17-month-olds viewing face stimuli of White-British ethnicity*..............................335

Table 13.70. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for adults viewing face stimuli of White-British ethnicity.* 335

Table 13.71. *Summary table of the one-way ANOVA with factor Stimulus Condition using dwell time data.* 336

Table 13.72. *Summary table of paired t-tests contrasting each stimulus condition: 10-month-olds.* 337

Table 13.73. *Summary table of paired t-tests contrasting each stimulus condition: 15- to 17-month-olds.* 337

Table 13.74. *Summary table of paired t-tests contrasting each stimulus condition: Adults.* 338

Table 13.75. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for Japanese participants.* 338

Table 13.76. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for British participants.* 339

Table 13.77. *Summary table of the one-way ANOVA with factor Stimulus Condition using fixation data*...339

Table 13.78. *Summary table of paired t-tests contrasting each stimulus condition: British participants.* 339

Table 13.79. *Summary table of paired t-tests contrasting each stimulus condition: Japanese participants.* 340

Table 13.80. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 10-month-olds.* 340

Table 13.81. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 15- to 17-month-olds.* 340

Table 13.82. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for adults*.. 341

Table 13.83. *Summary table of the one-way ANOVA with factor Stimulus Condition using fixation data*.. 341

Table 13.84. *Summary table of paired t-tests contrasting stimulus conditions: 10-month-olds.* 342

Table 13.85. *Summary table of paired t-tests contrasting stimulus conditions: 15- to 17-month-olds.* 342

Table 13.86. *Summary table of paired t-tests contrasting stimulus conditions: Adults.* 342

Table 14.1. *Pearson's correlation coefficients and corresponding p-values for the relationship between disengagement latencies and proportional mouth looking using dwell time data.* 343

Table 14.2. *Pearson's correlation coefficients and corresponding p-values for the relationship between disengagement latencies and proportional mouth looking using fixation time data.* 345

Table 14.3. *Pearson's correlation coefficients and corresponding p-values for the relationship between cognitive control task performance and proportional mouth looking using dwell time data.* 346

Table 14.4. *Pearson's correlation coefficients and corresponding p-values for the relationship between cognitive control task performance and proportional mouth looking using fixation time data.* 348

List of Abbreviations

ANOVA	Analysis of Variance
AQ	Autism Quotient
ASC	Autism Spectrum Conditions
CDI	Communicative Development Inventory
DCCS	Dimensional Change Card Sort
EEG	Electroencephalogram
FFA	Fusiform face area
fMRI	Functional magnetic resonance imaging
Hz	Hertz
I-VT	Velocity-Threshold Identification
ICC	Intraclass Correlation Coefficient
IQR	Interquartile Range
KLT	Kanade-Lucas-Tomasi
LED	Light-Emitting Diode
LSAS	Liebowitz Social Anxiety Scale
LSAS-SR	Self-report version of Liebowitz Social Anxiety Scale
M	Mean
Mdn	Median
OFA	Occipital face area
RFT	Random Field Theory
ROI	Region-of-interest
SD	Standard Deviation
SE	Standard Error
SMI	SensoMotoric Instruments
SMI ETG	SensoMotoric Instruments Eye Tracking Glasses
STS	Superior temporal sulcus
TFCE	Threshold-Free Cluster Enhancement
VR	Virtual Reality

Chapter 1

General Introduction

1.1. Chapter Overview

To contextualise the empirical studies in this thesis, the present chapter will first briefly discuss the origins of specialised face perception and highlight the importance of postnatal social experience in the development of expert face processing. A review of the literature on cultural influences on face perception – and face scanning in particular – will then be provided to illustrate the significant role of the postnatal environment. It will be concluded that current interpretations are restricted to screen-based paradigms that lack the social presence and visual complexity of 'real-world' social interactions, and the need for naturalistic social interaction paradigms is highlighted with relevant empirical evidence. Current accounts that attempt to explain cultural differences in face scanning will then be outlined before going on to conclude that our understanding of how the postnatal social environment can modulate the development of specialised face perception is still considerably limited. A developmental perspective on cultural differences in face scanning will then be provided and current gaps in the literature will be highlighted. Finally, critical outstanding questions will be outlined to illustrate the aims of this thesis.

1.2. Introduction

The human face represents one of the most important visual stimuli in our everyday life, allowing us to identify others, infer emotional states, and participate in shared attention (Bruce & Young, 1998; Haxby, Hoffman, & Gobbini, 2000; Hoffman & Haxby, 2000). Research has consistently demonstrated an attentional bias for faces compared to other stimuli in our environment, showing that we not only look significantly longer at faces but also rapidly detect them within cluttered scenes (Bindemann, Burton, Hooge, Jenkins, & de Haan, 2005; Bindemann & Lewis, 2013; Johnson, Dziurawiec, Ellis, & Morton, 1991; Langton, Law, Burton, & Schweinberger, 2008; Lewis & Edmonds, 2005; Ro, Russell, & Lavie, 2001; Theeuwes & Van der Stigchel, 2006). As a result, it has been suggested that faces are processed qualitatively differently compared to objects (Farah, Wilson, Maxwell Drain, & Tanaka, 1995; Valentine, 1988). This claim is corroborated by neuropsychological studies involving prosopagnosic patients who show impairments in the recognition of familiar faces but not objects (for a review see Farah, 1996). Electroencephalography (EEG) studies have additionally revealed the N170 component that can be elicited by faces, but not other objects, suggesting face-selective responses at a neural level (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Eimer, 2000). Functional magnetic resonance imaging (fMRI) studies have also identified several cortical regions involved in typical face perception, demonstrating increased neural activity during face processing tasks, including the lateral fusiform gyrus (or *fusiform face area*, FFA; Kanwisher, McDermott, & Chun, 1997), the superior temporal sulcus (STS; Hoffman & Haxby, 2000), and the inferior occipital gyrus (or *occipital face area,* OFA; Haxby et al., 2000). Face processing also requires secondary functions involving, for instance, the auditory cortex and the amygdala for speech and emotion processing, respectively (Haxby et al., 2000). Altogether, this points to an extensive neural network with various functional components for specialised face processing.

Given this functional network of neural substrates, there has been a long-standing debate about the origins of such perceptual and neural specialisation. On the one hand, it has been proposed that such functionally distinct brain regions imply the existence of innate, biologically-determined face 'modules' (Farah, Rabinowitz, Quinn, & Liu, 2000; Kanwisher et al., 1997). On the other hand, it has been demonstrated that relevant brain areas such as the FFA show increased activation in response to not only faces, but also artificial non-face stimuli (*Greebles*) after participants had acquired sufficient expertise with this stimulus category (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). This consequently suggests an experience-dependent mechanism for the emergence of a specialised face processing system (Nelson, 2001). Findings from developmental studies support this experience-dependent account. In particular, face processing has been shown to become more specialised with age (Cohen Kadosh, 2011; Di Giorgio, Turati, Altoè, & Simion, 2012; Frank, Amso, & Johnson, 2014; Frank, Vul, &

Johnson, 2009; de Haan, Pascalis, & Johnson, 2002), with improvements in behavioural proficiency linked to changes in the neural organisation involving the face-specific areas (Aylward et al., 2005; Peelen, Glaser, Vuilleumier, & Eliez, 2009). Altogether, evidence points to the importance of postnatal developmental processes in the emergence of specialised face perception skills (Cohen Kadosh, 2011; Morton & Johnson, 1991).

An experience-dependent account suggests that face perception is subject to postnatal environmental influences, and is therefore malleable during its development. Le Grand, Mondloch, Maurer and Brent (2003) demonstrated that individuals who were visually deprived during infancy due to a congenital cataract in the left eye (i.e., visual deprivation affected input to the right hemisphere where face processing is predominantly manifested; Kanwisher et al., 1997; G. McCarthy, Puce, Gore, & Allison, 1997; Rossion et al., 2000) did not achieve expert face perception. Recent studies have also examined postnatal modulations on face perception by exploiting naturally occurring variations in environmental influences, such as cultural or ethnic diversity. For instance, developmental studies have demonstrated that, over the course of the first year of life, face recognition abilities become increasingly fine-tuned to familiar "own-race" faces, but not unfamiliar "other-race" faces (Anzures, Pascalis, Quinn, Slater, & Lee, 2011; Anzures et al., 2013; Kelly et al., 2009, 2007) – a phenomenon coined *other-race effect* (Malpass & Kravitz, 1969; see Meissner & Brigham, 2001 for review). Such a process of *perceptual narrowing* (Nelson, 2001) reflects an adaptive mechanism for fine-tuning to familiar, socially relevant information.

Although most research has typically assumed that underlying processes are culturally invariant (Han & Northoff, 2008), the notion of perceptual narrowing suggests that postnatal cultural experience modulates the development of expert face processing. Indeed, eye tracking studies have recently revealed that visual strategies to achieve face perception tasks differ between cultural groups, pointing to significant influences of postnatal social experience on face scanning (e.g., Blais, Jack, Scheepers, Fiset, &

Caldara, 2008). In the following, a more detailed review of the literature on cultural differences in face scanning will be provided.

1.2.1. Cultural differences in face scanning

In their seminal study, Blais et al. (2008) examined face scanning strategies by recording eye movements of Western Caucasian and East Asian participants who were asked to learn and recognise identities with neutral facial expressions. Findings revealed that Western Caucasians exhibited greater triangular scanning of the eyes and the mouth, whereas East Asians showed more fixations on central face regions (i.e., the nose; see Figure 1.1). Crucially, the groups did not differ in their face recognition performance, indicating cultural modulations on *effective* visual strategies to extract task-relevant information. These eye movement patterns have also since been replicated in other studies examining face recognition (Kelly, Liu, et al., 2011; Kelly et al., 2010; Kita et al., 2010; Rodger, Kelly, Blais, & Caldara, 2010).

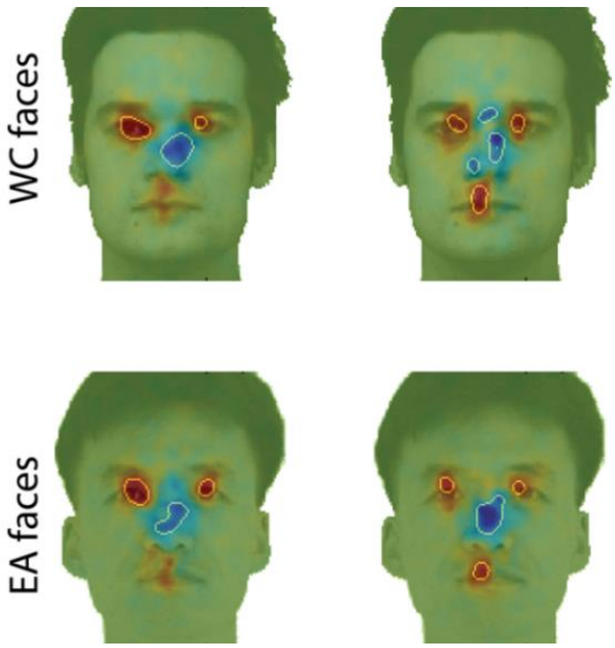


Figure 1.1. Western Caucasians showed greater triangular scanning of the eyes and mouth (red), whereas East Asians fixated more on the face centre (blue). Source: Blais et al. (2008).

Cultural differences in scanning strategies also manifested when participants were observing emotionally expressive face stimuli (Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Jack, Caldara, & Schyns, 2012; Senju, Vernetti, Kikuchi, Akechi, & Hasegawa, 2013; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). When categorising faces by emotional expression, Western Caucasian adults showed significantly more mouth looking while East Asians exhibited increased scanning of the eye region (Jack et al., 2009). These patterns were also observed in a task-free paradigm (i.e., *free-viewing paradigm;* e.g., Adolphs et al., 2005; Chua, Boland, & Nisbett, 2005; Frank, Vul, & Saxe, 2012) that used dynamic videos of faces (Senju, Vernetti, Kikuchi, Akechi, & Hasegawa, 2013; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). Although it remains unclear whether the differences in scanning strategies between neutral and emotionally expressive face stimuli were due to differences in stimulus characteristics or task demand (face recognition versus emotion categorisation or free-viewing), the findings altogether point to significant cultural influences on eye movement strategies across different face processing tasks. A more detailed discussion on proposed explanations that account for such cultural differences in face scanning will be provided below in Section 1.2.3.

1.2.2. Face perception in the 'real world' and the relevance for cross-cultural comparisons

Empirical evidence on cultural differences in face scanning – and on face perception in general – is currently predominantly restricted to findings from screen-based studies that employed static images of faces. In such highly controlled lab environments, retinal stimulation is constrained by removing task-irrelevant visual distractors, turning off lights, presenting stimuli on a computer screen that is directly positioned in front of participants, and restricting natural head and body movements (e.g., with the use of a head- or chinrest; Holmqvist et al., 2011). Stimulus size is also smaller than the faces that are encountered in everyday life (Holmqvist et al., 2011), and participants do not interact with the presented faces. However, the faces that we typically see are dynamic, and face processing is particularly relevant during human social interactions. Recent studies on face perception have indicated that face looking behaviour manifests differently within more naturalistic settings (Foulsham, Walker, & Kingstone, 2011; Freeth, Foulsham, & Kingstone, 2013; Laidlaw, Foulsham, Kuhn, & Kingstone, 2011; D. W.-L. Wu, Bischof, & Kingstone, 2013, 2014), pointing to the need to study cultural differences in face scanning during social interactions. In the following, the differences between screen-based and 'real-world' settings will be outlined in more detail, and relevant studies will be presented to illustrate that interpretations of cross-cultural findings cannot be limited to screen-based studies, but also require naturalistic paradigms.

First, static face images displayed on a two-dimensional screen lack dynamic features and environmental distractors, both of which are common to everyday situations. Visual saliency such as low-level motion can predict gaze locations during dynamic scene viewing (Itti, 2005; Mital, Smith, Hill, & Henderson, 2010), raising the possibility that more salient facial features (e.g., a moving mouth during speech) could modulate face scanning behaviour. In addition, studies examining processes underlying emotional facial expressions have suggested that, compared to static displays, dynamic features can result in improved performance when judging expressions (Ambadar, Schooler, & Cohn, 2005), and lead to different neural activations (Kilts, Egan, Gideon, Ely, & Hoffman, 2003) and fewer fixations on facial features during emotion recognition (Blais, Fiset, Roy, Saumure Régimbald, & Gosselin, 2017). With respect to cultural differences, Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al. (2013) presented participants with dynamic videos depicting a smiling actor, and found that East Asians fixated the eyes, whereas Western Caucasians looked more at the mouth, replicating the findings by Jack et al. (2009) who employed static stimuli. Although this suggests that dynamic features do not necessarily override cultural effects, this would need to be confirmed within more naturalistic paradigms.

Secondly, participants in screen-based studies do not interact with the faces, and this lack of any social presence disregards the influence of sociocultural norms that may affect face looking behaviour (Foulsham et al., 2011; Gobel, Kim, & Richardson, 2015; Laidlaw et al., 2011; D. W.-L. Wu et al., 2013, 2014). For instance, while participants can look at face images on screen for as long as they wish, excessive gazing at strangers in everyday situations can be considered inappropriate. More recently, a greater emphasis has therefore been placed on studying social attention within a social context (Kingstone, 2009; Kingstone, Smilek, & Eastwood, 2008; Richardson & Gobel, 2015; Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012; Risko, Richardson, & Kingstone, 2016), given that eye movements within naturalistic settings not only serve to extract relevant visual information (the *encoding function*), but also signal one's mental state to another person (the *signalling function*; Gobel et al., 2015; Risko et al., 2016). Indeed, visual attention to faces has been shown to decrease significantly when participants sat in the same room with another individual, compared to when participants saw a videotape of the same individual (Laidlaw et al., 2011). Furthermore, Foulsham et al. (2011) found that passers-by close to the participant were fixated less frequently in a 'real-world' setting than on screen, suggesting that social norms can influence face orienting behaviour. With respect to face scanning, Gobel et al. (2015) examined the influence of social rank by manipulating participants' belief about whether or not they could later be seen by the face targets that were presented on screen. When participants believed that they would later be seen by a target of higher social rank, they fixated the eyes less compared to instances when participants believed they would not be seen. For targets of low social rank, the reverse pattern was observed. These findings therefore indicate that the mere belief of social presence, coupled with existing sociocultural norms, can modulate face scanning behaviour. In a cross-cultural study, Gobel, Chen and Richardson (2017) manipulated the perceived interpersonal context by presenting Western Caucasian and East Asian participants with dynamic faces on screen that either engaged in direct eye contact (i.e., looked directly into the camera) or gazed away. Cultural differences in face scanning were only observed in the eye contact condition, with Western Caucasians showing increased fixations on the eye region whereas East Asians exhibited central face scanning of the bridge and nose. Compared to the eye contact condition, East Asians tended to look less at the nose when gaze was averted, while Western Caucasians showed fewer fixations on the eyes and more on the nose, and no group differences were found. These scanning patterns were suggested to reflect the sociocultural norm of gaze avoidance in East Asian cultures that acts as a sign of respect (Sue & Sue, 1990; discussed in more detail in Section 1.2.3 below).

Altogether, the results suggest that sociocultural norms, and social presence more generally, can influence face scanning strategies, and this can manifest differently between cultures. To date, however, this has not been directly examined within a social interaction paradigm. Given that ultimately the goal is to understand face perception processes in everyday life – as opposed to within a lab-based environment – naturalistic social interaction paradigms are required to determine the extent to which current evidence generalises to 'real-world' settings. In addition, identifying the factors that modulate face scanning behaviour within naturalistic settings will lead to an improved understanding of face processing more generally. To address this gap in the literature, this thesis will present two face-to-face interaction studies with adults (Chapter 3 and Chapter 4) that will investigate cultural differences in naturalistic face scanning. The findings from the two experiments will provide a more detailed description of the manifestation of cultural differences in face scanning and allow the formulation of novel predictions that can explain how postnatal social experience modulates face perception. As outlined in the following section, underlying mechanisms that could explain how culture influences face scanning remain largely unknown. The current literature continues to describe face scanning behaviour, but how postnatal social experiences interact with the development of specialised face processing is unclear.

1.2.3. Explaining cultural differences in face scanning

An explanation that has been initially suggested to account for the triangular versus central face scanning patterns of Western Caucasians and East Asians, respectively, is based on sociocultural norms in eye contact behaviour (Argyle, Henderson, Bond, Iizuka, & Contarello, 1986; Blais et al., 2008; Gobel et al., 2017; McCarthy, Lee, Itakura, & Muir, 2006, 2008). In particular, previous studies have reported an avoidance of prolonged eye contact in East Asians (Argyle et al., 1986; A. McCarthy et al., 2006, 2008), which has been suggested to function as a sign of respect (Sue & Sue, 1990). Argyle et al. (1986) employed self-report measures to show that British and Italian participants rated the importance of maintaining eye contact during conversations higher than individuals from Japan or Hong Kong. Furthermore, A. McCarthy et al. (2006, 2008) examined whether reduced eye contact actually manifests in East Asians by analysing video recordings of participants during thinking processes (e.g., solving abstract mathematical tasks), and showed that Japanese individuals indeed engaged in less eye contact than Canadians and Trinidadians. The central face scanning bias observed in East Asians has therefore been proposed to reflect the sociocultural norm of gaze avoidance – by focusing on the nose region, direct eye contact can be avoided and relevant information from the eye region can still be derived using extrafoveal vision (Miellet, He, Zhou, Lao, & Caldara, 2012; Miellet, Vizioli, He, Zhou, & Caldara, 2013). However, current evidence on reduced eye contact in East Asians is based on self-reports (Argyle et al., 1986) and video recordings of participants engaged in abstract thinking (A. McCarthy et al., 2006, 2008); no quantitative evidence currently exists in support of gaze avoidance within naturalistic face-to-face interactions. Crucially, the notion of reduced eye contact contradicts the finding demonstrating that East Asians predominantly scan the eye region of emotionally expressive faces. In addition, as will be discussed below, early cultural differences in face scanning can also be observed during the first year of life (e.g., Geangu et al., 2016), and it is unlikely that sociocultural norms can already modulate scanning strategies in infancy (Tomasello, Kruger, & Ratner, 1993).

Another proposed explanation is based on cultural differences in perceptual styles more generally (e.g., Blais et al., 2008; for a review see Nisbett & Masuda, 2003; Nisbett & Miyamoto, 2005). In particular, cross-cultural studies on scene perception

(Chua et al., 2005; Masuda & Nisbett, 2006) and scene description (Masuda & Nisbett, 2001) revealed that Western Caucasian individuals tended to extract focal information (*analytic* perception), whereas East Asians attended to visual stimuli as a whole (*holistic* perception). This has been proposed to arise from cultural differences in societal construal, with Western cultures placing greater importance on individualism and independence from others, while East Asian societies are more collectivist and value harmonious interdependence (Greenfield, Keller, Fuligni, & Maynard, 2003; Markus & Kitayama, 1991). Nisbett and Masuda (2003) consequently suggested a "causal chain running from social structure to social practice to attention and perception to cognition" (p.11170). With respect to face scanning strategies, an account based on cultural differences in perceptual styles is consistent with Western Caucasians focusing on individual facial features and East Asians fixating centrally. Given that critical visual information required for successful face recognition has been shown to be predominantly located in the eyes, and to a lesser extent also in the mouth region (e.g., Gosselin & Schyns, 2001; Henderson, Williams, & Falk, 2005; Schyns, Bonnar, & Gosselin, 2002; Williams & Henderson, 2007), it has been suggested that East Asians may use their extrafoveal vision more effectively, thereby allowing them to extract relevant visual information by fixating the nose (Caldara, 2017; Miellet et al., 2012, 2013). This was supported by studies that either obscured central vision (*Blindspot technique;* Miellet et al., 2012; Miellet, Zhou, He, Rodger, & Caldara, 2010; Rayner & Bertera, 1979) or peripheral vision in a gaze-contingent manner (*Spotlight technique;* Caldara, Zhou, & Miellet, 2010). When central vision was masked in a face recognition task, Western Caucasians tended to shift their gaze to the face centre; however, no changes in scanning strategy was observed for East Asians, suggesting greater use of extrafoveal vision for face recognition (Miellet et al., 2012). Similarly, when peripheral vision was masked, both Western Caucasians and East Asians fixated the eyes and mouth when visual information from the face was highly constrained (2° or 5°). However, a shift toward a central fixation bias could be observed for East Asians when both the eyes and mouth were visible (at 8°; Caldara et al., 2010). With respect to scanning strategies for emotionally expressive faces, increased scanning of the eye region such as that observed in East Asians does not necessarily contradict the notion of greater use of extrafoveal vision. In particular, Jack et al. (2012) used computational modelling techniques to show that East Asians represented the intensity of emotional expressions with movements of the eyes, whereas Western Caucasians used other face regions. If visually informative regions are not spatially distributed across the face, but restricted to local areas, extrafoveal vision may thus not be required to extract the information to achieve the relevant face processing task. East Asians may therefore fixate the eye region directly rather than deploy central face scanning strategies that allow extracting visual information from multiple facial features.

Although cultural differences in perceptual styles can account for the observed eye movement patterns when viewing both neutral and emotionally expressive faces, several outstanding questions remain. First, participants were engaged in a face recognition task (e.g., Blais et al., 2008) or were asked to categorise emotionally expressive faces (Jack et al., 2009; but see Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). Having an explicit task may have influenced scanning strategies in a top-down fashion (Yarbus, 1967); however, as outlined above (Section 1.2.2), gaze behaviour in naturalistic social settings also involves a signalling function. An account solely based on visual strategies (the encoding function) therefore unlikely explains cultural differences in naturalistic face scanning. Secondly, several studies have implied the possibility that viewing patterns not only depend on the participant's cultural background, but are also modulated by the ethnicity of the face shown (Fu, Hu, Wang, Quinn, & Lee, 2012; Goldinger, He, & Papesh, 2009; Wheeler et al., 2011; E. X. W. Wu, Laeng, & Magnussen, 2012; Xiao, Xiao, Quinn, Anzures, & Lee, 2013). An account based on cultural differences in perceptual styles cannot, however, accommodate such findings. Finally – and crucially – it remains unclear why cultural differences in perceptual style exist in the first place.

A study that examined potential sources that give rise to cultural differences in attentional strategies suggested a role of the physical environment (Miyamoto, Nisbett, & Masuda, 2006). In particular, scenes of small, medium, and large Japanese cities were found to contain more elements than those of American cities. When priming participants with Japanese scenes, both Japanese and American individuals exhibited a more holistic attentional strategy. Conversely, those primed with American scenes attended more to focal objects. In other words, visual experience with the postnatal environment such as the physical infrastructure may modulate attentional strategies and differentially affect cultural groups. With respect to face scanning, Geangu et al. (2016) suggested that visual experience with the caregiver's facial expressions may provide an early source for cultural differences. Geangu et al. (2016) revealed cultural differences in scanning strategies already at 7 months of age (a more detailed discussion will be provided in Section 1.2.4 below) – Japanese infants showed more scanning of the eye region and British infants demonstrated more mouth looking for emotionally expressive face stimuli – and proposed that this may have resulted from differences in the facial expressivity of caregivers. For instance, East Asian mothers have previously been found to show less emotional expressivity than Western Caucasian mothers (Fogel, Toda, & Kawai, 1988), and the locations of visually informative regions for emotional expressions may therefore differ between cultures – a suggestion that converges with the results by Jack et al. (2012) who found the visually informative regions to be located in the eye region for East Asians and in the mouth area for Western Caucasians. Geangu et al. (2016) also suggested that parental practices throughout childhood may additionally reinforce learned attentional strategies. For instance, it is possible that more distinct and informative mouth movements used in the facial expressions of Western Caucasian compared to East Asian parents could reinforce greater mouth looking. This account thereby considers a developmental mechanism that could give rise to the observed cultural differences in face scanning, providing insight also into possible sources for cultural learning. Although it remains unclear how infants may gradually adopt increasingly culturally-relevant scanning strategies, it is possible that the development of more advanced executive function in the first year of life (Diamond, 1985; Diamond, 2002; Rose, Feldman, & Jankowski, 2012) could allow infants to exert greater top-down influence on visual attention. Such developmental improvements in attentional control could, in turn, help infants to visually disengage from irrelevant features (e.g., irrelevant salient features such as mouth movements) and attend to more informative regions (e.g., the eyes during emotional displays). Altogether, this suggests that developmental processes are likely involved in the emergence of cultural differences in face scanning, but – as outlined in more detail in the following section – very little is currently known about the developmental trajectory of scanning strategies. To better understand cultural differences in face scanning, the following section will now provide a developmental perspective.

1.2.4. Cultural differences in the development of face scanning

Although evidence points to postnatal developmental processes for cultural differences in face scanning, previous studies largely investigated eye movement patterns of adults, and little is known about the underlying developmental trajectory. To date, only two cross-cultural studies on face scanning have been conducted with infants or young children. Geangu et al. (2016) showed that 7-month-old British and Japanese infants exhibited scanning strategies consistent with those of adults when viewing static, emotionally expressive faces presented on screen: whereas British infants showed more looking at the mouth, Japanese infants fixated the eye region. These results are also supported by findings from Senju, Vernetti, Kikuchi, Akechi and Hasegawa (2013) who used dynamic, emotionally expressive stimuli, and revealed similar patterns in British and Japanese children aged between 1 and 7 years old. In addition to these two studies that examined face scanning during infancy and early childhood, Kelly, Liu, et al. (2011) also studied the developmental trajectory for scanning strategies of British and Chinese 7- to 12-year-olds using static stimuli depicting faces with neutral expressions, and found that viewing patterns corresponded with those of adults from the respective culture. They also revealed that this correspondence strengthened with age, again pointing to developmental influences on specialised face perception.

Several additional studies also presented infants with dynamic, neutral faces; however, given that the primary aim of these studies was to examine perceptual narrowing (cf., Nelson, 2001) only one cultural group was tested. These studies demonstrated that, with increasing age, 6- to 10-month-old Western Caucasian infants looked longer at the eyes and less at the mouth of faces from their own ethnicity, while scanning for faces of unfamiliar (Black) ethnicity remained similar across age groups (Wheeler et al., 2011; Xiao et al., 2013). Chinese 4- to 9-month-old infants showed fewer fixations to internal facial features (particularly the nose) of unfamiliar, dynamic Western Caucasian face stimuli with neutral expression (Liu et al., 2011). Unlike the studies by Geangu et al. (2016) or Senju, Vernetti, Kikuchi, Akechi and Hasegawa (2013), these infant studies suggested that viewing patterns were also modulated by the ethnicity of face stimuli. However, since cultural groups were studied in isolation, it is difficult to draw conclusions about any cultural differences in the development of face scanning.

Altogether, existing developmental studies point to a consistent finding that cultural differences in scanning strategies for faces likely emerge by 10 months of age, with face ethnicity possibly modulating scanning patterns. However, only one cross-cultural study to date has examined face scanning behaviour in the first year of life (cf., Geangu et al., 2016), and several outstanding questions remain. First, the study examined face scanning in a single age group only (at 7 months), but mapping the developmental trajectory will require a cross-sectional or longitudinal design. Secondly, given that stimulus characteristics affect scanning strategies, it remains unclear how cultural differences manifest when infants scan dynamic and/or neutral faces. Finally, as discussed in Section 1.2.2, the goal is ultimately to understand face perception processes as they occur in the 'real world'; however, to date no study has examined face scanning behaviour of infants during social interactions. Adopting a developmental framework will help identify the time course of emerging cultural differences in face scanning, which in turn will cast a light on possible mechanisms that can explain how postnatal social experience modulates the development of specialised face perception.

1.2.5. Outstanding questions and thesis aims

This chapter presented previous empirical work investigating cultural differences in face scanning and emphasised the need to examine face perception within naturalistic social interactions. The importance of conducting cross-cultural infant studies – both screen-based and within naturalistic settings – to characterise the developmental trajectory of cultural differences in face scanning was also discussed. Altogether, this would allow insight into possible mechanisms that can explain how postnatal social experience may affect the development of specialised face processing.

This thesis will present a series of eye tracking studies which will investigate cultural differences in adults' and infants' face scanning behaviour within screen-based and naturalistic settings. Before introducing the empirical studies of this thesis, Chapter 2 will first briefly introduce some general principles of eye tracking and discuss the methodological challenges associated with mobile eye tracking, a technique that would permit the study of face-to-face interactions. Specifically, the lack of cross-cultural face-to-face interaction studies that adopt eye tracking techniques is likely in part a reflection of the significant technical and practical challenges underlying data pre-processing and analysis. To enable a series of social interaction studies using eye tracking techniques, Chapter 2 will therefore introduce a semi-automatic computational tool that can dynamically track faces, classify gaze points, and conduct traditional regions-of-interest analysis as well as a novel data-driven analysis. Chapter 3 will then present a face-to-face interaction study with adults by adopting dual head-mounted eye tracking techniques in Western Caucasian and East Asian dyads in order to establish whether and how cultural differences in face scanning manifest during live dyadic interactions. Chapter 4 will present an adapted face-to-face interaction study with methodological improvements to clarify interpretations that emerged from the findings in Chapter 3. Having established that cultural differences in adults' face scanning also manifest during social interactions, Chapter 5 will then describe a cross-sectional, cross-cultural screen-based eye tracking study that examines the developmental trajectory of face scanning in British and Japanese infants and adults. The study will also explore the role of executive function in the emergence of culture-specific face scanning to identify possible mechanisms that can generate new predictions for how the postnatal social environment shapes face perception. To investigate the extent to which infants' face scanning strategies generalise to naturalistic social contexts, Chapter 6 will describe and compare face scanning strategies of British and Japanese 10-month-old infants engaged in face-to-face interaction play. Finally, Chapter 7 will discuss findings from all the empirical studies presented in this thesis and consider the theoretical implications within the wider literature.

Chapter 2

Methods Development for ‘Real-World’ Eye Tracking

The statistical data-driven analysis (Monte Carlo permutation test) introduced in Section 2.3.4.2 was implemented in collaboration with and with major contribution of Dr Raffaele Tucciarelli.

2.1. Chapter Overview

To provide an overview of the methodologies used for the studies in this thesis, the present chapter will first introduce some general principles of eye tracking techniques. Given that this thesis examined cultural influences on infants' and adults' face scanning strategies within both screen-based and naturalistic settings, each study required a different eye tracker to appropriately address the relevant research question. An overview of available eye tracking systems will therefore be provided, outlining also the rationale behind the model selections for this thesis and describing associated calibration and data pre-processing procedures. The focus of this chapter, however, will be to present novel pre-processing and analysis methods for 'real-world' eye tracking data. As outlined in Chapter 1, current evidence on cultural differences in face scanning is largely restricted to screen-based eye tracking paradigms, likely due to the methodological challenges underlying mobile eye tracking. To enable a series of social interaction studies, this chapter will present a novel semi-automatic gaze annotation tool and, for the first time, introduce a data-driven analysis method for 'real-world' eye tracking data. These computational methods were applied to the face-to-face interaction studies in this thesis (Chapters 3, 4, and 6).

2.2. Introduction

2.2.1. General principles of eye tracking

Previous studies examined face looking behaviour by manually coding video recordings of participants' gaze locations based on estimations of head and eye positions (e.g., A. McCarthy et al., 2006, 2008). In the developmental literature, studies also adopted habituation paradigms whereby infants were first repeatedly presented with the same face stimulus until looking time decreased below a specified threshold, indicating that infants habituated to the stimulus. They were then presented with the familiar face as well as a novel face, and total looking time to each stimulus was recorded. If infants scanned the novel face longer, recognition would be inferred (Aslin, 2007). Such habituation paradigms have been adopted to examine the emergence of the 'other-race effect' across the first year of life (e.g., Kelly et al., 2009, 2007).

Although these approaches provided first insights into cultural modulations on face perception, these methods suffer from poor spatiotemporal resolution and also involve manual coding that is often subjective (Aslin & McMurray, 2004). However, with the advent of more advanced technology, precise and objective measurements of gaze locations can now be recorded using eye tracking techniques. Most contemporary systems compute gaze locations using the *pupil-corneal reflection technique*, with eye trackers emitting near infra-red light to create a reflection on the cornea and measuring the relative distance to the pupil centre (Duchowski, 2007; Gredebäck, Johnson, & von Hofsten, 2009; Holmqvist et al., 2011). Currently available eye trackers can collect gaze samples ranging from 30 Hz (i.e., data samples per second; e.g., Positive Science, www.positivescience.com) to 2000 Hz (e.g., EyeLink 1000 Plus, SR Research Ltd) and with a spatial accuracy as low as 0.25° (i.e., degree visual angle) to 0.50°. However, the selection of an appropriate eye tracker cannot be solely based on technical properties such as the spatiotemporal resolution, but also requires consideration of the study paradigm and population (Holmqvist et al., 2011). The next section will introduce different types of eye tracking systems and will outline some relevant advantages and disadvantages in order to illustrate the rationale behind the model selections for this thesis.

2.2.2. Differences in eye tracking systems and model selections for the present thesis

Historically, eye trackers were mechanical and therefore intrusive, requiring a coil to be attached to the anaesthetised eyeball (Delabarre, 1898). Although such systems are still available, most modern eye trackers rely on the non-intrusive pupil-corneal reflection technique (Duchowski, 2007; Gredebäck et al., 2009; Holmqvist et al., 2011). Generally, such systems can be divided into *static eye trackers* (e.g., SMI Red, Tobii TX300, EyeLink 1000) and *head-mounted eye trackers* (e.g., Positive Science, Tobii Glasses, Pupil Labs, SMI glasses; Holmqvist et al., 2011). Static systems are employed mainly for screen-based paradigms within laboratory settings and are far more commonly used than head-mounted systems given the challenges inherent to mobile eye tracking data (a more detailed discussion will be provided below in Section 2.3). Depending on the specific eye tracker model, static systems may require the use of a head- or chinrest (*tower-mounted eye trackers*; e.g., EyeLink 1000; SR Research Ltd), thereby ensuring high-quality data by keeping participants' heads fixed in position (Holmqvist et al., 2011). In other words, such tower-mounted eye trackers do not tolerate head movements and thus cannot be applied to study eye movements during naturalistic social interactions, nor can they be used in infants who cannot comply to verbal instructions and therefore do not sit still (Aslin & McMurray, 2004). This type of eye tracker was therefore not used for the work in this thesis.

A static system that tolerates head movements and can thus be used for infant populations is based on *remote eye tracking* (Holmqvist et al., 2011; e.g., SMI Red, Tobii TX300). Typically, remote eye trackers are attached underneath a monitor that displays the experimental stimuli providing a suitable method to record eye movements for screen-based paradigms in both infants and adults. For this reason, the remote eye tracker Tobii TX300 (Tobii Technology, Sweden) was used in the cross-sectional screen-based study in Chapter 5. As will be discussed below, the monitor can also be removed if this provides a better approach to address the relevant research question.

However, screen-based remote systems cannot be employed to study face scanning behaviour during naturalistic social interactions that allow free head and body movements; head-mounted eye trackers are instead required. For adults, head-mounted eye trackers typically involve glasses that are fit with near infra-red light emitting LEDs, a small camera that records the eye, and a scene camera that captures the participant's point-of-view. The eye tracking glasses are connected to a portable recording unit or directly to the computer in order to save the gaze data along with the scene camera videos. If two head-mounted systems are available, dual eye tracking can also be performed (i.e., two participants can be recorded simultaneously). Although some manufacturers (e.g., Pupil Labs) provide software that can automatically synchronise two separate recordings, most eye tracking systems require manual synchronisation. This can be achieved by, for instance, using a clapperboard within the dyad's field-of-view and taking the moment of clap as the mutual starting time (e.g., Ho, Foulsham, & Kingstone, 2015; see also Chapter 3). Head-mounted eye tracking glasses therefore allow the study of cross-cultural differences in adults' face scanning strategies during naturalistic social interactions, and were used for the studies in Chapters 3 and 4. In particular, two Positive Science headgears were used for the dual eye tracking study in Chapter 3 (see Figure 2.1 on page 47). This system was selected since two sets were available – a practical consideration – and since this model only minimally obstructs the face region compared to other available systems – an important factor that must be taken into account when investigating face scanning behaviour.

Figure 2.1. Participant wearing the Positive Science head-mounted eye tracker.

However, as will be discussed in more detail in Chapters 3 and 4, findings from the dual eye tracking study revealed some methodological limitations relating to the use of the Positive Science gear; for instance, although having considered the extent to which this system may obstruct the face region, results suggested the possibility that participants were nevertheless visually distracted by the eye camera arm. The study in Chapter 4 therefore recorded eye movements of only one person within each dyad, and the SMI eye tracking glasses (SMI ETG; SensoMotoric Instruments, Germany; see Figure 2.2) were employed given their superior spatiotemporal resolution relative to the Positive Science system (see Chapter 4 for more details).

Figure 2.2. Adult wearing the SMI eye tracking glasses.

With respect to developmental populations, only one head-mounted eye tracker currently exists for infants (Positive Science). This system requires infants to wear a beanie before the eye tracker can be attached using hook-and-loop fasteners (see Figure 2.3). This method has been successfully used for developmental studies examining locomotor development (Kretch, Franchak, & Adolph, 2014) or sustained and joint attention in infants (Yu & Smith, 2013).

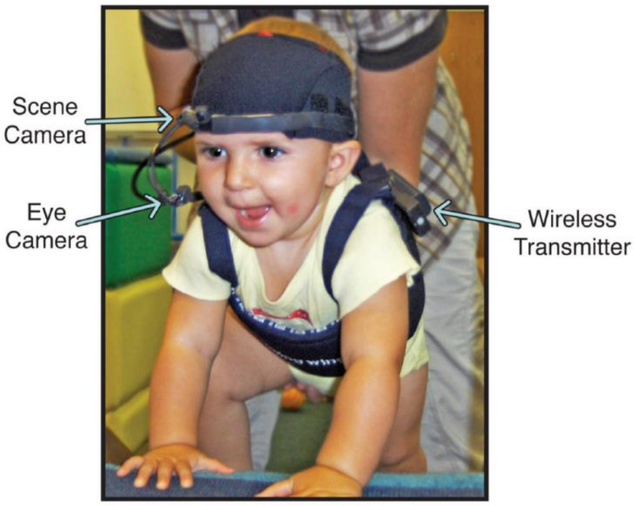


Figure 2.3. Infant wearing the Positive Science eye tracker. Source: Franchak, Kretch, Soska, & Adolph (2011).

Although the Positive Science headset would in theory provide a suitable technique to study cross-cultural differences in infants' face scanning during naturalistic social interactions, the piloting stage for the face-to-face interaction study in Chapter 6 resulted in a 72.73% (8 of 11) drop-out rate since infants did not tolerate the beanie or eye camera arm. In contrast to previous studies that successfully adopted head-mounted eye tracking, infants in the pilot study were younger and therefore less likely to tolerate the equipment. Crucially, infants in the pilot study were not given any engaging toys as distractors, which would have kept their hands away from the eye camera arm. This was implemented given that the presence of toys would significantly reduce face looking time (Yu & Smith, 2013) such that scanning behaviours could not be examined sufficiently. In light of the high drop-out rate, the Positive Science system was not chosen and instead the remote eye tracker Tobii TX300 (Tobii Technology, Sweden) was selected and used in standalone mode. In particular, although the TX300 represents a static system, the monitor attached to the eye tracker can be removed (see Figure 2.4; see also Chapter 6 for more details on the experimental set-up). A person can thus sit behind the eye tracker and naturally interact with the infant, providing a suitable method to study face-to-face interactions that infants can tolerate.

Figure 2.4. Post-hoc calibration procedure using the Tobii TX300 in standalone mode to study infants' naturalistic face scanning patterns.

2.2.3. Calibration

Prior to the start of any eye tracking study, each participant is required to complete a calibration procedure. This typically involves asking the participant to fixate a number of points that are presented sequentially. To ensure precise gaze measurements, the calibration points should be distributed across and shown at the same level as the stimulus area (e.g., the screen, or the face region for dyadic interactions; Holmqvist et al., 2011). The relevant eye tracking software can then fit a function to calculate gaze locations based on calibration information such as eye shape, position, distance relative to the tracker, or luminance in the testing room (Duchowski, 2007; Holmqvist et al., 2011).

In screen-based paradigms, the calibration points are presented in a grid that typically includes five or nine points (Holmqvist et al., 2011). Unlike adults, however, developmental populations do not comply with experimental instructions and additional methods are required when calibrating infants (Gredebäck et al., 2009). Given that the data quality of gaze recordings is highly dependent on the initial calibration performance (Holmqvist et al., 2011), it is important to set up the eye tracker correctly and keep infants' visual attention on-screen. An infant-friendly video can be presented while the experimenter sets up the eye tracker (see Figure 2.5 for a screenshot of the video used for the screen-based study in Chapter 5).

Figure 2.5. Screenshot of the video used to maintain infants' visual attention on-screen while the eye tracker is being set up.

In addition, the calibration points should be infant-friendly to maintain visual attention. For this thesis, colourful, inward-turning spirals were presented and accompanied with an auditory stimulus (see Figure 2.6 on page 51).

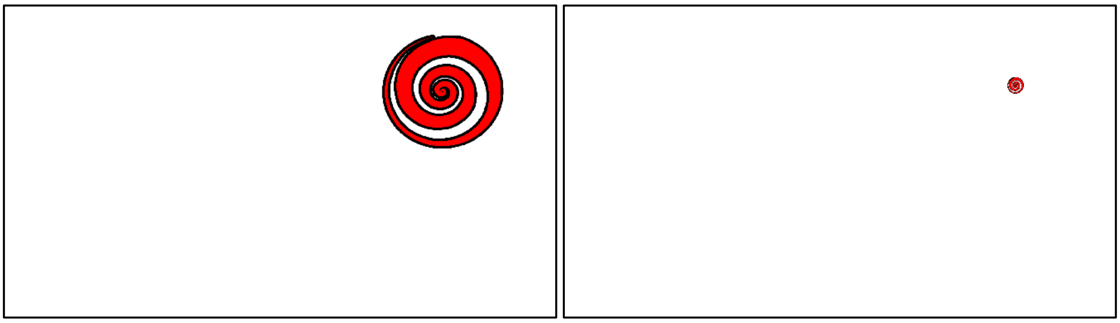


Figure 2.6. Screenshots of a calibration point at the start (left) and at the end of the presentation after spiralling inward (right).

Calibration procedures in more naturalistic settings rely on the same principles as those used for screen-based paradigms but differ slightly with respect to their implementation. Given that no screen is available, participants are typically asked to fixate a calibration object that is located in a fixed position or held by the experimenter (Holmqvist et al., 2011). The eye tracking software can then perform the relevant gaze estimations either on-line (e.g., *Tobii Studio*; Tobii Technology, Sweden) or off-line (e.g., *Yarbus*; Positive Science). For the adult studies presented in this thesis, calibration objects were held by the experimenter (Chapter 3) or the research assistant (Chapter 4; details on the calibration procedure will be given in the relevant chapters). For the face-to-face interaction study with infants (Chapter 6), a small, squeaky toy was held by the research assistant, and the experimenter manually triggered the recording of each calibration point via keyboard press as soon as the infant was judged to be looking at the calibration toy (Figure 2.4).

2.2.4. Pre-processing eye tracking data

Raw eye tracking data is noisy and is not always classified into oculomotor events (e.g., fixation, blinks, saccades). Several pre-processing steps are typically required to reduce noise in the raw data and classify gaze points by oculomotor events. In the following, a brief description of some pre-processing steps that were applied to the eye tracking data collected for this thesis will be provided.

To reduce noise, a bilateral *smoothing* algorithm can be applied to the raw eye tracking data (Holmqvist et al., 2011; Saez de Urabain, Johnson, & Smith, 2015). This can be done by averaging the gaze data from both eyes and removing jitter by considering the spatial position of several data points immediately before and after the current gaze point (e.g., Frank et al., 2009). In addition to smoothing, *interpolation* can be applied to fill very brief gaps of missing data that may have resulted from instabilities during gaze data recordings. Crucially, interpolation should only be performed when interruptions are short; for instance, interpolating gaps with a latency of less than 150 milliseconds will avoid filling in a saccade-fixation-saccade sequence or an instance of blinking (Holmqvist et al., 2011; Wass, Smith, & Johnson, 2013). Interpolation should also only be conducted when the last data point before and the first data point after the missing gap are minimally displaced from each other (cf., Saez de Urabain et al., 2015; Wass et al., 2013). Both smoothing and interpolation were performed for the studies presented in Chapters 4, 5, and 6, and relevant parameter values will be reported in the individual chapters. For Chapter 3, however, such pre-processing steps were not applied since the low temporal resolution of the eye tracker (30 Hz) resulted in visible displacements when the data was initially smoothed, and the option to interpolate data is not provided by the manufacturer's software *Yarbus* (Positive Science).

To examine eye movement behaviour, various dependent measures can be extracted from eye tracking data (for a review see Holmqvist et al., 2011). With respect to cultural differences in face scanning, previous studies have predominantly used either raw (or pre-processed) eye tracking data (e.g., Senju, Vernetti, Kikuchi, Akechi, & Hasegawa, 2013; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013) or fixation data (Blais et al., 2008; Jack et al., 2009; Liu et al., 2011; Wheeler et al., 2011) to calculate total visit durations in each region-of-interest such as the eyes or mouth or, alternatively, to perform data-driven analysis (a detailed discussion on these analysis methods will be provided below in Section 2.3). Whereas raw (or pre-processed) data includes all gaze data collected by the eye tracker (including, e.g., saccades), fixations represent only those periods when the eyes are stable, i.e. when visual encoding and cognitive processing are proposed to occur (see Holmqvist et al., 2011; Saez de Urabain et al., 2015). For this thesis, raw (or pre-processed) data and/or fixation data were used to examine face scanning strategies.

Given that terminologies and definitions for different types of eye movement data vary considerably (Hessels, Niehorster, Nyström, Andersson, & Hooge, 2018; Holmqvist et al., 2011), this thesis will henceforth refer to raw (or pre-processed) data as *dwell measures* and will refer to periods when the eyes are stable between saccades as *fixation measures*. To identify fixations within the raw (or pre-processed) eye tracking data, two types of event detection algorithms can be applied: a *dispersion-based algorithm*, which will identify a fixation if the positions of consecutive gaze points are within a specified positional threshold (e.g., 50 pixels) and above a minimum fixation duration (e.g., 100 milliseconds), and a *velocity-based algorithm*, which will identify a fixation if consecutive gaze points are below a specified velocity threshold (e.g., 20° per second) and above a minimum fixation duration (Holmqvist et al., 2011). For the present thesis, fixations in Chapter 4 (face-to-face interaction with adults) were identified using the dispersion-based algorithm that is provided by the manufacturer's software *BeGaze* (SensoMotoric Instruments, Germany), and fixations in Chapter 6 (face-to-face interaction with infants) used the velocity-based algorithm that is provided by the manufacturer's software *Tobii Studio* (Tobii Technology, Sweden). All relevant parameter values will be reported in the relevant chapters. Fixation analysis was not conducted for Chapter 3 (face-to-face interaction with adults using dual eye tracking) since the manufacturer's software *Yarbus* (Positive Science) does not provide the possibility to extract fixations.

Although these automatic algorithms provide a rapid approach to fixation detection, they are highly sensitive to noise, which in turn can affect the number and duration of detected fixations (Saez de Urabain et al., 2015). This is especially an issue when data quality is highly variable *between* experimental groups since this may introduce a systematic bias in the number and duration of fixations. For instance, eye tracking data from infants is significantly lower than that of adults (Saez de Urabain et al., 2015), thereby making cross-sectional studies prone to systematic biases. To account for this, the cross-sectional screen-based study in Chapter 5 used the semi-automatic fixation detection tool *GraFIX* (Saez de Urabain et al., 2015), which first applies an automatic algorithm to smooth and interpolate gaze data and detect fixations, before providing a moderation stage whereby the user can manually flag, delete, or modify fixations that were judged to be incorrectly detected by the automatic algorithm. Further details and relevant parameters will be provided in Chapter 5.

2.3. Data Analysis

After having provided a brief overview of eye tracking techniques, the following sections will describe in detail how the data collected for this thesis was processed and analysed. As mentioned above, previous studies have adopted a regions-of-interest approach or conducted data-driven analysis to examine cultural differences in face scanning. For the social interaction studies (Chapters 3, 4, and 6), both of these methods were adopted. However, the regions-of-interest approach requires gaze classification that is highly time-consuming for 'real-world' eye tracking data, and no data-driven method has been developed to date. The remainder of this chapter will therefore provide a detailed discussion on the challenges associated with 'real-world' eye tracking data to contextualise the solution developed for this thesis, namely a semi-automatic gaze annotation tool that can also perform data-driven analysis using statistical methods adopted from the neuroimaging literature.

2.3.1. Background: Regions-of-interest analysis

2.3.1.1. *Screen-based regions-of-interest analysis*

Screen-based paradigms that study face scanning typically involve a regions-of-interest analysis whereby individual facial features are first defined using rectangular or elliptical outlines (e.g., Oakes & Ellis, 2013; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013; Shic, Macari, & Chawarska, 2014; Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013; Wilcox, Stubbs, Wheeler, & Alexander, 2013; for a detailed overview of methods for regions-of-interest analyses see Holmqvist et al., 2011). Dependent measures such as dwell time or fixation time are then calculated separately for each face region to provide insight into scanning strategies. By spatially pooling gaze data in this way, the regions-of-interest approach can be statistically sensitive to detecting differences in eye movement behaviour between groups or conditions, particularly compared to other analytical methods (e.g., data-driven methods, which will be described in more detail below in Section 2.3.3). Most screen-based paradigms to date have employed static images or short dynamic videos as face stimuli, making it straightforward to define face regions. Although each stimulus requires the manual selection of individual face regions, this only needs to be performed a single time given that every participant is presented with the same stimulus set (Holmqvist et al., 2011). This regions-of-interest approach was applied in the screen-based face scanning study presented in Chapter 5.

After having briefly introduced regions-of-interest analysis for screen-based paradigms of face scanning, the following section will outline the additional challenges for 'real-world' eye tracking techniques and will explain how these were addressed for the work presented in this thesis.

2.3.1.2. *Challenges for 'real-world' eye tracking*

Unlike screen-based paradigms, investigating face scanning behaviour in naturalistic settings using eye tracking techniques poses various additional challenges. Scene recordings obtained from head-mounted eye trackers differ between individuals given that each participant exhibits unique head movements. With respect to dyadic face-to-face interaction paradigms, the location of the conversational partner's face can also never be known *a priori* since the face changes size and position in every frame. The location of the face must therefore be determined on a participant-by-participant and frame-by-frame basis, and this has typically been done manually whereby the gaze data is superimposed onto the scene recordings so that the coder can step through each frame to classify gaze points by event type (e.g. 'face look'; Franchak, Kretch, Soska, & Adolph, 2011; Yu & Smith, 2013). To facilitate this manual coding process, dedicated software is available. For instance, *Datavyu* (Datavyu Team, 2014) provides an interface that replays scene recordings, with keyboard shortcuts allowing for quick event coding. The resulting output file then provides relevant dependent measures such as overall dwell time on the face. Several eye tracking manufacturers also provide manual gaze annotation support. The SMI software *BeGaze* (SensoMotoric Instruments, Germany), for instance, replays scene recordings and allows the coder to map each fixation onto a reference image by clicking the corresponding position in the image.

Although such dedicated software considerably facilitates the manual coding process and typically results in highly accurate gaze classification, annotating data in this way is time-consuming and often requires multiple coders (Holmqvist et al., 2011). To address this issue, rapid automatic methods involving machine learning algorithms can provide an alternative way to detect and track faces, although only very few studies have employed such an approach (e.g., Bambach, Franchak, Crandall, & Yu, 2014). In their parent-child interaction study, Bambach et al. (2014) used a skin colour detection algorithm to identify face and hand regions, and subsequently distinguished faces from hands by detecting the black head-mounted camera worn by the person visible in the scene recordings. However, although such fully automatic methods significantly reduce processing time, accuracy rates for gaze classifications are often insufficient; in the given study (Bambach et al., 2014), accuracy rates ranged between 67% and 75%. Failures in automatic detection and tracking can result from multiple factors. For instance, recordings typically contain high motion and blurring due to the continuous head movements of the participant, resulting in low-quality footage that complicates feature detection. Partially occluded faces may also be missed, and irrelevant objects may be flagged as the target. Although accuracy rates were at above-chance level, inaccurate gaze classification can occur in a high number of frames and likely affect study findings and interpretations. In sum, automatic algorithms significantly reduce processing time, but at the expense of lower classification accuracy compared to the manual approach.

2.3.1.3. *Proposed solution for 'real-world' eye tracking*

A fully manual approach to gaze annotation is accurate but highly time-consuming. A fully automatic approach reduces processing time but lacks sufficient accuracy. The proposed method in this thesis exploits the advantage of each approach by combining automatic face detection and tracking algorithms that can process scene recordings and classify gaze points rapidly, coupled with a moderation stage that allows the user to interfere in case automatic algorithms fail. The eye tracking data obtained in Chapters 3, 4, and 6 was annotated using this proposed semi-automatic method. The following sections will now introduce this MATLAB-based tool by first describing the underlying algorithms and outlining the gaze annotation process, before going onto explaining data extraction and statistical methods for the regions-of-interest analysis of 'real-world' eye tracking data.

2.3.2. Semi-automatic regions-of-interest coding and analysis

2.3.2.1. *Face detection*

In computer vision, face detection refers to a process whereby faces are identified in given digital images (Viola & Jones, 2004; Yang & Huang, 1994). This can be achieved using a *classifier*, which comprises a two-step process of *training* and *classification*. Specifically, the classifier is first trained on a large set of images that are known to contain or not contain the target object, i.e. the face. Once the classifier is trained (using, e.g., optimisation algorithms; Felzenszwalb, Girshick, McAllester, & Ramanan, 2010), it can be run on new input images with unknown content to categorise them as positive face images or negative non-face images.

For the work presented in this thesis, faces were located using the Viola-Jones detector (Viola & Jones, 2001), which is built into the Computer Vision System Toolbox in MATLAB (Version R2015a, MathWorks). This detector is widely employed due to its high computational efficiency, which is achieved using a *Haar-like algorithm*, the notion of *integral images* (Viola & Jones, 2001), an adaptive algorithm based on *AdaBoost* (Freund & Schapire, 1997), and the concept of *cascading classifiers*. Each component is briefly outlined below.

Computing image properties for face detection on a pixel-by-pixel basis is computationally expensive. Using Haar-like features (Figure 2.7), the Haar-like algorithm instead examines rectangular image sections (rather than individual pixels) to calculate the difference in summed pixel intensity of adjacent sections. For instance, faces are typically characterised by a darker eye region coupled with lighter cheeks and a lighter nose, which can be represented as a Haar-like feature (see Figure 2.8 on page 59). By summing the pixel intensity of both the lighter and darker region, the sums can be subtracted to generate a single value that characterises that image region. The concept of integral images – a simple algorithm that can rapidly compute such intensity summations across the entire image – facilitates this process (for further details see Viola & Jones, 2001).

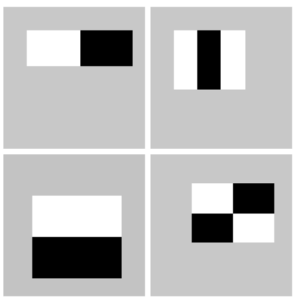


Figure 2.7. Examples of Haar-like features.

Figure 2.8. The upper face region represented as a Haar-like feature. Source: Viola & Jones (2001).

The adaptive algorithm (based on *AdaBoost*) then iteratively combines individual weak classifiers to establish a stronger classifier, and additionally computes weightings such that a high classification accuracy can be achieved. Finally, the concept of cascading classifiers refers to prioritising critical key features over other relevant ones when determining if a face is present or absent. Given that most image regions typically do not contain a face, it is more efficient to discard regions immediately if critical features are absent. Only regions that contain critical features are processed further in the next stage, with regions that pass all stages being categorised as containing a face.

For the work in this thesis, the face detector was implemented in MATLAB. The user first loaded in the input video, i.e. the scene recording of the eye tracker. Face detection was visualised using a rectangular bounding box that surrounded the face (Figure 2.9 on page 61), and the coordinates of the four vertices were stored for each frame. For the present thesis, the user also loaded an additional comma-separated values (csv) file that contained the manually coded frame numbers or recording times that marked the start and end times of the periods relevant for analysis. This property was added to the script given that only very short video segments of the overall recording required pre-processing. The input csv-file therefore provided event markers to rapidly skip scene frames that were not relevant for analysis.

Although an implementation for face detection in MATLAB is readily available online[1], the provided script suffers various shortcomings and had to be addressed for the work in this thesis. The most limiting property concerned the requirement for a face to be present in the first frame of the input video in order to proceed to the second frame; however, not all recordings met this requirement so that the scripts were adapted to allow the user to skip the frame manually if no face was present. Another limitation that significantly affected data pre-processing concerned the accuracy of automatic face detection. Occasionally, a present face was not detected or, conversely, detection occurred when the given frame contained no face. The failure to detect faces accurately is especially an issue for scene frames from head-mounted eye trackers given that participants often exhibit head movements (unlike stationary film cameras), resulting in increased scene motion and blurring and therefore a greater difficulty for the automatic algorithm to detect the face. In addition, the rectangular bounding box could not be angled to align it with any face tilts (see Figure 2.9 on page 61).

[1] MathWorks (2018, August 9). *Face Detection and Tracking Using the KLT Algorithm*. Retrieved from https://uk.mathworks.com/help/vision/examples/face-detection-and-tracking-using-the-klt-algorithm.html

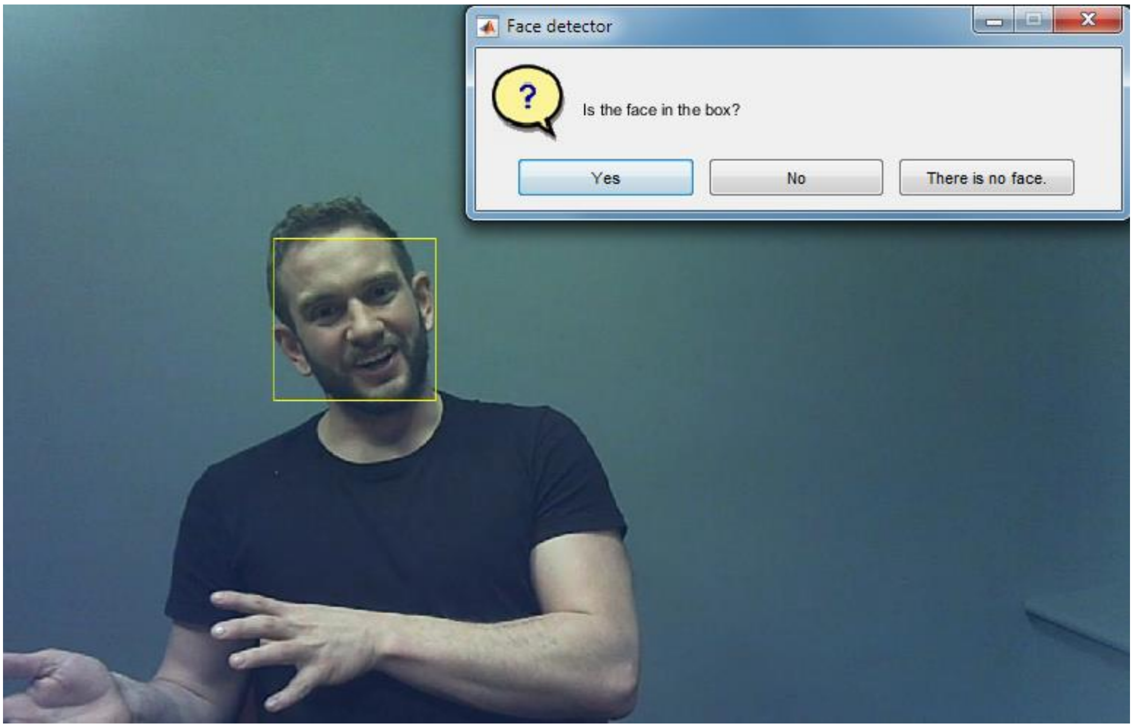


Figure 2.9. Screenshot of face detection process. The automatic algorithm attempts to fit a rectangular bounding box around the face, and the user is given the option to accept or reject the outcome.

To address these issues, the script was adapted to achieve higher accuracy and precision for face detection. A semi-automatic property was added to the script, allowing the user to indicate via a dialogue box whether a face was detected accurately (see Figure 2.9). If the user confirmed that the face was detected accurately, the script would proceed to the next scene frame. However, if the user disagreed, the face would be marked up manually by either dragging a rectangular box (for non-tilted faces) or marking four corner points to create a rectangular-like polygon (for tilted faces). The user was then asked again whether the bounding box accurately surrounded the face. If the user confirmed, the script would proceed to the next frame; alternatively, the script would loop back to the manual mark-up stage until the user agreed. A flowchart is depicted in Figure 2.10 on page 62 to illustrate the full coding process.

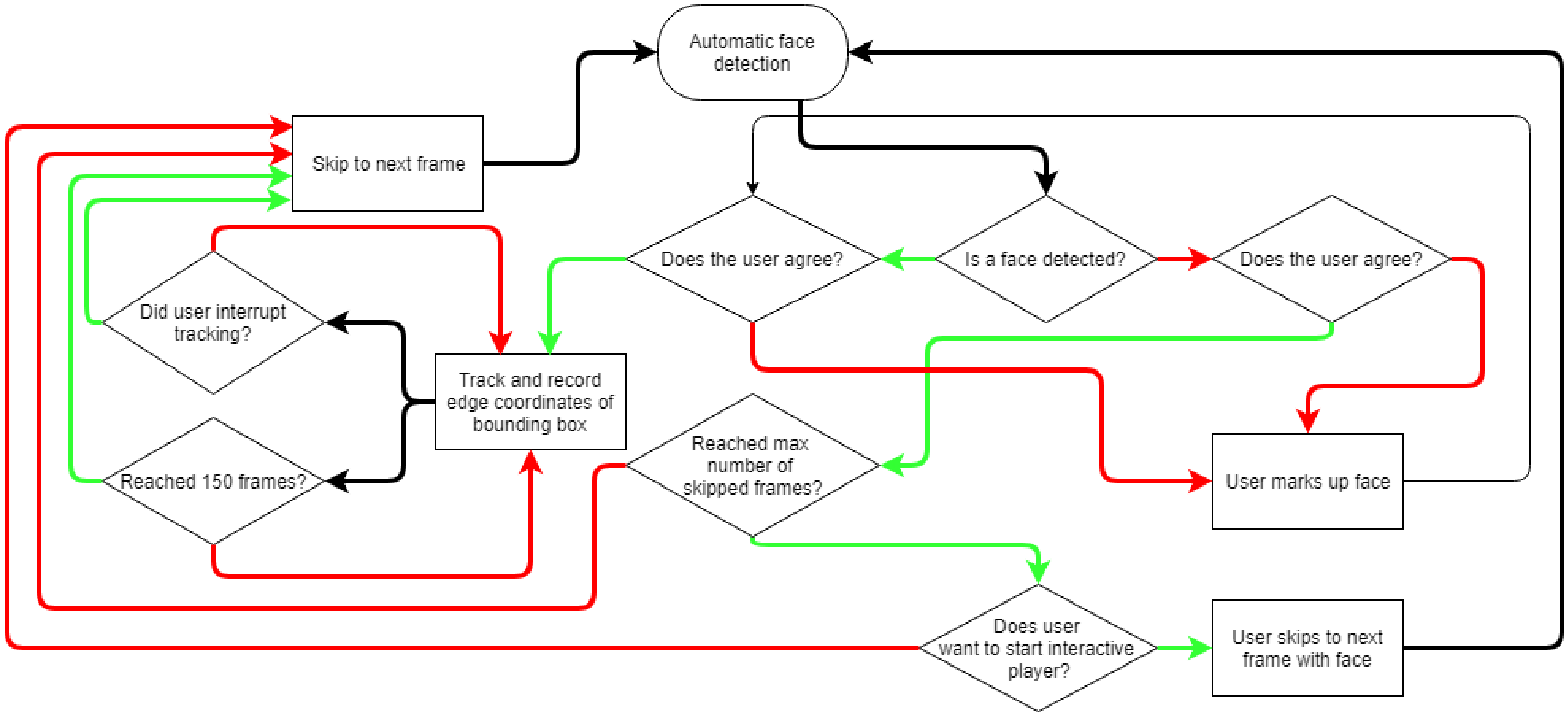


Figure 2.10. Flowchart visualising the semi-automatic face detection and tracking process. Green and red lines indicate 'yes' and 'no' responses, respectively.

For the studies presented in this thesis, the bounding box was checked and/or applied using the following guidelines: the upper and bottom edges were located along the middle of the forehead and just underneath the chin, respectively, while the side edges were aligned with the sides of the face including a small margin (Figure 2.11).

Figure 2.11. Close up of a randomly selected frame showing the manually coded face region based on pre-defined guidelines.

A small gap between the face outline was included to allow for measurement error. Given that manual selection introduces subjective judgement in marking up the face location, these guidelines were established to maximise consistency when defining the face region. The guidelines were also applied whenever the user was asked to agree or disagree with the fully automatic face detection performance. This was done since the automatic approach generated slightly different face locations every time.

Although more than one face can be *detected* within a given frame, the current script can only *track* a single face across frames (see Section 2.3.2.2 for details on face tracking – the process triggered once the face has been detected). Given that the relevant studies in this thesis included *dyadic* interactions (Chapters 3, 4, and 6), only one face was visible in the scene recordings at any given time such that the requirement for single-face tracking did not limit data processing. Finally, although the face was detected and tracked for the present studies, this semi-automatic tool can be applied for any other trained classifier (e.g., upper body) with minimal modifications to the current scripts. In addition, due to the possibility for manual coding, target objects without a trained classifier can also be detected and subsequently tracked.

2.3.2.2. *Face tracking*

Detecting the face in each frame is computationally inefficient and, for the purpose of the present thesis, would require considerable manual interference since automatic face detection often failed. An alternative method that can be applied after having successfully detected the face in a given frame is *feature tracking*. This process involves identifying distinct low-level feature points (based on, e.g., colour, shape, or texture) within the bounding box to subsequently relocate these points in the following frames. The geometric changes between the old and new points (e.g., shifts or rotations) are estimated and can then be applied to the rectangular bounding box. In this way, the face region can be located continuously without costly frame-by-frame face detection, providing a computationally superior approach particularly when processing a high number of frames.

For the work presented in this thesis, face tracking was achieved using the Kanade-Lucas-Tomasi (KLT) algorithm (Lucas & Kanade, 1981; Tomasi & Kanade, 1991), which is also built into MATLAB. Although other tracking methods are readily available, the increased scene motion and blurring in the head-mounted eye tracking recordings often resulted in unsuccessful face tracking since the relevant feature points could not be relocated in the following frame. For instance, the CAMShift (Continuously Adaptive

Meanshift; Bradski, 1998) algorithm[2] tracks target objects based on colour information that is unique to the target. With respect to faces, skin tone colour could be used for tracking; however, when attempting to process the scene videos, feature points were quickly lost following the face detection stage due to the high motion and blurring that distorted colour information.

The KLT algorithm, however, was able to relocate feature points continuously, thereby demonstrating its robust ability for face tracking that has also been confirmed in previous studies (e.g., Wagener & Herbst, 2001). The KLT algorithm attempts to find the displacement of a feature point between two subsequent frames to determine the new location of the point. This is achieved using optimisation techniques: the sum of squared intensity differences between features in the previous and the current frame is first computed, and the displacement of a single feature is then defined as the shift that minimised this sum of squared differences. The algorithm also produced an adaptive window, meaning that the position, size, and angle of the bounding box changed in line with the face region. As with the face detection stage, the four coordinates of the vertices of the bounding box were stored to log the face position for every frame. For the current implementation in MATLAB, a video player was used to visualise tracking behaviour (see Figure 2.12 on page 66).

[2] MathWorks (2018, August 9). *Face Detection and Tracking Using CAMShift*. Retrieved from https://uk.mathworks.com/help/vision/examples/face-detection-and-tracking-using-camshift.html

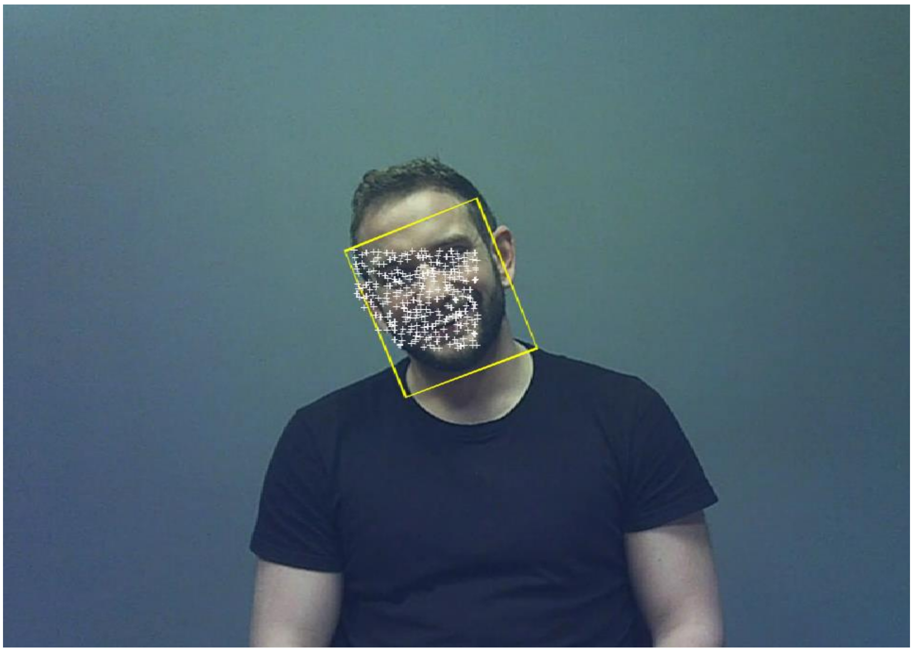


Figure 2.12. Screenshot of the video player that visualised the face tracking process.

A parameter that was required to be set manually by the user was the minimum number of feature points (i.e., the *threshold*) at which the algorithm should operate. A high threshold (i.e., a high number of points) ensures that many feature points are always available to estimate the location of the bounding box, resulting in a more spatially accurate and precise estimation of face regions. However, since the algorithm cannot always successfully relocate every single feature point such that some points are lost over the course of the tracking period, a high threshold results in a considerably shorter automatic tracking period relative to a low threshold. The threshold should therefore be set such that a suitable balance between tracking quality and processing time is achieved. Typically, the KLT algorithm only requires very few points (e.g., 5 points) to track a target object continuously. However, given that spatial accuracy and precision of the face regions were vital to investigate scanning behaviour for the current studies, the threshold was increased to 15 points in this thesis after having evaluated tracking performance of various higher and lower threshold values.

To improve tracking quality further, the original scripts were also modified to include an additional parameter, namely the maximum number of frames that should be processed before looping back to the face detection (and subsequent tracking) stage. Even with a points threshold, tracking quality declined as a function of tracking duration given that fewer points were available, such that the bounding box could no longer be accurately estimated. By adding a parameter that specified the maximum number of frames to be processed, tracking quality can be sufficiently maintained. For the present work, this parameter was set to 150 frames after having explored other values; this was equivalent to approximately 5, 6, and 7.5 seconds in Chapters 3, 4, and 6, respectively. After 150 frames (or earlier if the number of feature points fell below the threshold of 15 points), the script returned to the face detection stage (see flowchart in Figure 2.10 on page 62). A final property that was added to the original scripts to improve tracking quality included a pushbutton (Figure 2.13) that could be activated by the user at any time during the tracking period. This ensured that the user could manually stop the tracking period to return to the face detection stage in case the bounding box location could no longer be estimated accurately, even if the points threshold or the maximum number of frames had not yet been reached.

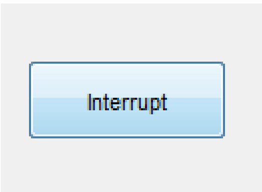


Figure 2.13. Screenshot of pushbutton that can be activated at any time during the face tracking period to return to the face detection stage.

Although the KLT algorithm is already implemented in MatLab[3], the provided scripts suffered some shortcomings that needed to be addressed to track faces accurately. Continuous tracking could only be performed while the frames contained a face;

[3] MathWorks (2018, August 9). *Face Detection and Tracking Using the KLT Algorithm*. Retrieved from https://uk.mathworks.com/help/vision/examples/face-detection-and-tracking-using-the-klt-algorithm.html

however, when the number of available feature points fell below the specified threshold, the algorithm stopped without processing the remaining frames. Similarly, if the face was no longer present in a frame (e.g., due to the participant turning their face away from the conversational partner), any subsequent frames could not be processed. To account for these scenarios, the original script was modified to automatically loop back to the face detection stage (see flowchart in Figure 2.10 on page 62). When a face was not present for at least 2 frames, the user was given the option to activate an interactive video player that allowed skipping forward to the next scene that contained a face. This parameter value can, however, be adapted by the user.

In sum, the modifications to the available scripts aimed to improve the spatial accuracy and precision of regions-of-interest coding by adding manual elements to the automatic face detection process. Following face detection, the user was required to confirm whether the automatic algorithm was sufficiently accurate in order to subsequently track the face using identified feature points. Additional properties were included to the present face detection and tracking procedure to increase performance accuracy (e.g., setting a points threshold for face tracking). Altogether, the semi-automatic nature of the face tracker with its added properties allowed for data coding and analysis with minimal manual interference, thereby reducing the time required for data pre-processing. The following sections will describe how the relevant eye tracking measures were extracted and analysed for the work in this thesis.

2.3.2.3. *Data extraction*

After having processed all frames relevant for analysis, the location of the face region was known and given by the edge coordinates of the bounding box. To gain a more detailed understanding of face scanning, the face area was subdivided into an upper and a lower part as a proxy for eye and mouth regions, respectively, in line with previous studies (e.g., Freeth & Bugembe, 2018; Vabalas & Freeth, 2016). This was computed by splitting the bounding box at the midline to obtain the top (upper face) and bottom half (lower face) of the face region (Figure 2.14). The following will describe how the dependent measures were extracted for each participant to conduct the regions-of-interest analyses presented in this thesis.

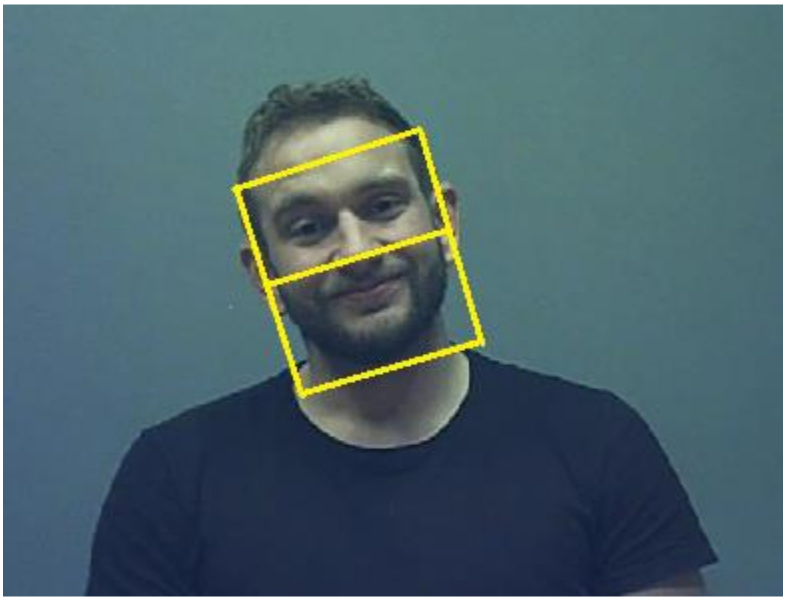


Figure 2.14. Close-up screenshot showing the midline of the bounding box that divides the face area into an upper and a lower region.

First, the raw (or pre-processed) eye tracking data file was loaded into MATLAB to extract the relevant columns for analysis, namely the *x*- and *y*-coordinates of the gaze points along with the corresponding timestamps. A csv-file that contained the manually coded frame numbers or recording times marking the analysis-relevant sections was also used to extract necessary rows in the gaze data file. By syncing the timestamps from the video and eye tracking data file, each gaze data point was associated with its corresponding scene frame. The gaze point could then be classified into an event type by checking whether or not the associated coordinates fell within the region of the bounding box (i.e., the face). When the coordinates fell within the face region, the gaze point was further classified as upper or lower face looking if it fell into the upper or lower half of the bounding box, respectively (Figure 2.14 on page 69). This gaze classification was implemented in MATLAB using the built-in `inpolygon`[4] function.

For every participant, a binary timeline was created for each region-of-interest (face, upper face, and lower face; see Figure 2.15): An entry was coded '1' if the corresponding gaze point fell within the region-of-interest, and '0' if it fell outside the region. An additional timeline described whether or not a gaze point was valid (coded '1') or invalid (coded '-1'). Entries were annotated invalid if data loss occurred (due to, e.g., blinks or track loss); this information was provided within the eye tracking data file, with coordinates during periods of data loss represented in a unique format (e.g., coordinates are set to '-1000' for the Positive Science system). A speech timeline was also added to annotate periods as listening (coded '0') or speaking (coded '1'). Speech information was provided in a csv-file that contained the manually coded start and end times or frames of the analysis-relevant period for every participant. Any entries associated with periods that were not relevant for analysis (e.g., short conversational periods during dyadic interactions) were set to '-1'. Each entry was also associated with the corresponding frame number and timestamp. An example timeline is given in Figure 2.15. For fixation analyses, a further timeline was added to annotate whether the gaze point was associated with a fixation (coded '1') or not (coded '0').

Timestamp (s)	11.331	11.365	11.399	11.435	11.468	11.502	11.538
Frame number	2390	2391	2392	2393	2394	2395	2396
Valid trial	1	1	1	1	1	1	1
Speech	0	0	0	0	0	0	0
Face	1	1	1	1	1	1	1
Upper face	1	1	1	1	1	1	1
Lower face	0	0	0	0	0	0	0

Figure 2.15. Snippet of coding timeline.

[4] MathWorks (2018, August 9). *Inpolygon*. Retrieved from https://uk.mathworks.com/help/matlab/ref/inpolygon.html

The dependent measures were then computed using the information provided in the timelines. For the present thesis, proportional dwell time and/or proportional fixation time (relative to the relevant recording period) were used as dependent variables to examine face scanning behaviour. Every participant thereby provided a proportional value between 0 and 1, with 0 indicating no looking at the region-of-interest, and 1 indicating continuous looking. A more detailed explanation of how the dependent measures were computed will be given in the relevant empirical chapters (Chapters 3, 4, and 6).

2.3.2.4. *Statistical analysis*

The proportional dwell and/or fixation time measures for each participant were written to a text (.txt) file for analysis in SPSS (Version 22, IBM). The data was statistically analysed using Analysis of Variance (ANOVA), with factors included in the model depending on the study design; detailed information on the analysis procedure will be provided separately in the relevant empirical chapters (Chapters 3, 4, and 6).

2.3.2.5. *Face detection and tracking performance*

Throughout the face detection and tracking stages, the gaze point was not visible, i.e. the eye movement data was not superimposed onto the scene camera videos. To ensure that the extracted dependent measures for the present analyses were not confounded by inaccuracies in face detection and tracking, manual checks were performed for 20% of the data collected for the study in Chapter 3 (face-to-face interaction with adults using dual eye tracking techniques). The data sets for the manual checks were randomly selected, with the East Asian and the Western Caucasian participant group each contributing 10% of data. Accuracy rates were computed by first collating all frames associated with the same event type (namely upper face, lower face, or non-face), separately for each data set. For each event type, the associated gaze points were then superimposed on the relevant frames and replayed for the user. When a gaze point was judged to be misclassified, the user was able to reclassify the point using simple keyboard shortcuts ('e' for upper face (eyes), 'm' for lower face (mouth), and 'o' for non-face (off). The timeline was immediately updated via these keypresses to reflect the changes in gaze classification. The accuracy was then computed by comparing the entries of the original timeline with the new timeline and calculating the proportion of true negatives and true positives relative to all entries. Mathematically,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP is number of correctly classified event hits (true positives), FP is number of incorrectly classified event hits (false positives), TN is the number of correctly classified event misses (true negatives), and FN is the number of incorrectly classified misses (false negatives). Results showed that the mean accuracy for detecting and tracking the face *prior* to any manual reclassification was $M = 99.56\%$ ($SD = 0.90\%$; upper face $M = 99.02\%$, $SD = 1.37\%$; lower face $M = 99.35\%$, $SD = 0.97\%$), reflecting a much higher accuracy than fully automatic procedures (e.g., 67% to 75% for detecting hands and faces in Bambach et al., 2014). Given the high accuracy, the remaining gaze data sets in this thesis were not manually reclassified, although the option is generally provided should the user wish to carry out classification checks.

The time required for semi-automatic gaze annotation was also evaluated using an infant data set collected for Chapter 6 (face-to-face interaction with infants) and compared to a manual coding approach. To code upper and lower face and non-face regions for 1 minute of recording time, the current semi-automatic method required 5 minutes and 16 seconds with minimal manual interference that was only required for face detection (but not tracking). Using a fully manual approach, gaze annotation took 11 minutes and 29 seconds (i.e., more than double the time compared to the semi-automatic approach).

2.3.3. Background: Data-driven analysis

2.3.3.1. *Screen-based data-driven analysis*

The regions-of-interest approach is widely employed to examine face scanning strategies, with head-mounted eye tracking studies typically coding the upper and lower half of the face for analysis (e.g., Vabalas & Freeth, 2016). By focusing the analysis on selected face regions, this approach is particularly useful when testing specific predictions about scanning behaviour of certain facial features since the spatial pooling of data makes it statistically sensitive to detecting differences.

However, pre-defining individual face regions can be prone to several limitations. For instance, the manual selection of face regions requires the subjective judgement of the coder about the size, position, and shape of each individual area. Differences in coder judgement could therefore potentially result in different findings. For the present thesis, pre-defined guidelines were established to achieve consistency in coding (see Section 2.3.2), although future studies will be required to follow the same guidelines in order to allow cross-study comparisons.

With respect to cultural differences in face scanning, the regions-of-interest approach can provide insight into specific predictions such as whether East Asian participants direct their gaze to the eye region more than Western Caucasians (Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). However, the binary nature of the regions-of-interest approach (i.e., upper versus lower face) may limit new insights into cultural differences in face scanning. Although East Asians have been found to scan the eye region more than Western Caucasian participants when presented with emotionally expressive faces, different studies show that East Asians exhibit a central gaze pattern of the nose when scanning neutral faces (Blais et al., 2008). With the regions-of-interest coding above, however, the nose falls onto the boundary between the two regions and gaze points directed at the nose fall onto either the upper or lower region. In addition, given that no study has previously conducted cross-cultural comparisons of live dyadic interactions, it is possible that scanning behaviour within face-to-face settings are characterised by entirely different eye movement patterns. Restricting analysis to pre-defined regions therefore enables a fruitful approach to test specific predictions but may hinder new insights and potentially mislead study interpretations.

An alternative method for analysing face scanning strategies that accounts for these limitations involves a data-driven approach that is free from the subjective judgement about the size, position, and shape of each face region. In their *iMap* toolbox, Caldara & Miellet (2011) implemented such a data-driven method to analyse screen-based eye tracking data. This involved presenting face stimuli that were matched in size and position, thereby allowing gaze data to be aggregated across time and/or stimuli to produce gaze density maps. These maps visualise gaze clusters, i.e. regions that are high in gaze density (see Figure 2.16 for an example) and can therefore inform about face scanning strategies without pre-defining facial features.

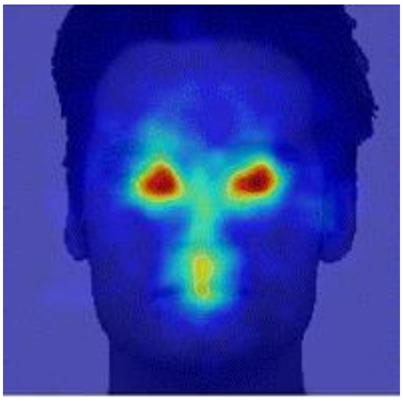


Figure 2.16. Data-driven visualisation of face scanning. Source: Caldara & Miellet (2011).

Such a data-driven approach therefore not only addresses the subjectivity issue underlying manual coding, but also makes no *a priori* assumptions about which face regions should be included in the analysis. Caldara & Miellet (2011) highlighted the usefulness of data-driven methods for exploratory purposes. Considering the gaze data illustrated in Figure 2.17 on page 75, the data-driven method highlights a prominent gaze cluster stretching from the nose region into the mouth area, whereas the regions-of-interest analysis cannot capture the intermediary area and in turn may lead to different study interpretations.

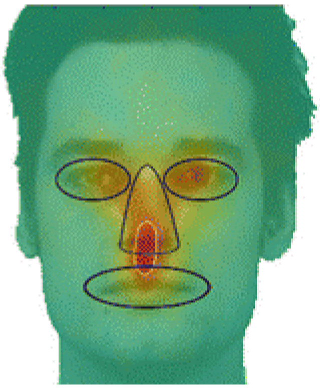


Figure 2.17. Interpretation of scanning behaviour can depend on the analysis method. The data-driven method highlights a single gaze cluster between the nose and mouth (in red), whereas the regions-of-interest approach cannot capture the intermediary area between the nose and mouth. Source: Caldara & Miellet (2011).

In sum, given that gaze data is spatially pooled in each face area when performing regions-of-interest analyses, this method may be statistically more sensitive for detecting group differences in scanning behaviour and therefore provides a useful way to test specific predictions. However, as this section outlined, a data-driven approach accounts for subjectivity in coding and also allows new insights into scanning strategies. The following sections will first describe the additional challenges for the data-driven approach when analysing gaze data from head-mounted eye trackers (or the Tobii TX300 used in standalone mode). To contextualise the proposed solution for data-driven analysis of 'real-world' gaze data, underlying theoretical concepts of cluster-based analysis will be briefly outlined before explaining in detail the solution for data extraction and subsequent data analysis.

2.3.3.2. *Challenges for 'real-world' eye tracking*

To date, no study has conducted data-driven analysis for 'real-world' eye tracking data. This is likely due to the methodological challenges underlying data extraction for statistical analysis. Unlike screen-based paradigms, 'real-world' eye tracking data cannot simply be collapsed across time and/or stimuli given that the position, size, and angle of the face changes with every frame. As described in Section 2.3.1.2, the common approach for processing gaze data from head-mounted eye trackers involves the classification of gaze points by categorical events (e.g., 'face look'). Such event-based coding consequently does not lend itself to data-driven analysis, which requires spatially sensitive gaze information to collapse the data across time and/or stimuli. Several eye tracking manufacturers do allow for manual, semi-automatic, or automatic data extraction methods that are appropriate for data-driven analysis. SMI's *BeGaze* (SensoMotoric Instruments, Germany) allows manual re-mapping of gaze data onto a reference image, as described in Section 2.3.1.2. Although this provides spatially sensitive, normalised data aggregated across time, the manual nature of this task can be highly time-consuming. Tobii's *Pro Lab* (Tobii Technology, Sweden) provides a semi-automatic or fully automatic approach whereby the algorithm detects the target object (via a given reference image) before re-mapping the corresponding gaze point onto the reference image. However, this processing approach is manufacturer- and model-specific and cannot be used in conjunction with other eye tracking systems, including the ones used in this thesis.

A further challenge to data-driven analysis of 'real-world' eye tracking data relates to the statistical analysis. Although data extraction methods are potentially available, as outlined above, no study to date has statistically analysed gaze density of 'real-world' eye tracking data. Current screen-based methods for analysis, such as *iMap* (Caldara & Miellet, 2011), were not developed to deal with 'real-world' eye tracking data. Consequently, there is currently no implementation of a statistical method that can conduct data-driven analysis with such eye tracking data.

2.3.3.3. *Proposed solution*

This thesis introduces a novel data-driven method for 'real-world' eye tracking data that can collapse gaze points across time and across participants to visualise and analyse gaze density. Specifically, linear transformations were applied to re-map gaze points onto a normalised face template in a fully automatic fashion. The method was implemented in MATLAB and is therefore manufacturer-independent such that it can be applied to data from any other eye tracking system. With respect to the statistical analysis, existing neuroimaging analysis techniques that aim to identify significant clusters were applied to the analysis of eye tracking data. In particular, group differences in gaze density were examined using *Monte Carlo permutation testing* (also named *approximate permutation test* or *random permutation test*; Nichols & Holmes, 2002). A detailed description of the data extraction and statistical analysis processes will be provided in the following sections.

2.3.4. Data-driven analysis of 'real-world' eye tracking data

2.3.4.1. *Data extraction*

To create density maps for the face region, the gaze points needed to be collapsed across time and across participants for each cultural group. However, given that the location, size, and angle of the face changed with every frame, and given that the gaze coordinates were expressed with respect to the scene frame, gaze points could not simply be collapsed without additional processing. The following will therefore describe a novel approach that re-expresses gaze coordinates with respect to the face region (as opposed to the scene frame), thereby accounting for changes in face location, size, and angle. The new coordinates can then be used to create data-driven gaze density maps of face scanning.

The gaze coordinates represented absolute coordinate values with respect to the entire scene frame. To collapse gaze points, the coordinates needed to be expressed with respect to the face region – i.e., the bounding box – independent of the location, size, or angle of the face. This was achieved by transforming the absolute coordinates of all gaze points that fell within the face region to coordinates relative to the bounding box. The following will explain the steps that were taken in more detail.

(1) To transform coordinates for the current purpose, the bounding box required a rectangular shape. Any bounding box that had a four-point polygonic shape (due to the manual marking of tilted faces; see Section 2.3.2) first needed to be transformed into a rectangle. This was achieved using the `minboundrect`[5] function in MATLAB, which fit a rectangle such that the four vertices were surrounded by the edges with minimal error (see Figure 2.18 for an example).

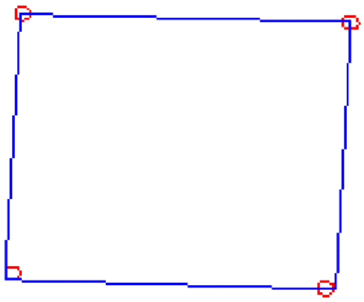


Figure 2.18. A rectangle (in blue) is fitted around the four vertices (in red) of a polygon.

(2) To simplify the re-mapping of coordinates, the rectangular bounding box and the corresponding gaze point were rotated for all frames containing a tilted face (i.e., a tilted bounding box). Rotation was performed such that the top and bottom edges of the bounding box aligned in parallel with the x-axis of the scene frame, i.e. such that the face was no longer tilted (see Figure 2.19 on page 79 for an illustration).

[5] MathWorks (2018, August 9). *A suite of minimal bounding objects*. Retrieved from https://uk.mathworks.com/matlabcentral/fileexchange/34767-a-suite-of-minimal-bounding-objects

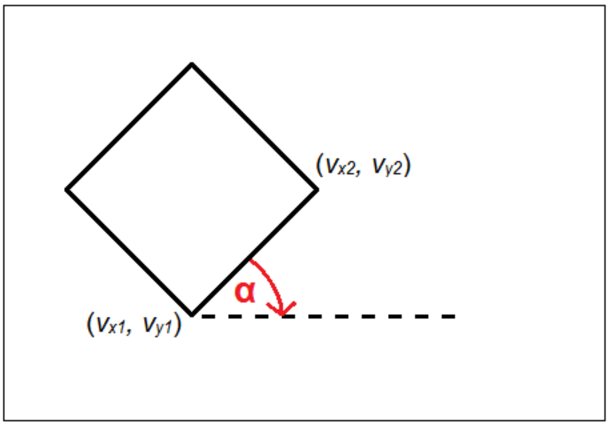


Figure 2.19. Rotating the tilted bounding face box by angle α such that the bottom edge (with endpoints (v_{x1}, v_{y1}) and (v_{x2}, v_{y2}) is parallel to the *x*-axis of the scene frame.

Given that the sole aim was to achieve a non-tilted bounding box without further requirements, the rotation did not need to be performed around a specific point. For each individual frame, this rotation was achieved by

- computing the angle α (see Figure 2.19) so that the bottom edge of the bounding box – with endpoints (v_{x1}, v_{y1}) and (v_{x2}, v_{y2}) – is parallel to the *x*-axis by solving the system of linear equations for *c*,

$$\begin{bmatrix} v_{x1} & 1 \\ v_{x2} & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} v_{y1} \\ v_{y2} \end{bmatrix},$$

- substituting c_1 to solve α using

$$\alpha = \arctan2(c_1, 1),$$

- using the angle α to set up a rotation matrix[6] *R*,

$$R = \begin{bmatrix} \cos(-\alpha) & -\sin(-\alpha) \\ \sin(-\alpha) & \cos(-\alpha) \end{bmatrix};$$

- applying matrix multiplication to rotate the bounding box and obtain the new coordinates of each shifted vertex (v_x', v_y'),

$$\begin{bmatrix} v_x{}' \\ v_y{}' \end{bmatrix} = R \begin{bmatrix} v_x \\ v_y \end{bmatrix};$$

[6] To perform clockwise rotations, the angle α is negative in this rotation matrix.

- applying the same operation for the original gaze coordinate and obtain the new gaze coordinates (x', y')

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = R \begin{bmatrix} x \\ y \end{bmatrix}.$$

(3) For all gaze points that fell within the face region, the gaze coordinates were expressed relative to the rotated bounding box, and setting its vertices to v_1 = (-1, -1), v_2 = (1,-1), v_3 = (1,1), and v_4 = (-1,1), such that the centre of the face (the nose tip) represented the origin (0,0) (see Figure 2.20).

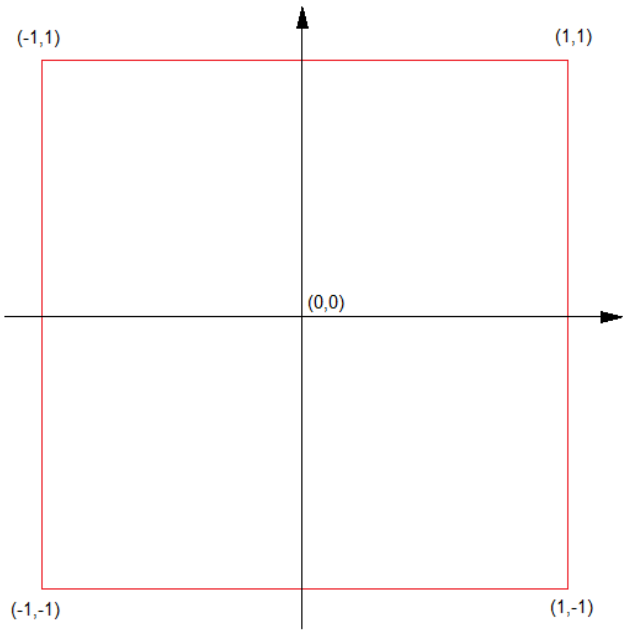


Figure 2.20. An illustration of the bounding box (in red) within a unified coordinate system, with the origin corresponding to the nose tip.

To ensure that the nose tip was indeed the centre of the face, methods were initially employed to account for any head turns that would have resulted in a shifted centre (e.g., the nose tip would not be the centre of a face that is shown in profile view). Specifically, the bounding box was divided into four quadrants with the true centre of the face (nose tip) representing the point of intersection, followed by linear transformations (such as stretching or squeezing) of each quadrant so that the centre of the unified coordinate system represented the true centre of the face. However, participants rarely performed head turns, and this method was therefore not applied to the work presented in this thesis.

(4) The user then set up a grid separately for each participant and for each experimental condition. This was done to map all relative gaze coordinates into a unified coordinate system. In this thesis, a grid of 100 x 100 was established, with the same vertices as the bounding box, i.e., $v_1 = (-1,-1)$, $v_2 = (1,-1)$, $v_3 = (1,1)$, and $v_4 = (-1,1)$.

(5) For each participant and experimental condition, the transformed gaze coordinates were mapped into the grid. Since all transformed gaze coordinates were relative values with respect to the same coordinate dimensions, their location within the 100 x 100 grid was found and filled in at the corresponding entry. The two-dimensional grid contained spatial, but no temporal information (i.e., gaze points were collapsed across time for each participant), given that the focus of this thesis was to examine the *spatial* distribution of scanning behaviour across the face. However, this step can be skipped for temporally-sensitive approaches such as scan path analyses (Holmqvist et al., 2011).

(6) In the final step, the resulting gaze density map was smoothed to take into account eye tracking measurement error and to consider that visual attention occurs not only at the gazed position but also within the immediate neighbouring region (Holmqvist et al., 2011). Typically, foveal visual attention is reported to be distributed within 1.5° to 2° visual angle (Holmqvist et al., 2011). For this thesis work, smoothing was therefore performed using a two-dimensional isotropic Gaussian kernel, with a kernel width that corresponded to 2° visual angle. Smoothing was achieved using the built-in MATLAB function `imgaussfilt`[7].

[7] MathWorks (2018, August 9). *Imgaussfilt*. Retrieved from https://uk.mathworks.com/help/images/ref/imgaussfilt.html

This procedure for data extraction to create individual gaze density maps was performed for each participant and each experimental condition. These maps were then used for statistical analysis, which will be introduced in the next section.

2.3.4.2. *Statistical analysis*

No study to date has employed statistical analysis methods for gaze density maps of 'real-world' eye tracking data, and only very few studies have adopted a data-driven approach for screen-based eye tracking data (e.g., *iMap* by Caldara & Miellet, 2011). These available (screen-based) approaches – as well as the proposed solution for this thesis – were typically inspired by the neuroimaging literature. The aim of functional magnetic resonance imaging (fMRI) studies, for instance, involves contrasting experimental groups or conditions to identify those brain regions that differ in neural activation patterns. Detecting such significant clusters requires the analysis of brain activation (density) maps containing a high number of voxels. Given the current aim to identify significant clusters that differ between groups and conditions in gaze density maps, appropriate statistical methods common in the neuroimaging literature can be adapted and applied to eye tracking data. The following will outline first the challenging nature of the multiple comparison problem and available (screen-based) methods that can address this issue, before going onto discussing the present statistical method for data-driven analysis of eye tacking data.

To identify regions in gaze density maps that significantly differ between two experimental groups, conducting an independent *t*-test separately for each pixel in the map introduces the *multiple comparison problem*. When running a single *t*-test, the alpha-level is typically set to 0.05; in other words, if the test were conducted 100 times, the null hypothesis would be, on average, wrongly rejected in five cases. However, for a map with 100 x 100 resolution (as in the maps used in Chapters 3, 4, and 6), a pixel-based analysis would result in 10,000 independent *t*-tests, giving around 500 pixels flagged as significant by chance. To adjust the alpha-level from a local scale (i.e., a single pixel) to a global scale (i.e., the entire map), several approaches can be adopted to address the multiple comparisons problem.

The Bonferroni correction method can compute an adjusted significance threshold which can be approximated by dividing the alpha-value by the number of tests. In the above example, this would result in an adjusted threshold of 0.05 / 10,000 = 0.000005. This threshold, however, can be considered too conservative due to the concept of *spatial correlation*. In neuroimaging, neighbouring voxel regions are not entirely independent from each other since similar regions tend to show similar neural response patterns; in other words, voxel activations are spatially correlated. Given that the Bonferroni correction method assumes that voxel activations are independent from each other, it computes an adjusted threshold that is too conservative, thereby reducing the statistical power to detect truly significant voxels. With respect to the present eye tracking data, the Bonferroni correction method would also be too strict given that gaze points closer in time are also closer in space, and as such gaze points are not entirely independent from another when collapsed across time. In addition, it is common to spatially smooth eye tracking data (as with neuroimaging data) to account for measurement error and to take into consideration that visual attention is continuously distributed around the focal gaze point. Smoothing therefore gives rise to spatially dependent eye tracking data, making the Bonferroni method an overly conservative approach.

An alternative method that takes into account spatial correlation is based on Random Field Theory (RFT; Adler, 1981; Worsley, Marrett, Neelin, Friston, & Evans, 1996), which also provided the theoretical framework for *iMap* (Caldara & Miellet, 2011). Specifically, the smoothness underlying activation maps is estimated, and the Euler characteristic (i.e., the number of clusters or "blobs" after thresholding; for further details see Brett, Penny, & Kiebel, 2003) is determined at varying thresholds. The threshold at which 5% (i.e., an alpha-level of 0.05) of equivalent statistical maps would occur under the null hypothesis can then be computed. RFT assumes an underlying

Gaussian distribution and requires sufficient smoothness of the given data. Although these assumptions can be met, the RFT approach may be unreliable under specific circumstances; paradigms with only a low number of participants, for instance, may not necessarily produce sufficiently smooth maps (Brett et al., 2003).

Another approach – and the one chosen for the work in this thesis – is non-parametric permutation testing (Nichols & Holmes, 2002), which has previously only been implemented in two screen-based studies by the same research group (Arizpe, Kravitz, Walsh, Yovel, & Baker, 2016; Arizpe, Kravitz, Yovel, & Baker, 2012). Permutation testing does not require underlying data to be normally distributed since it uses the observed data itself to generate a null distribution. This can be achieved by exchanging data across conditions or groups in all possible arrangements to compute the frequency distribution of test statistics (e.g., t-score). To illustrate, consider a paradigm with two participants A and B in one experimental group, and an additional two participants C and D in a different experimental group. By shuffling participants into all possible combinations, test statistics are calculated for AB (Group 1) vs CD (Group 2), AC (Group 1) vs BD (Group 2), and AD (Group 1) vs BC (Group 2) to obtain the underlying frequency distribution. This represents the null distribution of the observed data set, describing the distribution of test statistics if the participants' group allocations were entirely random. Naturally, these permutations are typically conducted for data sets with a much larger participant number. Conducting all possible permutations to establish the null distribution is, however, time-consuming and highly demanding in terms of computational power. An alternative approach is the *Monte Carlo method* (Manly, 1997), which approximates the null distribution by running permutations many times (typically in the order of several thousand iterations) – the larger the number of iterations, the better the approximation. For each iteration, the experimental data is first randomly swapped between groups or conditions, and a test statistic (e.g., t-value) is computed. This process is repeated many times, resulting in an approximate frequency distribution of the given test statistic. Using this distribution, the proportion of test statistics that resulted in larger values than the observed statistic (from the experimental data) can be computed. To achieve a significant group difference for the experimental data, the proportion of test statistics with larger values should therefore be minimal (e.g., less than 5%, or $p < 0.05$). This proportional value represents the *Monte Carlo significance probability*.

The only assumption required to be met for permutation testing is *exchangeability* (Nichols & Holmes, 2002), i.e. data needs to be exchangeable across conditions or groups. This can be an issue in neuroimaging due to the notion of *temporal autocorrelation,* whereby the amount of neural activation at one time point is significantly explained by the activation during the preceding time point (A. M. Smith et al., 1999). However, exchangeability of entire data sets from different participants – as in the present thesis – is still possible without violating assumptions.

For this thesis, the Monte Carlo permutation test was implemented in MATLAB using the *CoSMoMVPA* toolbox (Oosterhof, Connolly, & Haxby, 2016) and *FieldTrip* toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Although CoSMoMVPA and FieldTrip have been developed for neuroimaging data, the eye tracking data was arranged to suit the format that was necessary for analysis. The statistical analysis in this thesis involved *cluster-based* permutation tests. Specifically, neighbouring pixels (i.e., clusters) would first be identified if their test criterion (e.g., sum of *t*-scores) was greater than the critical value (t_{crit}) associated with a specified *p*-value threshold. This value was required to be set by the user, and a moderately strict threshold of 0.01 was chosen for this thesis. To examine which of these identified clusters survived multiple comparison correction, Monte Carlo permutation testing was performed as described above, with the frequency distribution representing test statistics at cluster-level (e.g., *t*-score at cluster-level).

2.4. Conclusion

In this chapter, a novel computational method was proposed that can perform highly accurate gaze classification of 'real-world' eye tracking data in a semi-automatic fashion and translate raw gaze coordinates into a unified coordinate system to create density maps. These two operations in turn allowed subsequent regions-of-interest analysis and, for the first time, data-driven analysis of 'real-world' eye tracking data. As outlined in this chapter, the regions-of-interest approach pools data for each face area and can therefore be statistically sensitive in testing specific predictions about group differences in scanning behaviour. Data-driven methods, on the other hand, are free from any subjective judgement in region selection and therefore represent a useful way to analyse eye movement data when pre-defining regions may limit interpretations and new insights. These novel tools were used to facilitate the gaze annotation process and to conduct a *spatially-sensitive* statistical analysis of the data collected for the naturalistic face-to-face interaction studies presented in this thesis. Detailed findings will be outlined in Chapters 3, 4, and 6, demonstrating also the applicability of the proposed methodology with different eye tracking systems (Positive Science, SMI, and Tobii).

Chapter 3

Cultural Differences in Face Scanning During Live Dyadic Interactions I

3.1. Chapter Overview

As highlighted in Chapter 1, eye tracking studies have identified significant differences in face scanning strategies between cultures, thereby challenging the idea that face perception processes are universal (e.g., Blais et al., 2008; Jack et al., 2009; Miellet et al., 2012; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). However, these studies were conducted within highly controlled lab settings in which participants were typically presented with static face stimuli on screens. Given that viewing behaviour can be modulated by factors such as low-level motion (Itti, 2005; Mital et al., 2010) or social presence (Foulsham et al., 2011; Laidlaw et al., 2011), it is unclear whether current findings also generalise to naturalistic settings. In this chapter, dual head-mounted eye tracking techniques were used to examine cultural differences in face scanning strategies during live dyadic interactions. To contextualise this work, the following sections will first briefly recap the literature on cultural differences in face scanning to highlight the importance of conducting naturalistic face-to-face interaction paradigms before going on to introducing the current study.

3.2. Introduction

3.2.1. Cultural differences in face scanning

Although behavioural, cognitive, and neural processes underlying face perception have typically been assumed to be universal (Han & Northoff, 2008), eye tracking studies have found significant differences in face scanning strategies between Western Caucasians and East Asians. When asked to learn and recognise face identities with neutral expressions, Western Caucasians exhibited a greater triangular scanning pattern of the eyes and mouth, whereas East Asians showed more fixations on the central face region (i.e., the nose; Blais et al., 2008). When categorising (Jack et al., 2009) or free-viewing emotionally expressive face stimuli (Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013), Western Caucasians scanned the mouth region significantly more than East Asians, who instead fixated the eye region. Although it is unclear whether the differences in scanning patterns between neutral and emotionally expressive faces resulted from differences in stimulus characteristics or task demand (recognition versus emotion categorisation or free-viewing), the studies point to cultural modulations in eye movement behaviour across different face processing tasks.

3.2.2. Limitations of screen-based paradigms

To date, cross-cultural comparisons on face scanning have used screen-based paradigms that presented adult participants with static face stimuli (except Gobel et al., 2017, and Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013, who used dynamic videos). Restricting studies to screen-based paradigms, however, can limit interpretations of current evidence when ultimately the goal is to understand ‘real-world’ behaviour. Static images displayed on a two-dimensional screen do not contain the dynamic properties and environmental distractors that are common to ‘real-world’ conditions. Studies on dynamic scene viewing have demonstrated that low-level motion can predict gaze location (Mital et al., 2010). In their cross-cultural study, Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al. (2013) employed dynamic face stimuli and found that East Asians fixated the eyes whereas Western Caucasians looked more at the mouth, replicating the results obtained by Jack et al. (2009) who used static face images. Although this suggests that the addition of dynamic properties does not override cultural effects, this would need to be confirmed within a face-to-face interaction paradigm.

Moreover, the lack of any social presence in screen-based paradigms implies that sociocultural norms were not considered in previous studies; while participants can look at static faces on screen for as long as they wish, it is considered socially inappropriate to gaze excessively at strangers in ‘real-world’ situations. As mentioned in Chapter 1, visual orienting to faces has been shown to decrease significantly when participants sat in the same room with another person, compared to when participants saw a videotape of the same individual (Laidlaw et al., 2011). With respect to face scanning, Gobel et al. (2017) found cultural differences only when the dynamic faces on screen made direct eye contact with the participant (i.e., looked directly into the camera), with East Asian participants scanning the nose and Western Caucasians fixating the eyes. No cultural differences were found, however, when the faces averted their gaze. Previous studies have indeed reported an avoidance for prolonged eye contact in East Asians (Argyle et al., 1986; A. McCarthy et al., 2006, 2008), which is proposed to act as a sign of respect (Sue & Sue, 1990). Although this would suggest reduced gaze scanning in East Asian participants, current findings on gaze avoidance are based on self-reports (Argyle et al., 1986) or video recordings of dyadic interactions during which participants were engaged in abstract thinking (A. McCarthy et al., 2006, 2008). Consequently, although the notion of gaze avoidance has been widely reported, no quantitative evidence exists that describes eye contact behaviour during naturalistic dyadic interactions. Given that reduced eye contact in East Asians may also contradict the cross-cultural findings showing increased scanning of the eye region for emotionally expressive faces (Jack et al., 2009), it remains unclear how eye movements manifest during face-to-face interactions. In the current study, the eye movements of both participants within a dyad were therefore recorded to examine both face scanning and eye contact behaviour.

Finally, participants in screen-based studies are usually instructed to look at the screen throughout the testing session. Given that typically only the face image appears on the screen, these paradigms cannot examine the extent to which individuals would visually orient to faces spontaneously within naturalistic social contexts. Although previous face-to-face interaction studies have shown decreased face looking when participants were speaking compared to when they were listening (e.g., Freeth et al., 2013) – suggesting that contextual factors inherent to dyadic interactions can modulate face orienting behaviour – these studies were conducted in Western cultures and no study to date has therefore examined whether naturalistic face orienting may differ between cultures.

3.2.3. The current study

The shortcomings of screen-based paradigms lead to several questions: How much do individuals engage in face looking during dyadic interactions, and does this differ cross-culturally? Do cultural differences in face scanning also exist under naturalistic conditions? If so, do these differences manifest in the same way as current screen-based findings, or do eye movements show an entirely different pattern?

To answer these questions, the aim of the current study was to identify an appropriate face-to-face interaction paradigm that could examine cultural differences in face looking and scanning within a naturalistic social context. The eye movements of both participants within a dyad were recorded using head-mounted eye tracking techniques. Based on previous face-to-face interaction studies (cf., Freeth, Foulsham, et al., 2013), it was predicted that participants would show increased face looking during listening compared to speaking periods. As there is currently no evidence to suggest that face looking could be modulated by cultural influences, no group differences were expected. However, it was predicted that East Asian participants would scan the upper face region proportionally more, whereas Western Caucasian individuals were expected to engage in increased mouth looking. These predictions were based on the findings from screen-based studies with emotionally expressive face stimuli (cf., Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013) and the assumption that participants in dyadic interactions would exhibit facial expressions of emotion (e.g., happy faces). Although increased upper face scanning in East Asian participants appears to contradict the notion of gaze avoidance, it is theoretically possible to find both greater upper face scanning and decreased eye contact in East Asian participants. Specifically, eye contact would additionally require *simultaneous* upper face scanning of both participants within a dyad, making eye contact dependent on temporal eye movement dynamics. However, given that current evidence on cultural differences in eye contact is limited, this question remained exploratory.

The above predictions were tested using regions-of-interest analyses (see Chapter 2). As explained in Chapter 2, regions-of-interest analyses are statistically sensitive for detecting underlying differences due to the spatial pooling of eye tracking data, and therefore provide a useful approach when specific predictions are in place. However, given the novelty of this cross-cultural study, additional Monte Carlo permutation tests (see Chapter 2 for more details) were employed to examine face scanning in a more spatially sensitive manner. This served to examine cultural differences in face scanning in an exploratory fashion that did not restrict analysis to pre-defined areas, thereby allowing for potential new insights into scanning strategies within naturalistic social contexts. Finally, in line with previous face-to-face interaction paradigms that recorded autistic traits (e.g., Freeth et al., 2013; Hessels, Holleman, Cornelissen, Hooge, & Kemner, 2018; Vabalas & Freeth, 2016), this study also included the Autism Quotient (AQ) (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; see Appendix A) to ensure that any cultural differences in face orienting or scanning were not solely driven by differences in autistic traits. The AQ is a 50-item self-administered questionnaire available in various languages including English, Japanese, and Chinese (i.e., the native languages of the participants in the current study). Each of the five subscales – namely social skill, attention to detail, attention switching, imagination, and communication (Baron-Cohen et al., 2001) – is assessed with ten items, and possible responses include 'definitely agree', 'slightly agree', 'slightly disagree', and 'definitely disagree'. One point is given for responses that indicate atypical behaviour, and scores are then summed to calculate the total score for every participant. Possible outcome scores therefore range from 0 to 50. Although the AQ does not serve as a diagnostic tool, 80% of autistic adults but only 2% of control participants obtained a score of at least 32 points (Baron-Cohen et al., 2001), making this a suggested cut-off threshold. The AQ was included in the current study based also on previous findings demonstrating increased autistic traits in East Asians relative to Western Caucasians (Kurita, Koyama, & Osada, 2005; Wakabayashi, Baron-Cohen, Wheelwright, & Tojo, 2006). Given that previous screen-based studies have revealed a relationship between increased autistic traits and reduced face and eye scanning (e.g., Chen & Yoon, 2011; Freeth et al., 2013), this would suggest decreased upper face scanning in East Asian participants, contrary to the current predictions outlined above. However, several live dyadic interaction studies were not able to replicate this relationship between autistic traits and face scanning strategies (e.g., Freeth et al., 2013; Vabalas & Freeth, 2016; but see Hessels et al., 2018), raising the possibility that these screen-based findings may not necessarily generalise to naturalistic social contexts. For the current study, it was therefore predicted that although autistic traits scores would differ between cultural groups, a relationship between autistic traits and face scanning behaviour was not expected.

3.3. Methods

3.3.1. Participants

Forty East Asian and 40 Western Caucasian adults were tested at Birkbeck, University of London. Every participant took part in one dyadic interaction, creating a sample of 20 dyads per cultural group. Two individuals were paired to form a dyad if they spoke the same native language and if they signed up for the same experimental session. An additional six Western Caucasian dyads and two East Asian dyads were excluded due to corneal reflection track loss of at least one participant within a dyad ($N = 7$) or misunderstanding of task instructions ($N = 1$).

To account for possible gender differences both in gaze scanning within one culture (Shen & Itti, 2012) and in face scanning between cultures (Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013), groups were gender-matched. Specifically, each cultural group consisted of 10 same-gender dyads (5 male-male, 5 female-female) and 10 mixed-gender dyads. Each cultural group therefore consisted of 10 females interacting with another female participant, 10 males interacting with another male participant, 10 males interacting with a female participant, and 10 females interacting with a male participant.

Western Caucasian participants (*M* = 26.95 years, *SD* = 9.12 years, 18-53 years) were born and raised in the UK, Ireland, USA, or Canada, were of White ethnicity, had never lived in a country outside Western Europe, USA, or Canada, and indicated English as their native language. East Asian participants (*M* = 26.35 years, *SD* = 7.14 years, 18-55 years) were born and raised in Mainland China or Japan, were of either Chinese or Japanese ethnicity, had never lived in a country outside East Asia before coming to the UK, and indicated Mandarin or Japanese as their first language. To minimise possible acculturation effects, only East Asian participants were included in the study who recently immigrated to the UK (stay in the UK: *M* = 5.7 months, *SD* = 2.1 months). The cultural groups did not differ in age ($t(78) = 0.33$, $p = 0.744$), and all participants had normal or corrected-to-normal vision and hearing. Participants were recruited using posters and participant databases of local universities. Given the challenge in recruiting East Asian participants who only recently immigrated to the UK, study adverts were also circulated in a local language school, on social networking platforms and community websites aimed at Japanese immigrants, and through word of mouth.

The study lasted approximately one hour, and each participant received £8 for their time. This study was approved by the local ethics committee of the Department of Psychological Sciences, Birkbeck, University of London. Each participant provided written informed consent prior to the study.

3.3.2. Apparatus

Eye movements were recorded using two Positive Science head-mounted eye trackers (www.positivescience.com), at a sampling rate averaging 30Hz. Since only one adult and one infant headset was available, the infant headgear was adapted for use in adults (see Figure 3.1 on page 95). The headgears only differed in their physical set-up, with all cameras, optics, and illuminators identical and therefore not differentially affecting data quality. The adult headgear consisted of glassless frames while the 'infant' headset was mounted using an elastic band. Each headset included an infra-red LED, one eye camera for monocular gaze tracking, and a scene camera fitted with a wide-angle lens (field-of-view 84.28° horizontally and 69.25° vertically). Scenes were recorded at an average of 30 frames per second and at 640 x 480 resolution. Each eye tracker was connected to a MacBook that recorded and saved the data, and an additional laptop was used to transmit sound to the neighbouring room to monitor participants.

Figure 3.1. Positive Science headset for adults (left) and adapted 'infant' headgear (middle, right).

3.3.3. Procedure

Each dyad was welcomed in the preparation room where the experimenter explained the study description, presented the eye tracker, and collected written informed consents. The experimenter spoke in English, but participants were instructed to communicate only in their native language with each other. To ensure a naturalistic interaction, participants were informed that the content of their speech would not be used for analysis. The participants were also informed that the current study aimed to examine cultural differences in face perception, but it was not made explicit that face scanning strategies were being investigated. Informal interviews after the testing session confirmed that participants were not aware of the present study aims. After the experimenter mounted the headsets, the dyad was guided to the adjacent testing room and participants were asked to sit at a table opposite each other at approximately 1 metre distance. While the experimenter prepared the recording set-up, participants were asked to complete a demographic questionnaire.

To complete a five-point calibration procedure, each participant was asked to fixate a small object held by the experimenter who was monitoring the calibration process using a MacBook. Prior to each experimental task, re-calibration was performed to avoid any drift. After successful calibration, dyads received task instructions in written form and in their native language. This ensured that all participants fully understood the procedure. Task instructions were translated from English to Mandarin or Japanese by a relevant native speaker and proof-read by an additional native speaker. Once participants finished reading the instruction for the first task, a clapperboard was used within the dyad's field-of-view to synchronise the eye tracking data in the analysis stage, and syncing was repeated prior to every experimental task. The experimenter left the room prior to the start of each task to ensure that the dyadic interaction was not influenced by a third person, and returned after each task for re-calibration, re-syncing, and to provide instructions for the next task. The experimenter remained silent throughout the study, providing only instruction sheets, except when a participant raised a question.

For the first task (*Introduction)*, participants were asked to introduce themselves and were encouraged to speak for at least 30 seconds to obtain sufficient data for analysis. Participants were instructed that they could mention their name, occupation, or hobbies, but that they were free to talk about anything as long as they were comfortable with sharing their personal details with their conversational partner. Dyads were free to have a conversation following the introduction task. The second task (*20 Questions)* consisted of two rounds of a guessing game in which one participant thought of an object while the other asked up to 20 questions to guess the object. Only *'Yes'*, *'No'* or *'I don't know'* were permitted answers. If a participant guessed the object correctly before reaching 10 questions, the experimenter returned and asked the dyad to play an additional round of *20 Questions*. This game was included in the paradigm to ensure that participants felt comfortable: since *20 Questions* has a very clear structure, it leads more easily to a dyadic interaction than a natural, free conversation. For the current study, face scanning strategies were not investigated for the *20 Questions* game. In the third task (*Story-telling*), each participant picked a coin from the table, looked at the year shown on the coin, and told the other participant about a personal event or experience that happened in that year. If participants could not remember a specific event or experience, they were free to talk about one from the year before or after. As with the first task, participants were asked to talk for at least 30 seconds to obtain sufficient data. After the third task, the experimenter stopped the recording. Finally, participants were asked to complete the Autism Quotient (AQ) (Baron-Cohen et al., 2001; see Appendix A), which was provided in the participants' native language and served to examine the relationship between face scanning and autistic traits in the general population.

3.3.4. Data pre-processing

3.3.4.1. *Data cleaning*

Yarbus (Positive Science) is a software that determines gaze coordinates by automatically tracking the pupil centre and corneal reflection in the eye video. If automatic tracking fails to accurately detect features in a frame, the bounding edges of the pupil and corneal reflection can be marked up manually for each frame. For the current study, manual selection was applied when the pupil centre or corneal reflection were positioned two or more pixels from the true centre (see Figure 3.2 on page 98; cf., Neesgaard, Senju, & Smith, 2017).

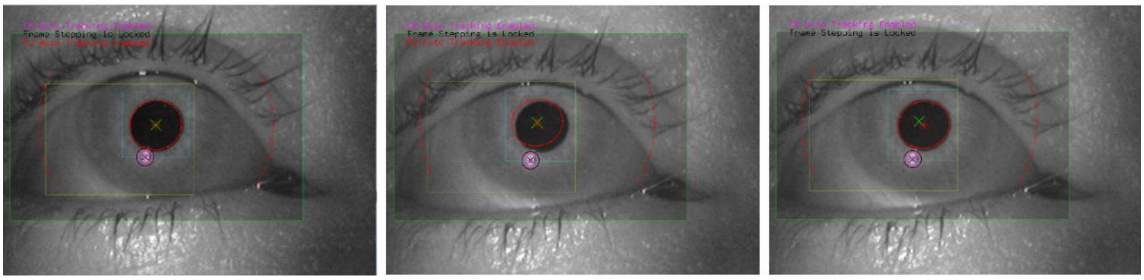


Figure 3.2. Correct automatic feature tracking (left), inaccurate pupil tracking (middle), and pupil centre position after manual correction (right).

3.3.4.2. *Coding of speaking and listening periods*

Separately for the introductory and story-telling task, the start time of the speaking period was defined as the first frame that contained audible speech (from the speaker) after the beginning of the task. Similarly, the start time of listening was coded as the equivalent time point in the listener's footage. The end time was determined when a participant stopped speaking. If participants spoke for longer than 30 seconds, their speech was cropped to ensure that participants contributed a similar amount of data (Freeth, Foulsham, et al., 2013). The start and end times of the listener were coded accordingly. If participants interrupted each other, the end time was counted as the last frame just preceding the interruption. Words or sounds such as "mhm" or "yeah" were not counted as interruptions. However, if the interruption occurred prior to a speaking/listening duration of 20 seconds, a second start and end time was used. Specifically, the second start time was coded as the start of the second sentence once the speaker resumed speaking. The reason for choosing the start of the second sentence rather than the first was based on the findings by Ho et al. (2015), who showed that individuals engaged in gaze aversion immediately after the start of speech. No third start/end times of speech were required or used.

3.3.4.3. *Regions-of-interest coding*

Face regions were coded semi-automatically using the detection and tracker tool presented in Chapter 2, with face regions being divided into upper and lower face.

3.4. Results

3.4.1. Regions-of-interest analysis

A 2 (Group: Western Caucasian, East Asian) x 2 (Speech: speaking, listening) x 2 (Task: introduction, story-telling) mixed ANOVA was conducted separately for face dwell time proportional to valid recording time, and upper face dwell time proportional to face looking overall. The data associated with speaking and listening periods were skewed in opposite directions and could not be corrected using data transformations, thereby violating the assumption of normality. Given that no equivalent non-parametric version exists, the 2 x 2 x 2 mixed ANOVA was conducted, and any significant effects were followed up or confirmed using appropriate non-parametric tests.

Dwell time on the face was calculated as a proportion to valid overall dwell time (with a cut-off at 30 seconds per task). Valid dwell time did not include periods of data loss (due to, e.g., blinks) to ensure that only periods were included for which it was known whether or not participants engaged in face looking. No significant group differences were found for data loss (East Asians: $M = 17.35\%$, $SD = 8.61\%$; Western Caucasians: $M = 16.88\%$, $SD = 12.03\%$; $t(78) = 0.20$, $p = 0.843$), indicating that both groups contributed a similar amount of data for the present analysis.

3.4.1.1. *Face looking*

Both groups looked more at the face of their conversational partner during listening compared to speaking periods (Table 3.1 on page 101). Dwell time on the face was additionally greater during the introduction task compared to the story-telling task (Table 3.1 on page 101). These patterns were reflected in a significant main effect of

Speech ($F(1,78) = 278.50$, $p < 0.001$, $\eta_p^2 = 0.781$; confirmed using the Wilcoxon Signed-Rank Test: $Z = -7.68$, $p < 0.001$, $r = 0.607$), and Task ($F(1,78) = 85.49$, $p < 0.001$, $\eta_p^2 = 0.523$; confirmed using Wilcoxon Signed-Rank Test: $Z = -6.71$, $p < 0.001$, $r = -0.530$). East Asian participants overall also engaged in significantly more face looking than Western Caucasians (Group: $F(1,78) = 5.70$, $p = 0.019$; $\eta_p^2 = 0.068$; confirmed using the Mann Whitney U Test: $U = 502$, $p = 0.004$, $r = -0.321$). A significant Speech x Group interaction was also found ($F(1,78) = 6.40$, $p = 0.013$, $\eta_p^2 = 0.076$). The interaction was followed up using the non-parametric Mann Whitney U Test, separately for each speech condition and by using a Bonferroni-corrected alpha-level of 0.025. The post-hoc comparisons revealed that, relative to the Western Caucasian group, East Asian participants looked significantly more at the face when speaking ($U = 533$, $p = 0.010$, $r = -0.287$), but no group difference was found for periods of listening ($U = 610$, $p = 0.068$). The analysis additionally revealed a significant Speech x Task interaction ($F(1,78) = 15.54$, $p < 0.001$, $\eta_p^2 = 0.166$). Follow-up analysis with a Bonferroni-corrected alpha-level of 0.025 showed that participants engaged in more face looking during the introduction task than the story-telling task, both when speaking (Wilcoxon Signed-Rank Test: $Z = -6.40$, $p < 0.001$, $r = 0.506$) and when listening (Wilcoxon Signed-Rank Test: $Z = -4.81$, $p < 0.001$, $r = 0.380$). This reflected the task effect reported above, and also suggested that the difference in face looking was significantly different in the speaking condition relative to the listening condition. The remaining main effects and interactions were not significant (Task x Group: $F(1,78) = 3.59$, $p = 0.062$; Speech x Task x Group: $F(1,78) = 3.28$, $p = 0.074$). In sum, both cultural groups looked more at the face during periods of listening than speaking, and during the introduction task relative to the story-telling game. When speaking, the East Asian group also looked more at the face than Western Caucasian participants (see also Figure 3.3 on page 101).

Table 3.1. *Medians and interquartile ranges for face dwell time (in %).*

		East Asians	Western Caucasians
		Mdn (IQR)	*Mdn (IQR)*
Introduction	Speaking	68.85 (33.69)	52.35 (36.24)
	Listening	95.19 (6.64)	96.74 (5.47)
Story-telling	Speaking	50.59 (38.57)	32.60 (31.14)
	Listening	93.69 (12.94)	87.84 (18.97)

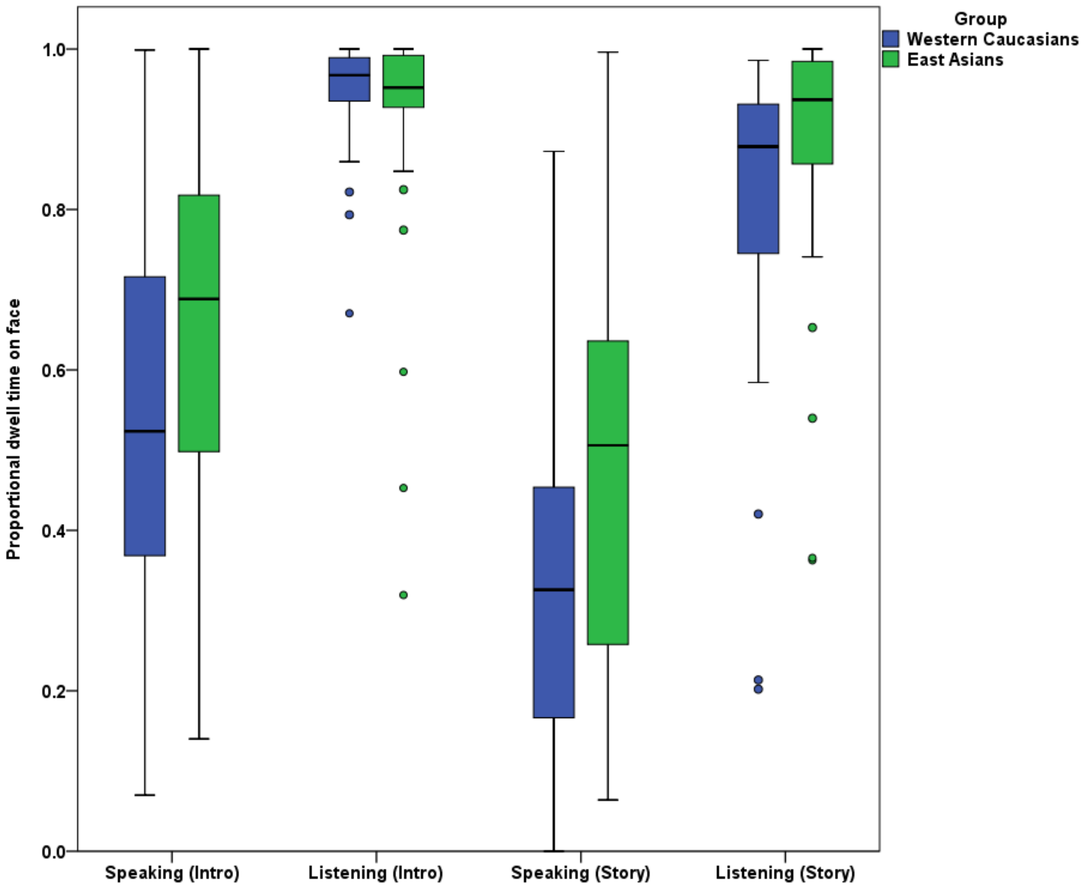


Figure 3.3. Proportional face dwell time for Western Caucasian and East Asian participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points).

3.4.1.2. *Upper face looking*

Upper face looking time was defined as the proportion of upper face dwell time relative to overall face dwell time. Unlike the results for face looking, the only significant effect for proportional dwell time of upper face scanning was Task, with greater upper face looking in the story-telling game than introduction task ($F(1,78) = 15.46$, $p < 0.001$, $\eta_p^2 = 0.165$; confirmed using Wilcoxon Signed-Rank Test: $Z = -4.11$, $p < 0.001$, $r = 0.325$; see Table 3.2). No other main effects or interactions were found (Speech: $F(1,78) = 3.56$, $p = 0.063$; Group: $F(1,78) = 0.16$, $p = 0.689$; Speech x Group: $F(1,78) = 0.045$, $p = 0.832$; Task x Group: $F(1,78) = 0.32$, $p = 0.576$; Speech x Task: $F(1,78) = 1.03$, $p = 0.314$; Speech x Task x Group: $F(1,78) = 0.55$, $p = 0.463$). In sum, both groups spent a significantly greater proportion of face looking time scanning the upper face region during the story-telling game than the introduction task (see Figure 3.4 on page 103).

Table 3.2. *Medians and interquartile ranges for upper face dwell time (in %).*

		East Asians	Western Caucasians
		Mdn (IQR)	*Mdn (IQR)*
Introduction	Speaking	74.34 (55.16)	76.15 (36.33)
	Listening	79.62 (36.73)	76.13 (28.41)
Story-telling	Speaking	87.63 (30.79)	91.33 (26.87)
	Listening	86.18 (22.02)	90.54 (22.40)

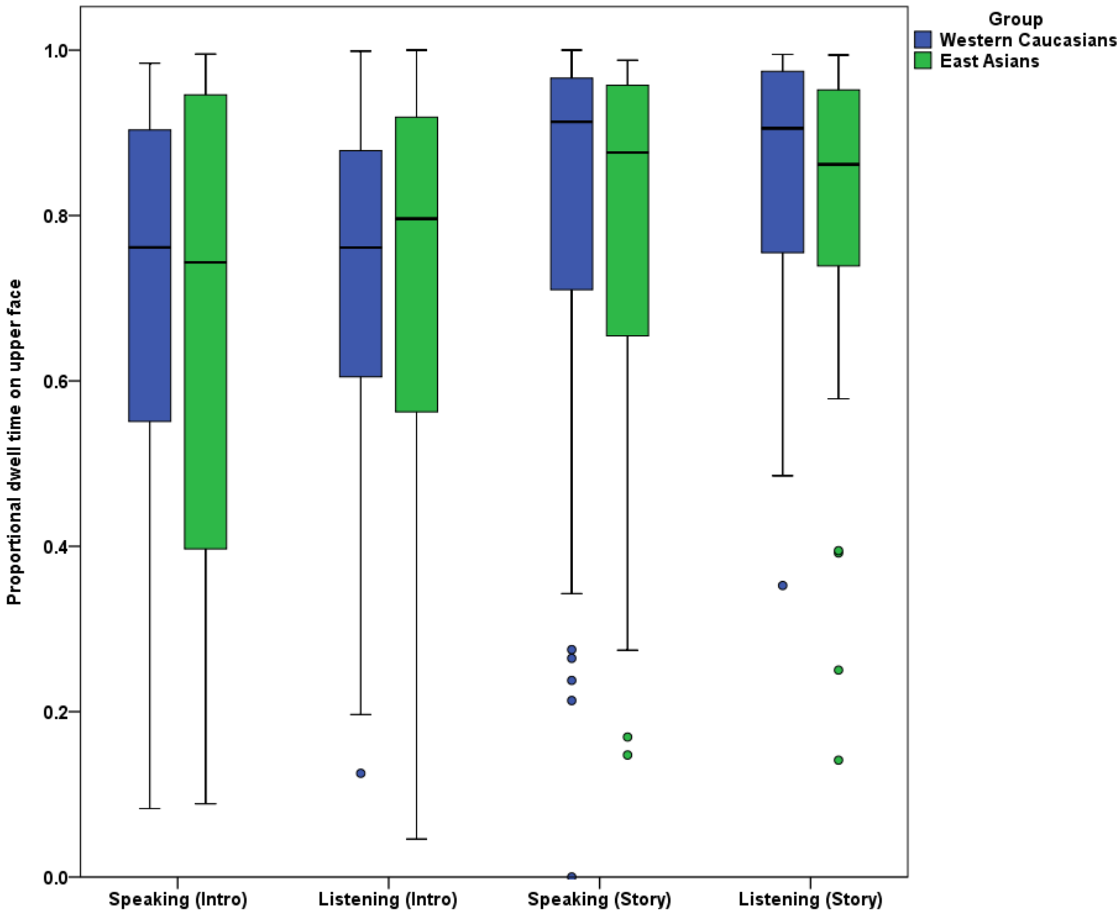


Figure 3.4. Proportional upper face dwell time for Western Caucasian and East Asian participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points).

3.4.2. Monte Carlo permutation test

To explore face scanning strategies in a spatially sensitive manner, Monte Carlo permutation tests (see Chapter 2) were conducted using the *CoSMoMVPA* toolbox (Oosterhof et al., 2016) and *FieldTrip* toolbox (Oostenveld et al., 2011). Analysis was performed separately for speaking and listening periods. The task effect found for upper face looking using the regions-of-interest analysis (see Section 3.4.1.2) also indicated that face scanning patterns differed between the introduction and story-telling task. The

Monte Carlo permutation tests were therefore conducted separately for each experimental task. Parameters were set to the following:

- an uncorrected *p*-value threshold of 0.01, and
- 10,000 iterations.

Figure 3.5 (below) and Figure 3.6 on page 105 illustrate descriptive heat maps showing group differences in gaze density for the introduction task and story-telling game, with a tendency of more looking at the eye region in East Asian adults and left face scanning for Western Caucasians.

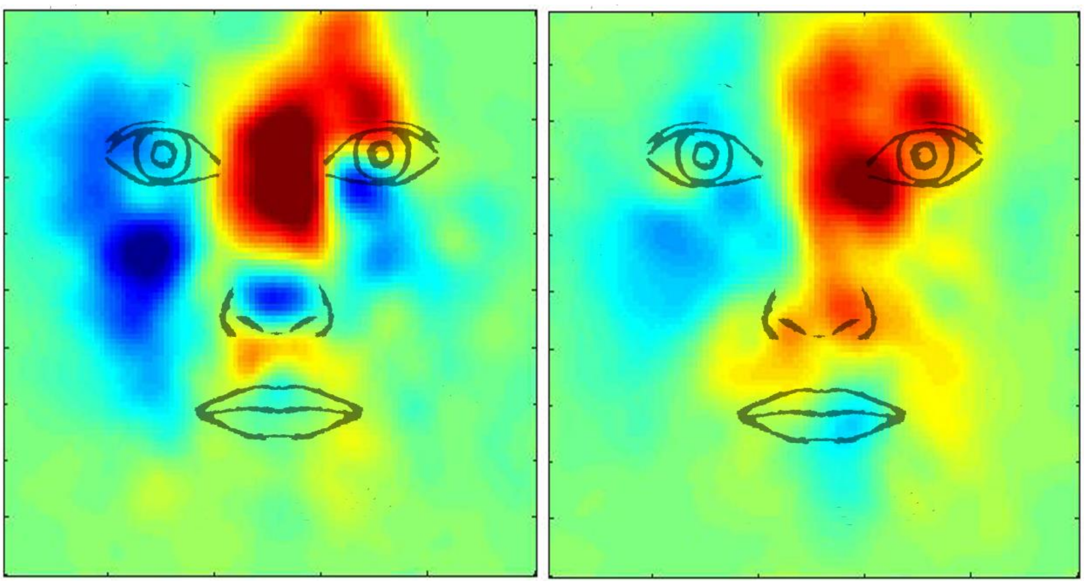


Figure 3.5. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for the introduction task, with red and blue colours depicting regions that East Asians and Western Caucasians scanned more, respectively.

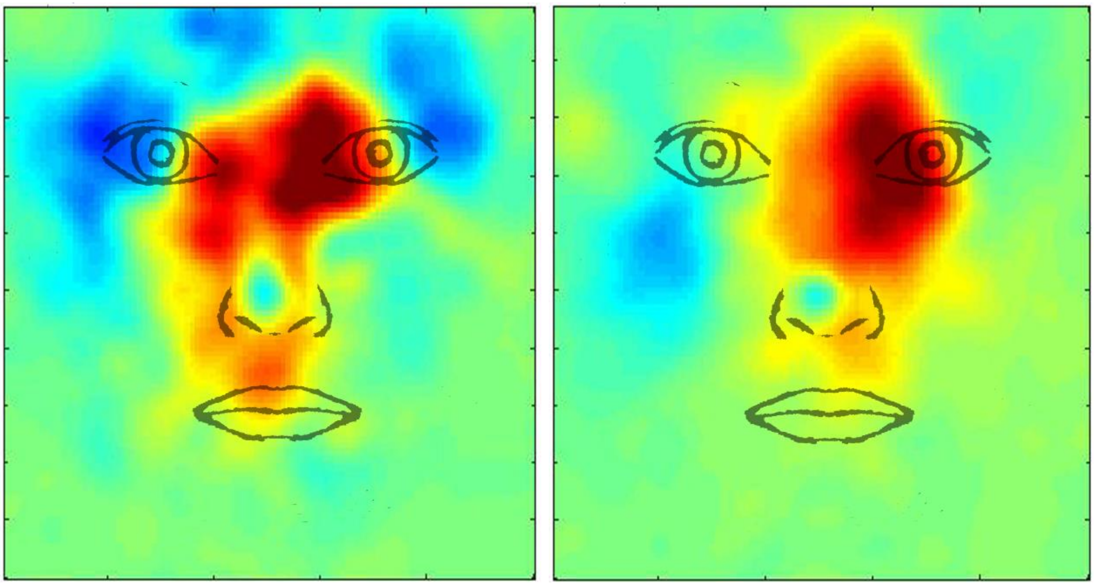


Figure 3.6. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for the story-telling game, with red and blue colours depicting regions that East Asians and Western Caucasians scanned more, respectively.

The statistical analysis revealed significant clusters for speaking but not listening periods (see Figure 3.7 and Figure 3.8, both on page 106, which represent the significance test). When speaking, East Asian participants scanned the eye region – particularly the area between the eyes – more than the Western Caucasian group for both the introduction and story-telling task. In contrast to the current predictions, Western Caucasian participants looked more at the left side of the face (from the observer's perspective) for the introduction task. This is in contrast with the findings from the regions-of-interest analysis, which revealed no group differences. However, it is possible that the gaze points from the Western Caucasian participants sufficiently fell into the upper face region so that less spatially sensitive analyses could not detect any such differences.

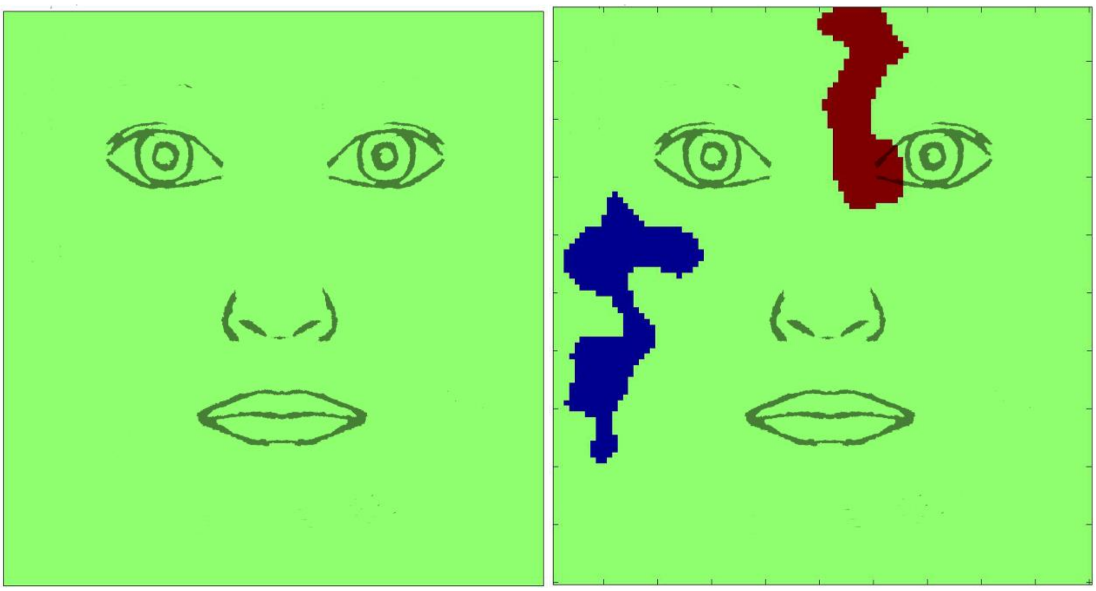


Figure 3.7. Gaze clusters for the introduction task during listening (left) and speaking periods (right), illustrating regions that were scanned significantly more by Western Caucasian (blue) and East Asian participants (red).

Figure 3.8. Gaze clusters for the story-telling game during listening (left) and speaking periods (right), illustrating regions that were scanned significantly more by Western Caucasian (blue) and East Asian participants (red).

3.4.3. Mutual gaze

Mutual gaze was defined as those periods during which both participants within the same dyad simultaneously looked at the upper face region. To code mutual gaze, each dyad's timelines were first synchronised by pairing entries that were closest in time. The start of a single instance of mutual gaze was considered to be the timing at which both participants of a dyad looked at the upper face region. The end of this mutual gaze instance was coded as the timing when at least one of the participants blinked or shifted their gaze away from the upper face region. Given that entries in the upper face looking timeline were coded '1' when participants scanned the upper face region and '0' when participants looked away from the upper face region (see Chapter 2), the entries of the dyad's timelines were simply multiplied. This resulted in a new timeline which denoted periods of mutual gaze as '1' and periods without mutual gaze as '0'. A 2 (Group: Western Caucasian, East Asian) x 2 (Task: introduction, story-telling) mixed ANOVA was conducted on mutual gaze time proportional to valid recording time. The factor Speech was not included in the model given that mutual gaze depended on both participants within a dyad, i.e. periods could not be split into speaking and listening periods. Unlike the ANOVAs conducted above (Sections 3.4.1.1 and 3.4.1.2), the data was normally distributed so that any significant effects did not need to be confirmed or followed up using non-parametric tests.

East Asian dyads numerically seemed to engage in more mutual gaze than the Western Caucasian group for both experimental tasks (see Table 3.3 on page 108). However, the analysis revealed no significant main effects and no significant interaction (Task: $F(1,38) = 1.30$, $p = 0.262$; Group: $F(1,38) = 3.02$, $p = 0.090$; Task x Group: $F(1,38) = 0.44$, $p = 0.512$).

Table 3.3. *Time spent on mutual gaze proportional to valid recording time (in %).*

	East Asian *Mean (SD)*	Western Caucasian *Mean (SD)*
Introduction	29.96 (18.61)	23.43 (11.92)
Story-telling	29.02 (18.31)	19.86 (12.19)

To examine whether the mean durations of mutual gaze instances differed between cultural groups, an additional 2 (Group: Western Caucasian, East Asian) x 2 (Task: introduction, story-telling) mixed ANOVA was conducted. The findings revealed no significant effects (Task: $F(1,38) = 2.85$, $p = 0.099$; Group: $F(1,38) = 0.01$, $p = 0.907$; Task x Group: $F(1,38) = 0.17$, $p = 0.682$; see also Table 3.4).

Table 3.4. *Means and standard errors for the mean durations of dyad's mutual gaze occurrence (in milliseconds).*

	East Asian *Mean (SE)*	Western Caucasian *Mean (SE)*
Introduction	152.96 (17.34)	149.98 (15.55)
Story-telling	171.02 (16.70)	179.71 (27.56)

3.4.4. Autism Quotient (AQ)

Western Caucasian participants obtained a significantly lower score on the AQ ($M = 14.88$, $SD = 6.97$, ranging from 5 to 36) than the East Asian group ($M = 18.05$, $SD = 5.49$, ranging from 7 to 35; $t(78) = 2.26$, $p < 0.026$, $d = 0.51$). Given the wide range for (upper) face looking measures and also AQ scores, a correlational analysis was performed to investigate the relationship between AQ scores and dwell time on the face or upper face. The correlational analysis was conducted separately for each group to account for the reversal paradox, which refers to a reversal in the directional relationship between two variables when pooling data from different experimental groups (see e.g., Kievit, Frankenhuis, Waldorp, & Borsboom, 2013). The correlational analysis was conducted separately for the introduction and story-telling task given that significant differences in face and upper face dwell time were found between those tasks. Results showed no significant correlations between AQ and (upper) face looking (all $p > 0.05$; see Table 3.5 for correlation coefficients).

Table 3.5. *Pearson's correlation coefficients for the relationship between AQ scores and (upper) face dwell time during speaking and listening periods.*

		East Asians	Western Caucasians
Speaking	AQ – Face looking (Intro)	-0.089	0.092
	AQ – Face looking (Story)	0.126	0.162
	AQ – Upper face looking (Intro)	0.033	0.236
	AQ – Upper face looking (Story)	-0.083	0.170
Listening	AQ – Face looking (Intro)	-0.254	-0.192
	AQ – Face looking (Story)	-0.098	-0.092
	AQ – Upper face looking (Intro)	0.053	0.111
	AQ – Upper face looking (Story)	-0.048	-0.014

3.5. Discussion

Cultural differences in face scanning have so far only been found using screen-based paradigms that lack the visual complexity and social presence common to naturalistic conditions. Given that ultimately the aim is to understand and explain behaviour in the 'real world', interpretations cannot be limited to screen-based paradigms. The current study therefore investigated cultural differences in scanning strategies of faces during live dyadic interactions, with findings summarised and discussed below.

The regions-of-interest analysis investigated the extent to which Western Caucasians and East Asians spontaneously oriented toward the face, while the data-driven approach served as a more spatially detailed investigation of participants' face scanning strategies. As predicted, both cultural groups looked more at the face of their conversational partner during periods of listening compared to periods of speaking. This is consistent with previous findings (e.g., Freeth et al., 2013) and – combined with the associated large effect size found in the current study – implies a robust speech effect. Several explanations could account for the difference in the amount of face orienting between periods of speaking and listening. Previous studies have demonstrated increased gaze aversion during cognitively demanding periods, possibly in order to reduce cognitive load (Doherty-Sneddon & Phelps, 2005). The decrease in face looking when speaking could therefore have served to reduce load given that periods of speaking were likely to be more cognitively demanding than periods of listening. Furthermore, greater face orienting behaviour during listening periods could have facilitated participants' decoding of speech (Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998) as well as served as a social signal to convey to the conversational partner that one is still listening (Risko et al., 2016). With respect to task effects, dwell time on the face was greater for the introduction task relative to the story-telling game. This could be explained by the task order whereby participants may have exhibited increased social signalling during the start of the testing session (i.e., the introduction period) when they were less familiar with each other. Another explanation could be the nature of the tasks; specifically, the story-telling game likely required greater cognitive effort since participants had to recall and describe a past event or experience. Gaze aversion may have helped to reduce cognitive load and in turn facilitated participants' memory recall (Glenberg, Schroeder, & Robertson, 1998), giving rise to the decreased face looking during the story-telling task compared to the introduction period. With respect to cultural differences, East Asian participants looked significantly more at the face of their conversational partner than the Western Caucasian group. This is in contrast with the prediction that no group differences in the amount of face orienting behaviour would exist. However, since this cultural difference in face orienting was only found for periods of speaking, but not listening, it raises the possibility that the Western Caucasian participants more likely averted their gaze away from the face in order to reduce cognitive load. No study to date has directly examined cultural differences in cognitive load, however, and this interpretation therefore remains speculative. Alternatively, increased face orienting in East Asian participants relative to the Western Caucasian group could indicate a greater tendency to socially signal to the conversational partner that one is still engaged in the conversation (Risko et al., 2016). Although this explanation should then also hold for periods of listening for which no cultural effect could be found, face orienting when listening to the conversational partner was characterised by a ceiling effect in both groups. This could therefore have masked any increased tendency for social signalling in the East Asian group. Such increased social signalling in East Asians could reflect an underlying cultural difference in face-to-face interactions or, alternatively, could have been specific to immigrant populations. In particular, the East Asian participants in the current study had only very recently moved to the UK, and it is possible – given the challenges for immigrants such as social isolation, language difficulties, or cultural assimilation (Stewart et al., 2008) – that their social interactions were qualitatively different from those of local citizens. Future studies will need to account for this by studying non-immigrant populations.

The regions-of-interest analysis additionally revealed no cultural differences in upper face dwell time, with both groups showing increased scanning in the story-telling game compared to the introduction task. Based on the regions-of-interest analysis, the present results do not support the prediction that the East Asian group would exhibit more upper face scanning than the Western Caucasian participants. However, the findings from the data-driven method showed significant gaze clusters for periods of speaking. Although no clusters were revealed for periods of listening – gaze was less spatially clustered compared to speaking periods – the descriptive heat maps (Figure 3.5

on page 104 and Figure 3.6 on page 105) illustrate that gaze distribution was similar across the speech conditions. For periods of speaking, the permutation test revealed a gaze cluster indicating that East Asian participants scanned the eye region – and specifically the area between the eyes – significantly more than the Western Caucasian group. The Western Caucasian participants, meanwhile, spent significantly more time looking at the left side of the face (from the observer's perspective) for the introduction task. Importantly, the associated gaze cluster stretched across the upper and lower face regions (see Figure 3.7 on page 106), such that the regions-of-interest approach was likely not spatially sensitive enough to capture this difference in scanning behaviour. This emphasises both the limitation of the regions-of-interest approach that makes *a priori* assumptions about the areas to include for analysis, and the usefulness of the data-driven method for exploratory purposes.

The increased scanning of the left side of the face observed in the Western Caucasian group during speaking periods in the introductory phase indicates that participants did not exhibit greater mouth looking as predicted. Furthermore, the findings suggest that the widely reported left-side bias in face perception (see Rhodes, 1985) may be more likely manifested in Western Caucasians. However, this interpretation needs to be treated cautiously for two reasons. First, although the descriptive heat map (Figure 3.6 on page 105) illustrates a weak gaze cluster on the left side of the face for the story-telling game, the effect could not be statistically confirmed (see Figure 3.8 on page 106) and therefore was not replicated. Secondly, the eye tracking hardware could have modulated scanning behaviour. Specifically, the eye camera arm of the Positive Science gear was located on the left side of the face (from the observer's perspective), raising the possibility of a systematic group difference in the extent to which participants were visually distracted by the hardware. This visual distraction would also provide a possible explanation for the absence of predicted mouth looking in the Western Caucasian group. Greater scanning of the left side of the face in Western Caucasians as a result of orienting toward the camera arm is also consistent with the scene perception literature, which demonstrated a significantly higher tendency for Western Caucasians to fixate focal, salient objects as opposed to East Asians who exhibit a more holistic scanning pattern (Chua et al., 2005). To make a stronger conclusion about cultural differences in face scanning strategies, further research will therefore be required that takes into account the possible interference of eye tracking hardware within face-to-face interaction paradigms.

The current study also examined cultural differences in mutual gaze given previous reports suggesting East Asians to engage in less eye contact than Western Caucasian individuals (Argyle et al., 1986; A. McCarthy et al., 2006, 2008). The regions-of-interest analysis revealed no significant cultural differences both in the time spent on mutual gaze overall and the mean duration of each instance of mutual gaze. Based on these results, the notion of gaze avoidance reported for East Asians could not be supported. As outlined in the introduction of this chapter, current evidence in support of such gaze avoidance is limited to self-reports (Argyle et al., 1986) or video recordings of dyadic interactions during which participants were engaged in abstract thinking (A. McCarthy et al., 2006, 2008), and it is therefore possible that cultural differences in eye contact behaviour may not exist for dyadic interactions. Alternatively, the results from the permutation test could suggest an explanation that would support both the notion of gaze avoidance in East Asians and the current regions-of-interest findings showing no cultural effect. Specifically, the data-driven method revealed that the East Asian group allocated more visual attention *between* the eyes when looking at the face. By scanning the intermediary region, East Asian participants would still be able to extract visual information from the eye region (cf., Jack et al., 2009) using their peripheral vision and possibly also signal to the other person that they were still engaged in the dyadic interaction (Risko et al., 2016), while simultaneously avoiding direct eye contact. Given that the data-driven findings were collapsed across time and thus cannot indicate precisely where East Asian participants looked in the face at the time of mutual gaze, further evidence would be required to confirm this interpretation.

In addition, the mean durations of each instance of mutual gaze in the current study ranged between 150 and 180 milliseconds. Given that each frame was manually corrected if automatic tracking of the pupil centre or the corneal reflection failed (see Section 3.3.4.1), these short durations cannot be due to flicker or other instabilities during data recording. The mutual gaze durations thereby represented significantly shorter durations than those reported in previous studies (e.g., Argyle et al., 1986; Binetti, Harrison, Coutrot, Johnston, & Mareschal, 2016). Binetti et al. (2016), for instance, found a preferred mean mutual gaze duration of 3.3 seconds when participants were asked to indicate their level of (dis)comfort while maintaining eye contact with a dynamic face identity displayed on a screen. Two possible explanations could account for the short mutual gaze duration in the current study. First, unlike screen-based paradigms, participants in the present paradigm were engaged in a live dyadic interaction and the social presence of the conversational partner may have influenced mutual gaze behaviour. This idea is supported by findings from a live dyadic interaction study with Western Caucasian participants that found eye contact to last for an average of 360 milliseconds (Rogers, Speelman, Guidetti, & Longmuir, 2018), i.e. a mean duration that is more consistent with the current results. Secondly, the end of a single instance of mutual gaze in the present study was defined when one participant within a dyad averted the gaze away from the face or when an eye blink occurred. Eye blinks were considered in this study since this indeed represents a temporary break in eye contact. Furthermore, eye blinks have not only been found to function as a purely biological mechanism (e.g., to protect the corneal surface; Ousler, Hagberg, Schnindelar, Welch, & Abelson, 2008), but have also been linked to cognitive processes (Hirokawa, Yagi, & Miyata, 2004; Holland & Tarlow, 1975) and are relevant in social interactions (Hömke, Holler, & Levinson, 2017). For instance, longer blinks more likely occur during mutual gaze and serve as social signals to indicate understanding (Hömke et al., 2017). Since previous studies measuring eye contact durations did not consider blinks as an end marker (e.g., Rogers et al., 2018), this likely increased the mean durations of mutual gaze.

In light of the possible confounding role of the eye tracker, further studies will be required to ensure that mutual gaze behaviour was not influenced by equipment hardware. The findings currently imply, however, that the widely reported notion of gaze avoidance in East Asian participants may need to be refined and that mutual eye gaze during face-to-face interactions occurs only very briefly.

The present study also examined whether a relationship existed between AQ scores and face orienting and scanning behaviour to ensure that any cultural effects were not simply modulated by autistic traits. In line with previous studies (Kurita et al., 2005; Wakabayashi et al., 2006) and with the current predictions, cultural differences were found for AQ scores, with East Asian participants obtaining significantly higher scores than the Western Caucasian group. In contrast to previous screen-based studies that demonstrated a link between increased autistic traits and reduced face looking (e.g., Chen & Yoon, 2011; Freeth et al., 2013), no significant correlations were found between AQ scores and (upper) face scanning in the current study using a live dyadic interaction paradigm. However, the present results are consistent with previous studies that also adopted face-to-face interaction paradigms (e.g., Freeth et al., 2013; Vabalas & Freeth, 2016; but see Hessels et al., 2018), suggesting that the influence of autistic traits may not necessarily extend to naturalistic social contexts. Given the possibility that the eye tracking equipment may have affected face scanning strategies, however, these results would need to be replicated to make stronger conclusions. Nevertheless, the findings currently suggest that autistic traits unlikely modulated face orienting and scanning behaviour.

In sum, the current study revealed evidence for cultural influences on face orienting behaviour and further demonstrated some support for group differences in face scanning strategies during naturalistic social interactions. To examine face scanning strategies in each group, semi-automatic coding methods were applied to conduct traditional regions-of-interest analysis. In addition, a novel data-driven approach was employed, which allowed for a more refined interpretation of the present findings compared to the regions-of-interest analysis. East Asians were found to orient more to the face of their conversational partner than Western Caucasians. Although this may possibly suggest a greater tendency for social signalling in East Asians, future studies will need to replicate this finding. With respect to face scanning strategies, the regions-of-interest analysis could not reveal any cultural differences, while the data-driven method developed for this thesis indicated that East Asian participants scanned the region between the eyes more than Western Caucasians, who in turn exhibited a tendency to scan the left side of the face. This emphasises the strength of applying data-driven methods, particularly when only limited evidence exists to formulate specific predictions that ultimately determine which areas should be defined for regions-of-interest analyses. As discussed in this chapter, however, the current findings must be interpreted with caution given the possibility that the eye movements of Western Caucasians participants were modulated by the eye tracking hardware. To address this limitation, a follow-up study with improved methodology was conducted and will be presented in the next chapter.

Chapter 4

Cultural Differences in Face Scanning During Live Dyadic Interactions II

4.1. Chapter Overview

The study findings presented in Chapter 3 revealed some cultural differences in face orienting and scanning during dyadic social interactions. However, several methodological limitations in the face-to-face paradigm made data interpretation difficult. This chapter will present a follow-up study with improved methodology to gain a better understanding of cultural differences in naturalistic face scanning during dyadic interactions. To contextualise the current study, the following sections will briefly summarise the findings obtained in Chapter 3 before going on to outline the methodological changes that were implemented with respect to the equipment, paradigm, and participants.

4.2. Introduction

As outlined in Chapter 1, existing evidence on cultural differences in face scanning is predominantly restricted to screen-based paradigms that typically employ relatively simple stimuli such as static images of faces. To investigate whether and how such cultural differences manifest within more naturalistic settings, the study presented in Chapter 3 examined face orienting and face scanning during social interactions of Western Caucasian and East Asian dyads. The findings revealed that participants from both cultural groups engaged in more face looking during periods of listening compared to periods of speaking, and in more face looking during the introductory task compared to the story-telling game. With respect to cultural differences, East Asian participants were found to look significantly more at the face of their conversational partner than Western Caucasians, although only for periods of speaking. In addition, whereas the regions-of-interest analysis did not reveal any cultural differences in face scanning strategies (upper versus lower face), the permutation tests identified increased scanning of the area between the eyes during speaking periods in East Asian relative to Western Caucasian participants. For the introductory task, the Western Caucasian group additionally exhibited greater scanning of the left side of the face (from the observer's perspective).

However, as discussed in Chapter 3, several methodological factors may have affected face orienting and scanning behaviour, and some interpretations that emerged from the findings in Chapter 3 therefore required further clarification. The following sections will provide a more detailed discussion of the relevant methodological factors that may have modulated the findings obtained in Chapter 3, and will outline how the changes were implemented for the current study. Specifically, this included methodological changes with respect to the choice of eye tracker model (see Section 4.2.1), the paradigm – a research assistant (and not a participant) acted as the social interaction partner (see Section 4.2.2) – and the participant groups, which consisted of local residents in the UK and in Japan, i.e. only non-immigrant populations were considered (see Section 4.2.3). In addition, a social anxiety traits measure was included as part of the current study (see Section 4.2.4). The current study thus changed several methodological factors at once, rather than modifying one parameter at a time that would allow insight into underlying influences on face orienting or scanning behaviour. However, a number of considerations motivated the decision to change various parameters at once for the current study. First, the aim was not to identify the influence of each factor (e.g., examining the effect of eye tracking hardware on face scanning, or comparing immigrants with local residents with respect to face orienting and scanning), but to examine cultural differences in face scanning during social interactions. Secondly, as will be discussed in more detail below, some aspects needed to be addressed to improve data quality overall (e.g., a research assistant rather than a participant was employed to control for the distance between conversational partners; see Section 4.2.2). Finally, practical considerations must be taken into account; given the highly time- and resource-intensive nature of conducting cross-cultural social interaction studies with head-mounted eye tracking techniques, it was decided that the various methodological factors should be changed at once for the current study.

4.2.1. Equipment changes

In the previous study (Chapter 3), two Positive Science eye trackers were used for the participants in each dyad. The Positive Science system was chosen since the sets would not obstruct the face as considerably as other available models – a crucial property given that the study looked at face scanning during a dyadic interaction – and two sets of the same system were readily available. Results indicated that Western Caucasians tended to scan the left side of the face (from the observer's perspective) more than East Asian participants. Although this could suggest cultural differences in the widely reported left-side bias in face perception (see Rhodes, 1985), this finding may have been confounded by the nature of the Positive Science gear. Specifically, given that the Positive Science system records in a monocular fashion with the eye camera arm being located on the left side of the face (from the observer's perspective), it is possible that Western Caucasian participants were more visually distracted by the equipment, leading to increased looking toward the left side of the face. Since previous studies on scene perception have reported a significantly greater tendency for Western Caucasians – relative to East Asians – to exhibit analytic strategies characterised by scanning focal, salient objects as opposed to more context-dependent holistic strategies (Chua et al., 2005), the Positive Science gear may have modulated scanning behaviour differently between groups. To address this, participants in the current study interacted with a local research assistant whose eye movements were not recorded, i.e. no equipment was obstructing the face that the participants were scanning. Although a different head-mounted eye tracking model could alternatively have been worn by two participants, all available systems occlude the eye region and therefore could potentially give rise to differential scanning behaviour. A local research assistant was therefore chosen to interact with participants. Given that only the participants' eye movements were recorded, the current study solely examined face orienting and scanning, and not eye contact behaviour.

Another methodological limitation concerned the scene video quality and properties. The Positive Science system used in the previous study was equipped with a wide-angle scene camera at 640 x 480 resolution. Although the face of the conversational partner was usually located in the centre of the scene where visual distortion of wide-angle lenses is less impactful compared to the scene periphery, a normal lens would ensure that faces are not distorted at any time. This would also increase the size of the conversational partner's face and thereby allow for a more detailed examination of scanning behaviour. In the present study, participants' eye movements were therefore recorded using SMI eye tracking glasses (SMI ETG; SensoMotoric Instruments, Germany), which were equipped with a scene camera fitted with a normal lens. The SMI ETG also provided higher spatial resolution at 1280 x 960 and temporal resolution at 60 Hz (compared to 30 Hz for Positive Science), giving more refined eye movement data. Although the SMI ETG heavily obstructed the eye region, it represented an appropriate system for the current study given that only the participant – and not the local research assistant – was wearing the eye tracker. Finally, SMI's software *BeGaze* (SensoMotoric Instruments, Germany) can extract fixation measures in addition to dwell time measures, unlike Positive Science's *Yarbus* (see Chapter 3). Given that a) previous studies on cultural differences in face scanning have used either dwell time (Senju, Vernetti, Kikuchi, Akechi, & Hasegawa, 2013; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013) or fixation time (e.g., Blais et al., 2008; Jack et al., 2009; Wheeler et al., 2011) as dependent measures, and b) it has been shown that cultural effects in face scanning may be analysis-dependent (Arizpe et al., 2016), the current study employed both fixation and dwell time analyses to ensure that key results would be replicated in both measures.

4.2.2. Paradigm changes

A local research assistant – rather than another participant as in the previous study – was chosen to interact with the participant to control also for content of speech and distance. Local research assistants were instructed to keep the content of their speech similar across participants. Additionally, although participants in the previous study were asked to sit in chairs that were positioned to achieve a distance of 1 metre between individuals, participants occasionally leaned forward or backward, thereby increasing or decreasing the size of the face captured by the scene camera. To ensure that face size remained consistent across individuals and groups in the current study, the local research assistants were trained to maintain the distance to the participant by leaning forward or backward.

4.2.3. Participant changes

The previous study compared Western Caucasian and East Asian participants from diverse Western and East Asian nationalities. Although no empirical evidence is currently available, greater within-group variability could potentially mask differences between cultural groups. In addition, the East Asian participants represented an immigrant population and it is currently unknown to what extent and within which timeframe immigrant populations culturally acclimatise when interacting with local residents, raising the possibility that face scanning behaviour observed in the East Asian sample could have been modulated as a function of residency duration. Importantly, however, it is possible that the participants, who only very recently emigrated, interacted differently relative to East Asian individuals who reside in their home country. Given that immigrants often encounter various challenges including social isolation, language difficulties, or difficulties in cultural assimilation (Stewart et al., 2008), the comparable situations between immigrants may increase social coherence and lead to qualitatively different interactions. To account for possible effects specific to immigrant populations or due to acclimatisation, the present study introduced stricter inclusion and exclusion criteria for participants (see Section 4.3.1 for details), comparing Japanese individuals residing in Japan with British or Irish individuals residing in the UK.

4.2.4. Measurement of social anxiety traits

Previous screen-based studies have pointed to a possible relationship between higher social anxiety and reduced face and gaze scanning (Horley, Williams, Gonsalvez, & Gordon, 2003; Moukheiber et al., 2010). Studies have also suggested that social anxiety traits may differ between cultures (cf., Freeth, Bullock, & Milne, 2013; Sugawara et al., 2012), with Japanese individuals typically obtaining higher scores (i.e., showing more social anxiety traits). This raises the possibility that cultural differences in face scanning could arise from group differences in social anxiety traits. In other words, Japanese participants would be expected to score higher on social anxiety measures and therefore exhibit reduced face and/or gaze scanning. This prediction conflicts with findings on cultural differences in face scanning of emotionally expressive faces, however, which showed that Japanese participants engaged in greater eye looking behaviour (Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). In line with this conflict, Wieser, Pauli, Alpers, & Mühlberger (2009) showed a link between higher social anxiety traits and increased heart rate without gaze avoidance. In light of this contradictory evidence, and to ensure that any possible cultural differences in this study were not modulated by social anxiety traits, relevant measures were employed in an exploratory fashion, as in previous studies adopting live face-to-face paradigms (Freeth, Foulsham, et al., 2013). In this study, the self-report version of the Liebowitz Social Anxiety Scale (LSAS-SR; Liebowitz, 1987; see Appendix B for the English version) was administered in addition to the AQ. The LSAS is available in English and Japanese and has been validated as a measure when administered both in English and Japanese (Heimberg et al., 1999; Sugawara et al., 2012). The LSAS is a 24-item questionnaire divided into two subscales relating to performance anxiety (13 items) and anxiety in social situations (11 items). Each item is rated twice, namely on how much anxiety or fear is experienced in a given situation, and how often the situation is avoided. Ratings are on a scale between 0 (no fear or anxiety/no avoidance) and 3 (severe fear or anxiety/severe avoidance), and scores from both subscales are combined and summed to compute the total score for each participant. The possible outcomes therefore range from 0 to 144 points, with a recommendation to classify social anxiety scores as follows (Liebowitz, 1987): insignificant (<55 points), moderate (55-64 points), marked (65-79 points), severe (80-95 points), and very severe (95+ points).

4.2.5. The current study

The current study was based on the same rationale and predictions as in the previous experiment (Chapter 3). Based on earlier studies, it was expected that participants engaged in increased face looking during listening than speaking periods (cf., Freeth, Foulsham, et al., 2013). The findings in Chapter 3 suggested cultural differences in face orienting behaviour; however, various methodological factors, which were modified in the current study, may have affected study findings. For instance, it is possible that dyadic social interactions amongst immigrant populations were qualitatively different than those of local residents, thereby resulting in increased face orienting behaviours (see Section 4.2.3). As in Chapter 3, the current prediction regarding cultural differences in face orienting was therefore based on the existing literature. Specifically, given that there is no available evidence in the literature to suggest cultural differences in face looking, no group differences were expected. With respect to scanning strategies within the face, it was predicted that Japanese participants would exhibit more upper face looking than British/Irish participants (cf., Jack et al., 2009). Conversely, British/Irish participants were expected to engage in greater mouth looking (cf., Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). As outlined in this chapter, methodological factors may have modulated the face scanning strategies found in Chapter 3; for instance, the eye camera arm may have visually distracted participants. For this reason, the current predictions were not based on the findings obtained in Chapter 3, but they were formulated based on the evidence from screen-based studies with emotionally expressive face stimuli and based on the assumption that the research assistants in the current study would exhibit emotionally expressive faces during the dyadic interaction (e.g., happy faces).

The predictions for overall face looking and cultural differences in the scanning of specific face regions were tested using regions-of-interest analysis. As explained in Chapter 2, regions-of-interest analyses are statistically sensitive when testing specific predictions given the spatial pooling of eye tracking data. However, an additional Monte Carlo permutation test (see Chapter 2 for more details) was employed to investigate cultural differences in face scanning in an exploratory manner that did not restrict analysis to pre-defined areas. For all predictions, it was expected that the dwell time and fixation time analysis would result in similar findings. Finally, although cultural differences in AQ and LSAS scores were expected, these differences were not expected to modulate face scanning patterns.

4.3. Methods

4.3.1. Participants

Thirty-six British/Irish and 34 Japanese adults participated in this study. British/Irish participants were tested at Birkbeck, University of London (UK), and Japanese participants were tested at Kyoto University (Japan). Seven British/Irish participants were excluded due to flickering gaze data ($N = 6$) or because the face of the research assistant was not captured in the scene recording ($N = 1$; this can occur when the participant tilts the head downward while gazing at the research assistant). Seven Japanese participants were also excluded from analysis due to flicker ($N = 6$) or because the participant had previously lived in a Western country ($N = 1$). The drop-out rate (19.4% and 20.6% for British/Irish and Japanese participants, respectively) was higher than in typical screen-based eye-tracking experiments. However, this largely resulted from the naturalistic face-to-face interaction paradigm during which participants often smiled or laughed. This in turn occluded the pupil or corneal reflections and resulted in flickering gaze data. The final sample for analysis therefore included 29 British/Irish (13 female, 16 male) and 27 Japanese (14 female, 13 male) adults.

British/Irish participants (M = 28.07 years, SD = 6.60 years, 19-40 years) were born and raised in the UK or Republic of Ireland, were of White-British or White-Irish ethnicity, had never lived in a country outside Western Europe, the USA, or Canada, and indicated English as their first language. Japanese participants (M = 21.70 years, SD = 2.77 years, 18-31 years) were born and raised in Japan, were of Japanese ethnicity, had never lived in a country outside East Asia, and indicated Japanese as their first language. Although the two cultural groups significantly differed in age ($t(54) = 4.64$, $p < 0.001$, $d = 1.26$), there is no evidence to suggest that age modulates face scanning for the age groups included in this study. However, additional correlational control analyses were conducted to ensure age did not significantly modulate any observed group effects.

All participants had normal or corrected-to-normal vision and hearing. The study lasted approximately 30 minutes and each participant received – in line with departmental regulations – £8 (London) or ¥1000 (Kyoto) for their time. The study was approved locally by the ethics committees of the Department of Psychological Sciences, Birkbeck, University of London, and the Department of Psychology, Kyoto University. Each participant provided written informed consent prior to the study.

4.3.2. Apparatus

Eye movements were recorded using SMI eye tracking glasses (SMI ETG; SensoMotoric Instruments, Germany) at a sampling rate of 60 Hz. An integrated scene camera fitted with a normal lens recorded the participant's field-of-view (60° horizontally and 46° vertically), with two integrated eye cameras and infra-red LEDs used for binocular gaze tracking. Scene recordings were captured at 24 frames per second and at 1280 x 960 resolution. The SMI ETG were connected to a portable recording unit that enabled the experimenter to control the recording sessions. A small object displaying concentric circles was used for calibration purposes (see Figure 4.1 on page 127). At 1 metre viewing distance, the diameter of the calibration object measured approximately 24.3°, and the centre of the object (i.e., the point that participants were asked to fixate) measured approximately 1.7°.

Figure 4.1. Small calibration object. Participants were asked to fixate the image centre.

4.3.3. Procedure

The overall procedure was similar to the previous study in Chapter 3, with only a few modifications due to the methodological changes that were outlined in the introduction. Participants were welcomed in the preparation room where the experimenter explained the study, collected written informed consent, and asked participants to fill out a demographic questionnaire. As in Chapter 3, participants were informed that the content of their speech would not be used for analysis to facilitate a naturalistic interaction. The participants were also informed that the current study aimed to examine cultural differences in face perception, and the notion of face scanning strategies was not mentioned. The participants were then guided to the testing room where the experimenter mounted the SMI ETG. The participant was seated at a table opposite the local research assistant at approximately 1 metre distance. British/Irish participants interacted with a British research assistant (British-White ethnicity) in English, and Japanese individuals interacted with a Japanese research assistant (Japanese ethnicity) in Japanese (Figure 4.2 on page 128). Both research assistants were male and in their mid-20s.

Figure 4.2. Participants' view of the local research assistant in the UK (left) and in Japan (right).

To complete a three-point calibration procedure, the research assistant held the calibration object in one hand (Figure 4.3) and asked the participant to fixate the image centre. During the piloting phase, participants often involuntarily shifted their gaze away from the calibration target and toward the eyes of the research assistant. The research assistants were therefore instructed to avoid eye contact by gazing downward during the calibration process (see Figure 4.3), and the experimenter, who controlled the recording unit, indicated to the research assistants when to move to the next calibration point. Prior to each experimental task, re-calibration was performed to avoid any drift. Following successful calibration, each participant was also asked to fixate the calibration target an additional five times just before and just after each experimental task. These gaze points were used as post-hoc checks to ensure that calibration was sufficiently accurate and precise.

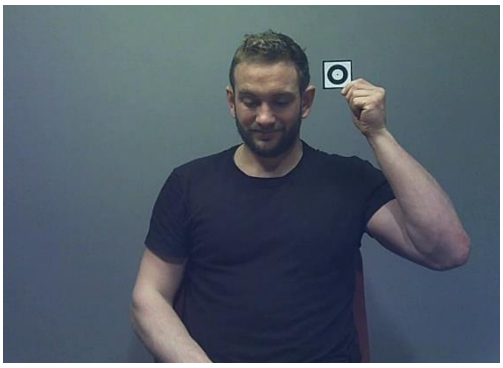


Figure 4.3. Participant's view during the three-point calibration procedure.

The research assistant then explained the forthcoming task in more detail (described below), after which the experimenter left to ensure that the dyadic interaction was not influenced by a third person in the room. The experimenter returned after each task for re-calibration.

The study tasks were identical to those described in Chapter 3. To briefly recap the tasks, participants were first asked to introduce themselves (*Introduction*) and were encouraged to speak for at least 30 seconds to obtain sufficient data for analysis. Participants were informed that they could mention their name, occupation, or hobbies, but that they were free to talk about anything as long as they were comfortable with sharing their personal details. The research assistants were also asked to introduce themselves, and the participant and research assistant were free to have a conversation afterwards. The second task consisted of two rounds of the guessing game *20 Questions* in which one person thought of an object while the other asked up to 20 questions to guess the object. Only '*Yes*', '*No*' or '*I don't know*' were permitted answers. An additional round of *20 Questions* was played if the object was guessed correctly before reaching 10 questions. As in Chapter 3, given that *20 Questions* has a clear structure, this game was included to facilitate a naturalistic interaction, but face scanning strategies were not analysed for this task. In the third task (*Story-telling*), the participant and research assistant each picked a coin from the table, looked at the year shown on the coin, and told the other person about a personal event or experience that happened in that year. As with the introductory task, participants were asked to talk for at least 30 seconds. After the third task, the experimenter stopped the recording. For the *Introduction* and *Story-telling* task, the research assistants kept their content of speech consistent across participants. The research assistants were also instructed to not interrupt the participant in order to facilitate later coding of speaking and listening periods (for more details see Section 4.3.4.2). Finally, participants were asked to complete the Autism Quotient (AQ) (Baron-Cohen et al., 2001; see Appendix A) and the self-report version of the Liebowitz Social Anxiety Scale (LSAS-SR; Liebowitz, 1987; see Appendix B). Both the AQ and LSAS were provided in the participants' native language and served to examine the relationship between face scanning and autistic traits/social anxiety traits in the general population.

4.3.4. Data pre-processing

4.3.4.1. *Data quality*

Spatial accuracy (or *offset*) is crucial for examining face scanning behaviour, particularly when contrasting two groups that may systematically differ in data quality. A vertical or horizontal gaze offset that is only present in one group but not the other could mask true group differences or, alternatively, falsely suggest cultural differences in face scanning. Given that ethnicity can affect data quality (Blignaut & Wium, 2014), post-hoc calibration points were presented twice for each experimental task in the current study: after initial calibration (prior to the start of the task) to ensure that the calibration procedure itself was accurate, and after the experimental task (prior to the next calibration procedure) to account for drift or slippage of glasses during the task. Associated gaze data was examined offline in *BeGaze* (Version 3.7; SensoMotoric Instruments, Germany) by overlaying gaze points onto the scene recordings and checking whether the crosshair fell onto the post-hoc calibration targets. If the data showed no offset prior to the start of the experimental task, no recalibration was performed. Drift or slippage of glasses was examined using the post-hoc calibration data collected at the end of the experimental task. When a consistent offset was detected (i.e., a *linear* offset was present), recalibration was conducted in *BeGaze* by automatically shifting all gaze coordinates to their accurate position. The accuracy of the recalibration procedure was then confirmed using the post-hoc calibration data at the end of the task.

4.3.4.2. *Coding of speaking and listening periods*

The same criteria as in Chapter 3 were applied (see Section 3.3.4.2).

4.3.4.3. *Regions-of-interest coding*

Face regions were coded semi-automatically using the detection and tracker tool presented in Chapter 2, and face regions were divided further into upper and lower face.

4.4. Results

4.4.1. Regions-of-interest analysis

A 2 (Group: British/Irish, Japanese) x 2 (Speech: speaking, listening) x 2 (Task: introduction, story-telling) mixed ANOVA was conducted on both dwell time and fixation time measures, separately for proportional face looking and proportional upper face scanning. The assumption of normality was violated and could not be corrected with data transformation since speaking and listening periods resulted in data that was skewed in opposite directions. Given that no equivalent non-parametric version exists, the 2 x 2 x 2 mixed ANOVA was nevertheless conducted, with any significant effects followed up or confirmed using appropriate non-parametric tests.

Dwell time on the face was calculated as a proportion of valid overall dwell time (with a cut-off at 30 seconds per task). Fixations were first extracted in *BeGaze* (Version 3.7; SensoMotoric Instruments, Germany), which requires fixations to have a minimum fixation duration of 50 milliseconds and a maximum dispersion of 0.5° visual angle. Fixation time on the face was then computed relative to the valid total fixation duration. Periods of data loss (due to, e.g., blinks or temporary track loss) were excluded from the valid time to consider only those periods during which it was possible to confidently state whether or not participants engaged in face looking. No significant group differences were found for data loss (Japanese: $M = 9.60\%$, $SD = 7.25\%$; British/Irish: $M = 6.62\%$; $SD = 5.15\%$; $t(54) = 1.78$, $p = 0.080$), suggesting that both cultural groups contributed a similar amount of data for the analysis presented below.

4.4.1.1. *Face looking*

4.4.1.1.1. *Dwell time analysis*

Both cultural groups spent a similar amount of time scanning the face of the research assistant when listening, and face looking decreased for speaking periods (see Table 4.1). Dwell time on the face was also overall greater during the introduction task compared to the story-telling task (Table 4.1). These patterns were reflected in a significant main effect of Speech ($F(1,54) = 209.98$, $p < 0.001$, $\eta_p^2 = 0.795$; confirmed using the Wilcoxon Signed-Rank Test: $Z = -6.51$, $p < 0.001$, $r = 0.615$), and a significant main effect of Task ($F(1,54) = 21.76$, $p < 0.001$, $\eta_p^2 = 0.287$; confirmed using Wilcoxon Signed-Rank Test: $Z = -2.84$, $p = 0.005$, $r = -0.268$). No other significant main effects or interactions were found (Group: $F(1,54) = 0.62$, $p = 0.435$; Speech x Group: $F(1,54) = 0.51$, $p = 0.478$; Speech x Task: $F(1,54) = 2.59$, $p = 0.113$; Task x Group: $F(1,54) = 0.07$, $p = 0.797$; Speech x Task x Group: $F(1,54) = 0.05$, $p = 0.830$). In sum, both cultural groups spent a similar proportion of the valid recording time scanning the face, with increased face looking during listening compared to speaking periods, and during the introduction task compared to the story-telling game (see Figure 4.4 on page 133).

Table 4.1. *Medians and interquartile ranges for face dwell time (in %).*

		Japanese	British/Irish
		Mdn (IQR)	*Mdn (IQR)*
Introduction	Speaking	37.47 (30.04)	49.61 (31.03)
	Listening	75.97 (28.67)	84.88 (23.00)
Story-telling	Speaking	27.47 (22.56)	32.39 (38.33)
	Listening	69.63 (21.34)	72.18 (17.33)

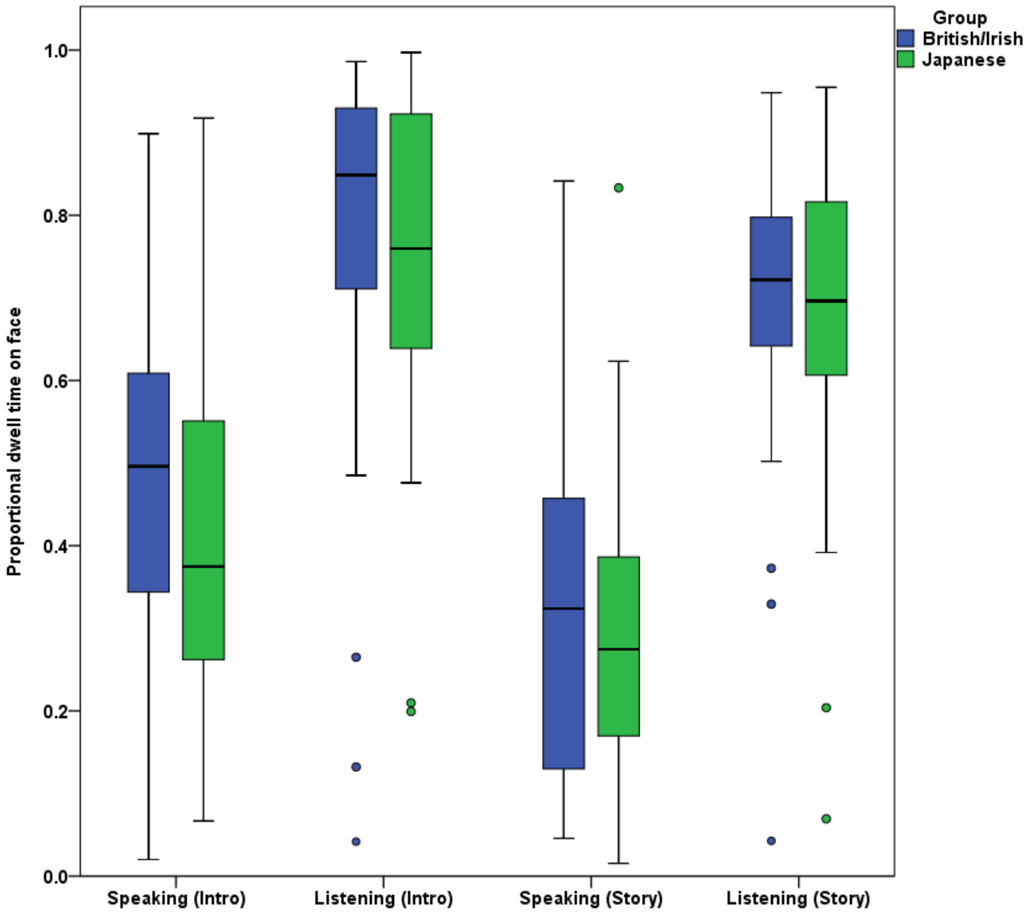


Figure 4.4. Proportional face dwell time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points).

4.4.1.1.2. *Fixation analysis*

Face fixation time was computed for each participant as a proportion of total fixation duration during the recording time (with a cut-off at 30 seconds per task). For listening periods, both groups spent a similar amount of time fixating the face of the research assistant, and face fixation time decreased during speaking periods (see Table 4.2 on page 134). Total face fixation duration was also greater for the introduction task relative to the story-telling task (Table 4.2 on page 134). These findings were manifested in a significant main effect of Speech ($F(1,54) = 182.95$, $p < 0.001$, $\eta_p^2 = 0.772$; confirmed using the Wilcoxon Signed-Rank Test: $Z = 6.42$, $p < 0.001$, $r = -0.607$) and Task

($F(1,54) = 25.56$, $p < 0.001$, $\eta_p^2 = 0.321$; confirmed using Wilcoxon Signed-Rank Test: $Z = -4.10$, $p < 0.001$, $r = 0.388$). A significant Speech x Group interaction was also found ($F(1,54) = 4.83$, $p = 0.032$, $\eta_p^2 = 0.082$); however, post-hoc comparisons using the non-parametric Mann Whitney U test showed that the speaking and listening condition did not differ between the two groups (speaking: $U = 287$, $p = 0.087$; listening: $U = 364$, $p = 0.652$). In other words, the *difference* in proportional face fixation duration between the speaking and listening condition was significantly greater in the Japanese than the British/Irish group, but no cultural differences existed when directly comparing the groups in each speech condition. There were no other significant main effects or interactions (Group: $F(1,54) = 0.09$, $p = 0.770$; Task x Group: $F(1,54) = 0.43$, $p = 0.513$; Speech x Task: $F(1,54) = 3.99$, $p = 0.051$; Speech x Task x Group: $F(1,54) = 1.04$, $p = 0.312$). In sum, total face fixation duration was higher during the introduction task compared to the story-telling game, and for periods when participants were listening to the research assistant (see Figure 4.5 on page 135), mirroring the findings for the dwell time analysis (see Section 4.4.1.1.1).

Table 4.2. *Medians and interquartile ranges for face fixation time (in %).*

		Japanese *Mdn (IQR)*	British/Irish *Mdn (IQR)*
Introduction	Speaking	43.95 (32.76)	63.89 (30.32)
	Listening	84.14 (18.64)	91.02 (22.28)
Story-telling	Speaking	31.05 (34.79)	39.77 (42.79)
	Listening	81.46 (19.54)	80.00 (24.79)

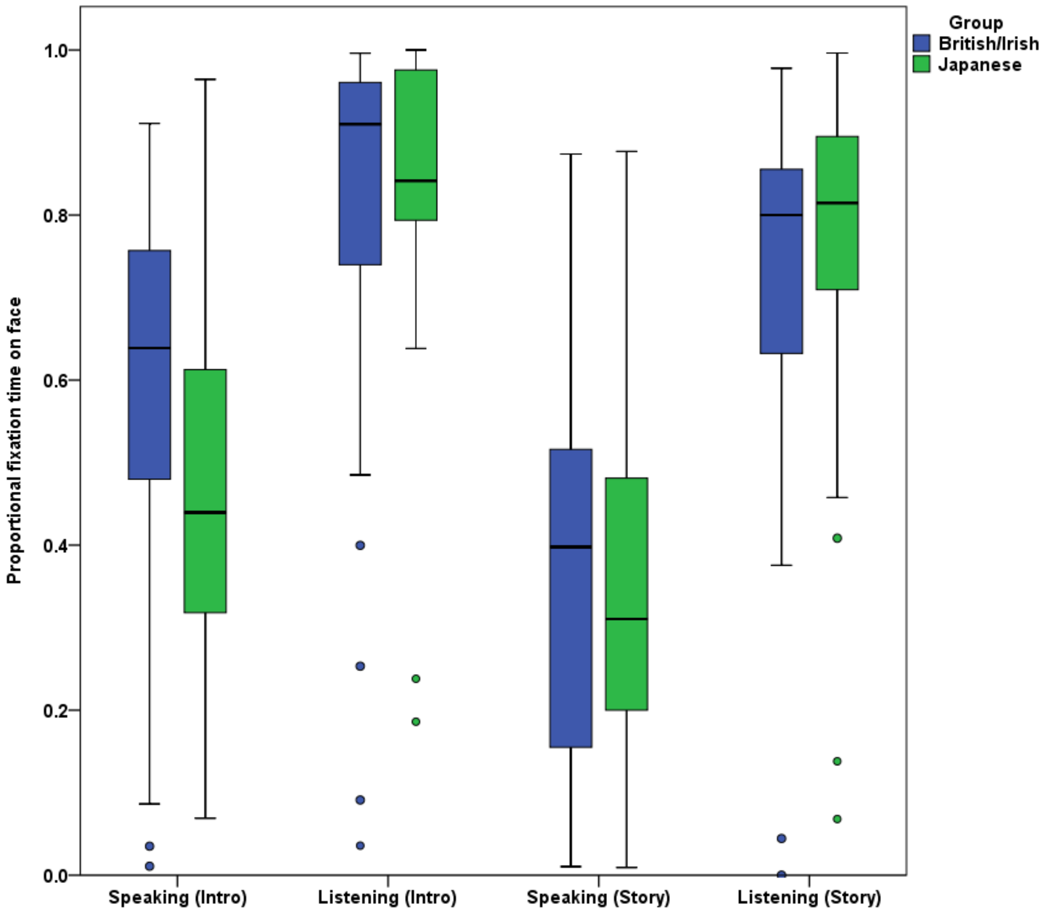


Figure 4.5. Proportional face fixation time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points).

4.4.1.2. *Upper face looking*

4.4.1.2.1. *Dwell time analysis*

Upper face looking time was calculated as a proportion of overall dwell time on the face. Unlike the findings for face looking, proportional dwell times for scanning the upper face region remained consistent across tasks (see Table 4.3 on page 136). In line with predictions, Japanese participants also spent a higher proportion of face looking time scanning the upper face region than British/Irish participants (Table 4.3 on page 136). This pattern was reflected in a significant main effect of Group (F(1,54) = 8.06,

p = 0.006, η_p^2 = 0.130; confirmed using the Mann Whitney U Test: U = 222, p = 0.005, r = -0.371). No other main effects or interactions could be found (Speech: $F(1,54)$ = 0.31, p = 0.579; Task: $F(1,54)$ = 3.09, p = 0.085; Speech x Group: $F(1,54)$ = 1.326, p = 0.255; Task x Group: $F(1,54)$ = 0.12, p = 0.727; Speech x Task: $F(1,54)$ = 2.71, p = 0.106; Speech x Task x Group: $F(1,54)$ = 0.60, p = 0.442). In sum, Japanese participants, relative to British/Irish participants, spent a significantly greater proportion of face looking time scanning the upper face region of the research assistant (see Figure 4.6 on page 137).

Table 4.3. *Medians and interquartile ranges for upper face dwell time (in %).*

		Japanese *Mdn (IQR)*	British/Irish *Mdn (IQR)*
Introduction	Speaking	70.23 (25.97)	48.96 (50.83)
	Listening	80.39 (21.71)	55.87 (62.53)
Story-telling	Speaking	66.81 (24.53)	56.27 (35.89)
	Listening	77.54 (29.10)	49.05 (52.81)

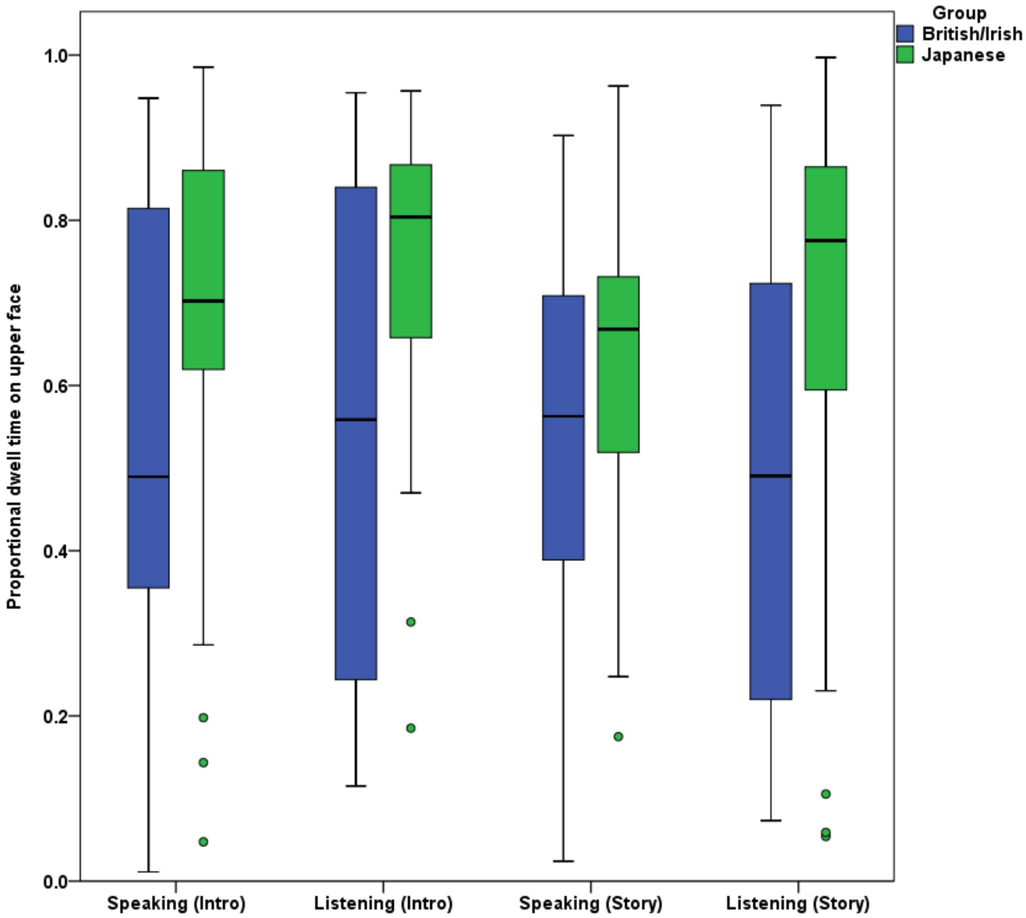


Figure 4.6. Proportional upper face dwell time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points).

4.4.1.2.2. *Fixation analysis*

Upper face fixation time was computed for each participant as a proportion of total fixation time on the face. Japanese participants spent a greater proportion of overall face fixation time scanning the upper face region than British/Irish participants (see Table 4.4 and Figure 4.7, both on page 138). As with the findings from the dwell time analysis, a significant main effect of Group was revealed ($F(1,54) = 10.47$, $p = 0.002$, $\eta_p^2 = 0.162$; confirmed using Mann Whitney U test: $U = 248$, $p = 0.019$, $r = -0.314$), but no other main effects or interactions were found (Speech: $F(1,54) = 0.10$, $p = 0.759$; Task: $F(1,54) = 3.52$, $p = 0.066$; Speech x Group: $F(1,54) = 1.83$, $p = 0.182$; Task x Group:

$F(1,54) = 0.01$, $p = 0.946$; Speech x Task: $F(1,54) = 0.89$, $p = 0.349$; Speech x Task x Group: $F(1,54) = 0.04$, $p = 0.853$).

Table 4.4. *Medians and interquartile ranges for upper face fixation time (in %).*

		Japanese *Mdn (IQR)*	British/Irish *Mdn (IQR)*
Introduction	Speaking	79.40 (31.72)	58.71 (61.12)
	Listening	84.66 (25.53)	53.10 (67.12)
Story-telling	Speaking	69.33 (42.66)	57.70 (46.03)
	Listening	78.25 (32.86)	49.26 (56.62)

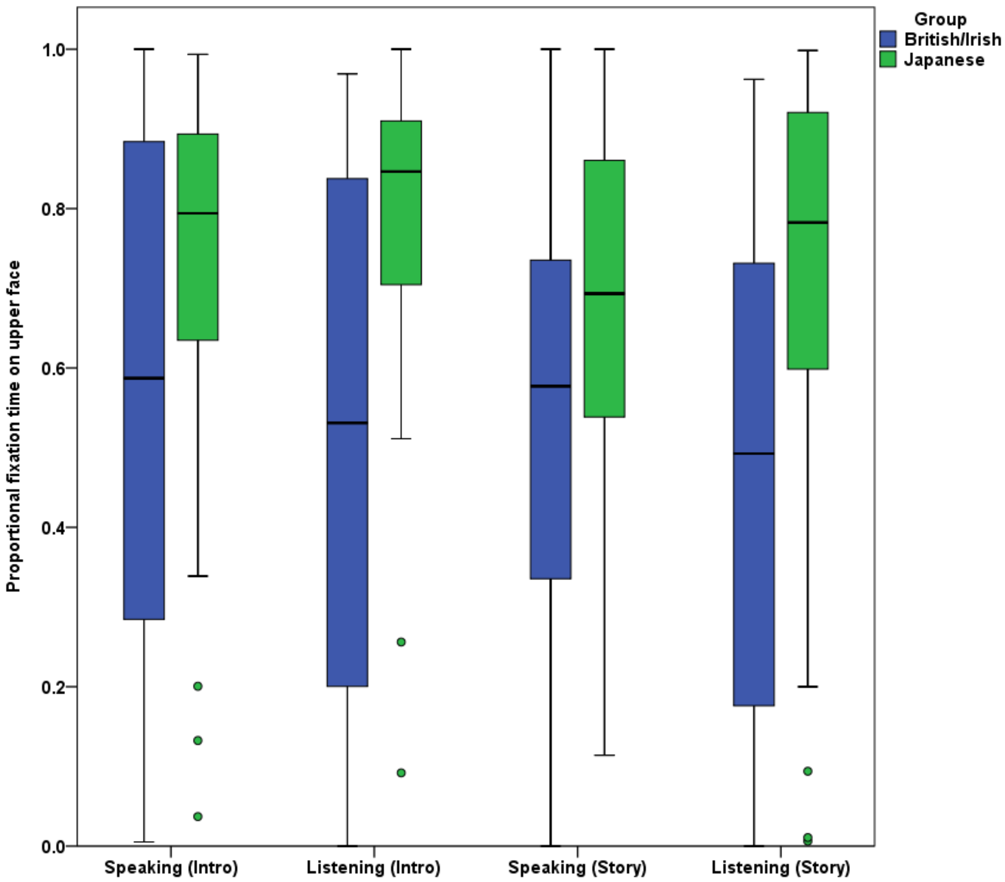


Figure 4.7. Proportional upper face fixation time for British/Irish and Japanese participants during listening and speaking periods. Whisker ends represent minimum and maximum values, with the exception of outliers (> 1.5 x IQR; represented by points).

4.4.2. Monte Carlo permutation test

To examine face scanning behaviour in a more spatially sensitive manner, data-driven Monte Carlo permutation tests (see Chapter 2) were conducted and implemented in MATLAB using the *CoSMoMVPA* toolbox (Oosterhof et al., 2016) and *FieldTrip* toolbox (Oostenveld et al., 2011). Analysis was performed separately for speaking and listening periods using either the dwell or fixation measures. For the present analysis, parameters were set to the following:

- an uncorrected p-value threshold of 0.01, and
- 10,000 iterations.

Initial analysis conducted separately for the introduction and story-telling task revealed no clusters that survived multiple comparison corrections. To ensure null findings were not simply due to a lack of statistical power resulting from insufficiently available eye movement data, measures were collapsed across tasks. Although the regions-of-interest analysis revealed significant differences between tasks (see Sections 4.4.1.2.1 and 4.4.1.2.2), this only applied to face looking overall and not upper face scanning. In other words, the experimental tasks modulated the amount of orienting behaviour toward the face but not the spatial allocation within the face, making it possible to collapse data across tasks.

Figure 4.8 and Figure 4.9, both on page 140, illustrate descriptive heat maps showing group differences in gaze density for dwell and fixation measures, respectively, with a tendency of more looking at the eye region in Japanese adults and mouth scanning for British/Irish individuals.

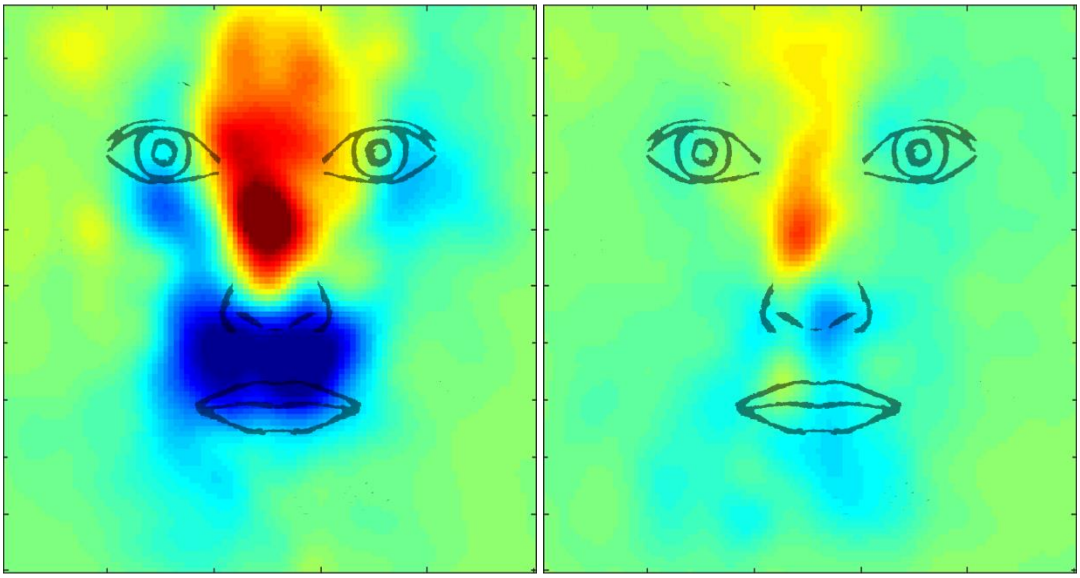


Figure 4.8. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for dwell measures, with red and blue colours depicting regions that East Asians and Western Caucasians scanned more, respectively.

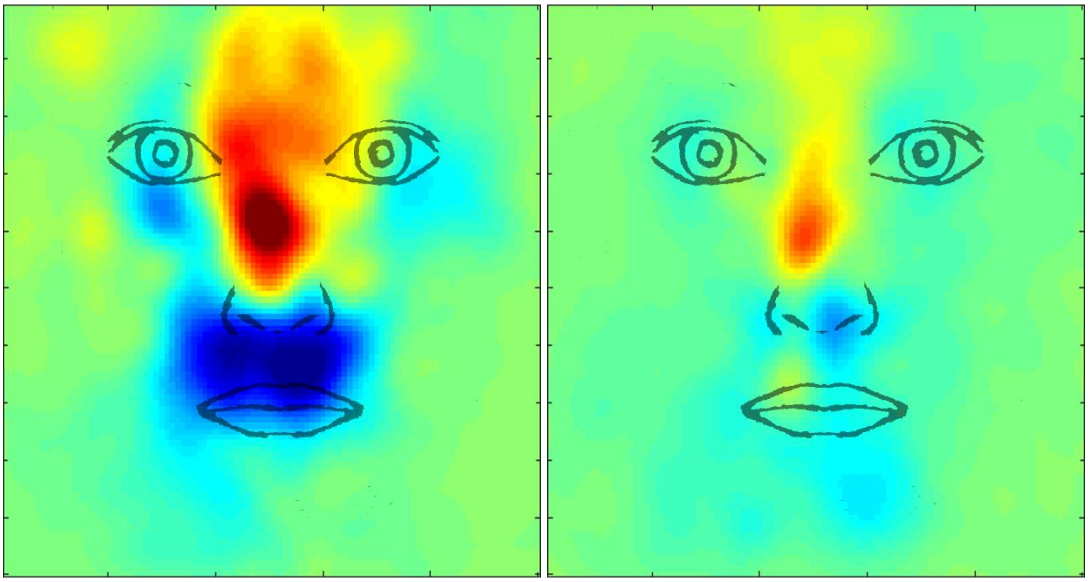


Figure 4.9. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for fixation measures, with red and blue colours depicting regions that East Asians and Western Caucasians scanned more, respectively.

The statistical analysis revealed significant clusters for listening but not speaking periods (see Figure 4.10, and Figure 4.11 on page 142). When listening to the research assistant, Japanese participants scanned the region between the eyes more frequently than the British/Irish group, mirroring the regions-of-interest findings suggesting significantly greater upper face scanning for Japanese participants (see Sections 4.4.1.2.1 and 4.4.1.2.2). British/Irish individuals scanned the mouth region more than the Japanese group. Dwell and fixation measures produced similar clusters (see Figure 4.10, and Figure 4.11 on page 142), though for the speaking condition no significant gaze clusters could be identified after multiple comparison correction.

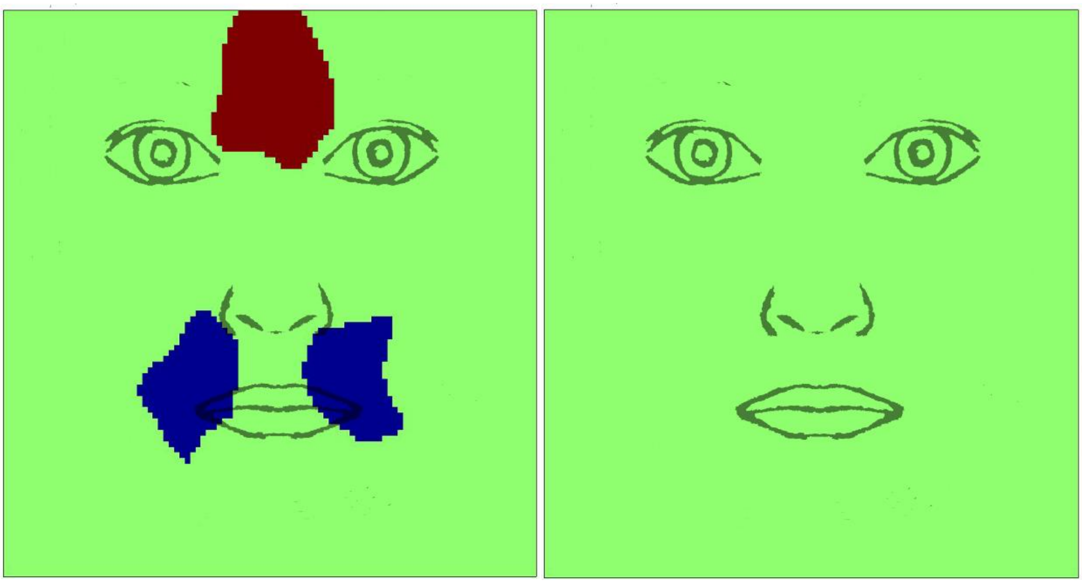


Figure 4.10. Gaze clusters for listening (left) and speaking periods (right) using dwell data, illustrating regions that were scanned significantly more by British/Irish (blue) or Japanese participants (red).

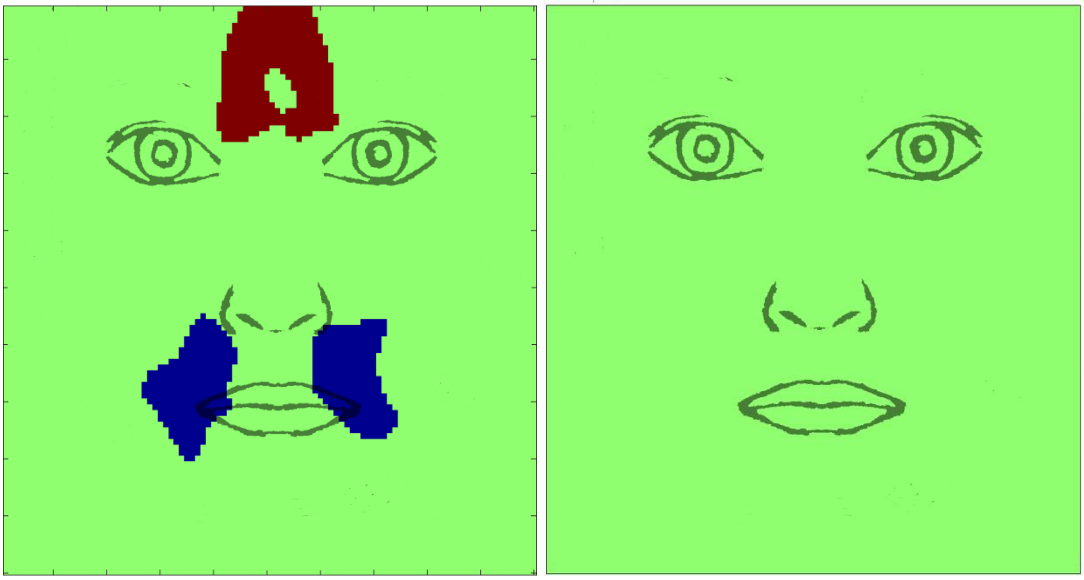


Figure 4.11. Gaze clusters for listening (left) and speaking periods (right) using fixation data, illustrating regions that were scanned significantly more by British/Irish (blue) or Japanese participants (red).

4.4.3. Autism Quotient (AQ)

British/Irish participants obtained a significantly lower AQ score (M = 14.00, SD = 6.28, ranging from 5 to 30) than Japanese individuals (M = 21.59, SD = 8.65, ranging from 10 to 39; $t(54) = 3.78$, $p < 0.001$, $d = 1.00$). A correlational analysis was conducted to examine any potential relationship between AQ scores and dwell/fixation time on the face or upper face. As in Chapter 3, analysis was performed separately for each cultural group to avoid the reversal paradox (see, e.g., Kievit et al., 2013). For the face looking measure, the correlational analysis was conducted separately for each task given that significant differences in face dwell/fixation time were found between the introduction and story-telling task (see Sections 4.4.1.1.1 and 4.4.1.1.2). In contrast, given that task effects were not present for upper face looking, the correlational analysis was conducted by collapsing the data across the two tasks. Results showed no significant correlations between AQ scores and dwell time on the (upper) face (all $p > 0.05$; see Table 4.5 on page 143). In addition, correlating AQ scores with (upper) face fixation time (as opposed to dwell time) also resulted in no significant correlations (all $p > 0.05$; see Table 10.1 in Appendix C).

Table 4.5. *Pearson's correlation coefficients for the relationship between AQ scores and (upper) face dwell time during speaking and listening periods.*

		Japanese	British/Irish
Speaking	AQ – Face looking (Intro)	-0.196	-0.104
	AQ – Face looking (Story)	-0.266	-0.250
	AQ – Upper face looking	-0.049	-0.276
Listening	AQ – Face looking (Intro)	0.076	-0.218
	AQ – Face looking (Story)	-0.264	-0.033
	AQ – Upper face looking	0.067	-0.207

4.4.4. Liebowitz Social Anxiety Scale (LSAS)

British/Irish participants had a significantly lower LSAS score ($M = 36.45$, $SD = 16.32$, ranging from 10 to 75) than Japanese individuals ($M = 48.19$, $SD = 21.07$; $t(54) = 2.34$, $p = 0.023$, $d = 0.62$, ranging from 10 to 94). A correlational analysis conducted separately for each cultural group and task revealed a significant negative correlation between LSAS scores and face dwell time during speaking periods in the introduction task (Japanese: $r = -0.415$, $p = 0.032$; British/Irish: $r = -0.423$, $p = 0.022$), suggesting that greater social anxiety traits were associated with decreased face looking. This was replicated when correlating LSAS scores with (upper) face fixation time (Japanese: $r = -0.442$, $p = 0.021$; British/Irish: $r = -0.360$, $p = 0.055$; see Table 10.2 in Appendix C). No other significant correlations were found for dwell time measures ($p > 0.05$; Table 4.6 on page 144). For the British/Irish group, a significant negative correlation between LSAS scores and face fixation time during speaking periods in the story-telling task was found ($r = -0.394$, $p = 0.035$; see Table 10.2 in Appendix C).

Table 4.6. *Pearson's correlation coefficients for the relationship between LSAS scores and (upper) face dwell time during speaking and listening periods.*

		Japanese	British/Irish
Speaking	LSAS – Face looking (Intro)	-0.415*	-0.423*
	LSAS – Face looking (Story)	-0.102	-0.286
	LSAS – Upper face looking	-0.147	0.049
Listening	LSAS – Face looking (Intro)	0.015	-0.279
	LSAS – Face looking (Story)	-0.224	-0.056
	LSAS – Upper face looking	-0.127	0.024

*$p < 0.05$

4.4.5. Control analysis for age

Given the significant difference in age between British/Irish and Japanese participants (see Section 4.3.1), an additional control analysis was conducted to examine any potential relationships between participant age and (upper) face dwell time. Correlational analyses conducted separately for each cultural group did not reveal any significant relationships (all $p > 0.05$; Table 4.7 on page 145). Significant correlations were also not found when using (upper) face fixation time (see Table 10.3 in Appendix C). This suggests that age unlikely modulated the present findings.

Table 4.7. *Pearson's correlation coefficients for the relationship between age and (upper) face dwell time during speaking and listening periods.*

		Japanese	British/Irish
Speaking	Age – Face looking (Intro)	0.327	0.138
	Age – Face looking (Story)	0.226	0.156
	Age – Upper face looking	-0.155	-0.254
Listening	Age – Face looking (Intro)	0.214	0.017
	Age – Face looking (Story)	0.150	-0.136
	Age – Upper face looking	0.154	0.053

4.5. Discussion

To date, cross-cultural comparisons on face scanning strategies have been restricted to screen-based paradigms which largely employed static images as face stimuli. However, factors such as low-level motion (from, e.g., a moving mouth), the social presence of another person, or visual distractors – all characteristics common to 'real-world' conditions – can affect face orienting behaviour (Foulsham et al., 2011; Laidlaw et al., 2011). It is therefore possible that individuals exhibit different scanning behaviour within a naturalistic social context. The present study aimed to examine cultural differences in face scanning strategies during dyadic face-to-face interactions, addressing also the methodological limitations of the experiment presented in Chapter 3 in order to clarify study interpretations. In the following, findings will be summarised and discussed with respect to the results from the previous study (Chapter 3) as well as the wider literature.

The regions-of-interest analysis was able to measure general face orienting behaviour, while the data-driven findings provided detailed insight into the spatial allocation of participants' eye movements within the face. Dwell time and fixation time analyses produced the same pattern of results, indicating that differences in dependent variables did not affect findings. In line with the results presented in Chapter 3, the current findings demonstrated greater overall face looking during periods of listening compared to periods of speaking for both cultural groups. This replicates findings from previous studies (e.g., Freeth, Foulsham, et al., 2013) and suggests that the speech effect is robust across different naturalistic face-to-face paradigms, which is supported by the large effect size observed in this study. Decreased face orienting during speaking periods is consistent with the finding that individuals tend to avert their gaze during more cognitively demanding periods to reduce cognitive load (Doherty-Sneddon & Phelps, 2005). In addition, increased face looking at the research assistant during periods of listening could have helped participants to decode speech (Vatikiotis-Bateson et al., 1998), as well as functioned as a social signal to indicate to the conversational partner that one is still listening (Risko et al., 2016). With respect to the experimental tasks, both cultural groups engaged in more face looking during the introduction period than the story-telling game, replicating results from the previous study (Chapter 3). As discussed in Chapter 3, this task effect could reflect greater social signalling when dyads were less familiar with each other, or indicate a need to avert gaze in order to reduce cognitive load given the more demanding nature of the story-telling game (Glenberg et al., 1998), which required participants to recall and describe a past event or experience. Consistent with the present predictions, no cultural effects could be found, suggesting that the amount of overall face looking during dyadic interaction was not modulated by culture. This is in contrast with the findings from Chapter 3, which revealed significantly more face looking in East Asians compared to Western Caucasians. In Chapter 3, this was discussed with respect to an increased need for Western Caucasians to avert gaze in order to reduce cognitive load, and a greater tendency for East Asians to exhibit social signalling. However, the study in Chapter 3 also suffered some methodological limitations as outlined in the introduction of this chapter, which could have affected face orienting; for instance, the East Asian participants in Chapter 3 tended to engage in more face orienting than the Japanese residents in the current study. As outlined in Section 4.2.3, immigrant populations (Chapter 3) may engage in qualitatively different social interactions compared to local residents (current study) due to their shared experience within a new cultural environment, thereby possibly leading to differences in social signalling and thus face orienting. The current study with its methodological improvements cannot support the findings from Chapter 3, questioning the notion of an increased need to reduce cognitive load in Western Caucasians or social signalling in East Asians *generally* by means of greater face orienting.

The regions-of-interest analysis also revealed that Japanese participants directed a significantly greater proportion of face looking time to the upper region compared to British/Irish individuals. This is in line with the current predictions, mirroring findings obtained from screen-based studies employing emotionally expressive face stimuli (e.g., Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). It is also possible that Japanese participants engaged in greater social signalling compared to the British/Irish group. Although greater upper face scanning in Japanese participants contradicts the gaze avoidance theory, which postulates decreased eye contact in East Asian populations (Argyle et al., 1986), the results from the permutation analysis pointed to the possibility that Japanese participants directed their gaze *between* the eyes. However, given that eye movements in this study were not recorded for both the participant and the research assistant, only limited interpretations can be made. Future studies will be required to examine the gaze avoidance theory in more detail.

The observed cultural differences could partly be explained by group differences in scanning strategies for emotionally expressive faces. Whereas East Asians fixate the eye region when decoding emotional expressions, Western Caucasians distribute their gaze across the face (Jack et al., 2009). Consistent with such cultural differences in eye movement behaviours, Jack et al. (2012) showed using computational modelling that East Asians represent the intensity of emotional expressions with movements of the eyes, whereas Western Caucasians use other face regions. Given that participants in the current study likely exhibited emotional facial expressions during their dyadic interactions, it may have been visually more informative for Japanese participants to scan the upper face region.

Conversely, it is possible that attending to the mouth region could have been visually more informative for British/Irish participants. While both Japanese and English native speakers benefit from attending to the mouth when decoding speech (Vatikiotis-Bateson et al., 1998), phonological differences between the English and Japanese languages could contribute to cultural differences in audio-visual speech perception and in turn cultural differences in face scanning. Lip-read information has been shown to be less ambiguous and therefore more informative in the English than Japanese language (Sekiyama & Tohkura, 1993); for instance, English consonants can be categorised into a higher number of consonant groups by lip-reading than Japanese consonants (Sekiyama, Tohkura, & Umeda, 1996). This is also consistent with the findings that visual cues – in addition to auditory ones – do not benefit Japanese second language learners in consonant perception (Hazan et al., 2006), and that Japanese individuals exhibit a significantly reduced McGurk effect (Sekiyama & Tohkura, 1991). It may therefore have been less informative for Japanese participants to look at the mouth than for British/Irish individuals, giving rise to the observed cultural differences in face scanning.

Findings from the permutation analysis also showed that only the listening but not the speaking condition resulted in cultural differences. Given that both cultural groups showed significantly less face orienting behaviour for the speaking than the listening condition, it is possible that the null finding resulted from insufficient eye movement data. This is supported by the descriptive heat maps (see Figure 4.8 and Figure 4.9, both on page 140), which were characterised by similar density patterns, but the speaking condition showed a much weaker effect.

Despite the similarities with screen-based findings on face scanning using emotionally expressive faces, it cannot necessarily be concluded that underlying mechanisms for scanning strategies extended to naturalistic face-to-face interactions.

Unlike screen-based paradigms, eye movements in naturalistic social settings not only serve to encode the stimulus or environment, but can also signal one's mental state to another person (Gobel et al., 2015; Risko et al., 2016). Further research that manipulates and thereby disentangles the encoding and signalling functions will be required to draw conclusions about the underlying mechanisms that drive the observed cultural differences in naturalistic face scanning.

The current study also included AQ and LSAS measures to investigate whether face scanning was associated with autistic or social anxiety traits. Cultural differences were found for both the AQ and LSAS measures, with Japanese participants exhibiting significantly higher autistic and social anxiety traits than British/Irish individuals. This is consistent with previous research indicating that Japanese individuals obtained higher AQ scores (Kurita et al., 2005; Wakabayashi et al., 2006) and LSAS scores (Freeth, Bullock, et al., 2013; Sugawara et al., 2012). In the current study, no significant correlations could be found between AQ scores and (upper) face scanning for either cultural group. Although this contrasts findings from screen-based studies showing that higher autistic traits are associated with reduced face looking (e.g., Chen & Yoon, 2011; Freeth, Foulsham, et al., 2013), the present results are consistent with several live face-to-face interactions (e.g., Freeth, Foulsham, et al., 2013; Vabalas & Freeth, 2016). Vabalas and Freeth (2016), for instance, did not find a relationship between autistic traits and the amount of upper/lower face scanning during dyadic interactions in British participants. With respect to Japanese individuals, no study to date has examined their face scanning using a live dyadic interaction, and as such no evidence is available regarding the relationship with autistic traits. The findings from this study therefore suggest that there is no significant relationship between (upper) face scanning and autistic traits in Japanese adults. In sum, the present results indicate that autistic traits unlikely modulated the observed cultural differences.

For the LSAS scores, both cultural groups showed a significant association between greater social anxiety traits and decreased face orienting during speaking periods in the introduction. This finding did not, however, extend to periods of listening or to the story-telling game. It is possible that the role of social anxiety traits was more relevant during earlier stages of the social interaction (i.e., the introduction period) when participants were still relatively unfamiliar with their conversational partner. In addition, decreased face orienting may have been considered more socially appropriate during speaking periods (e.g., gaze aversion when listening to another person could signal disinterest), thereby allowing participants to look at the face at a degree with which they were comfortable.

The nature of face-to-face interaction paradigms and associated cross-cultural comparisons are inherently characterised by shortcomings that are problematic to solve but need to be acknowledged. First, it is possible that the observed cultural differences in face scanning were specific to the local research assistant. Unlike the Japanese research assistant, for instance, the beard of the British research assistant could have elicited the observed lower face looking in British/Irish participants. Given that the increased scanning of the lower face was restricted to the area immediately surrounding the mouth region, the lack of eye movements directed toward the remaining lower face region suggests that participants may have focused on the mouth region *per se*. It is possible though that British/Irish participants showed increased scanning of the mouth since the beard may have made it more difficult to decode the research assistant's lip movements (i.e., speech). However, this account should then not hold for periods of listening, but the regions-of-interest findings demonstrated that increased mouth scanning was not dependent on speaking and listening periods. With respect to the findings from the permutation test showing no cultural differences for speaking periods, it is possible, as mentioned above, that eye movement data was not sufficient to capture group differences. Matching research assistants across cultures and ethnicities is practically impossible, though future studies could include additional research assistants in each culture to reduce the possibility of obtaining face scanning patterns that are specific to the research assistant, or adopt virtual reality (VR) designs whereby avatars could serve as conversational partners. These proposals for future studies will be discussed in more detail in Chapter 7.

Another shortcoming relates to the size of the face. Although the current paradigm employed SMI ETG and thereby increased the size of the face relative to the previous study and ensured that faces were not distorted due to the use of wide-angle lenses, the size of the face in the scene recording was still significantly smaller than that of screen-based stimuli. A possibility to increase the size of the face for dyadic interactions includes using the two-way video set-up with eye tracking as introduced in Hessels, Cornelissen, Hooge, & Kemner (2017), which can live stream the face of a participant at high resolution to the screen of the social interaction partner (and vice versa). With the use of half-silvered mirrors, this two-way video set-up also ensures that eye contact can be achieved when both participants look each other in the eye region. Although this provides a possibility to examine face scanning with higher data quality, it is not clear whether the physical barrier between participants could affect the perception of the social presence of another person, and in turn influence eye movement behaviour. Alternatively, increasing the size of the face could be achieved by using the head-mounted Pupil Labs eye tracking system which features exchangeable scene camera lenses, including those with a smaller field-of-view that can focus onto the face region. This may increase data loss depending on the participant's natural head position – if the head is tilted too far, the face cannot be captured – but would achieve high resolution without participants having to interact at an uncomfortably close distance. Given the focus on *group* differences in the current study, however, increasing the size of the face does not invalidate the present findings, but would only serve to improve data quality.

The present study addressed the methodological limitations of the experiment presented in Chapter 3, with findings demonstrating cultural differences in face scanning during naturalistic social interactions. Semi-automatic coding methods were applied to conduct traditional regions-of-interest analysis and a data-driven approach was employed to allow for a more refined interpretation. Although the overall amount of face looking during dyadic interactions was not modulated by culture, factors including speech (listening versus speaking) and task demands influenced how often participants gazed at the face, suggesting that face orienting behaviour is highly context-dependent. In terms of scanning behaviour within the face, Japanese participants gazed at the region between the eyes more while British/Irish individuals looked at the mouth, replicating findings from screen-based studies using emotionally expressive faces. The next chapter will present a study that aimed to characterise the developmental trajectory of cultural differences in face scanning, examining also possible cognitive explanations.

Chapter 5

Cultural Differences in the Development of Face Scanning and the Role of Executive Function

5.1. Chapter Overview

As reviewed in Chapter 1, findings from screen-based eye tracking studies have demonstrated cultural differences in scanning strategies (e.g., Blais et al., 2008; Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). Chapters 3 and 4 further revealed that such cultural differences are not limited to screen-based paradigms but can also be observed within more naturalistic dyadic interactions. However, it remains unclear when and how such cultural differences emerge. To address this gap in the literature, the current study presented faces on screen and adopted a developmental framework to examine scanning strategies of two infant groups, with an average age of 10 and 16 months, as well as an adult group. To investigate possible explanations at the cognitive level, the role of executive function was also investigated. The following sections will first briefly summarise several cross-cultural developmental studies on face scanning and executive function and discuss how executive function may be involved in the emergence of cultural differences in face scanning before introducing the current study.

5.2. Introduction

5.2.1. Cross-cultural developmental studies on face scanning

Cultural differences in adults' face scanning have been found both in screen-based eye tracking paradigms (e.g., Blais et al., 2008; Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013) and during live dyadic interactions (Chapter 4), suggesting that cultural modulations on scanning behaviour occur across different face processing tasks and experimental settings. Although these studies provided insight into the manifestations of cultural differences in eye movements during face viewing, underlying mechanisms remain largely unknown (for a more detailed discussion see Chapter 1).

As outlined in Chapter 1, despite evidence pointing to postnatal developmental processes for the emergence of cultural differences in face scanning, previous studies largely investigated eye movement patterns of adults and little is known about the developmental trajectory. Only two cross-cultural studies on face scanning have been conducted with infants or young children. In particular, face scanning patterns were shown to be consistent with those of adults already at 7 months of age when viewing static, emotionally expressive face stimuli (Geangu et al., 2016), with British infants exhibiting greater looking at the mouth and Japanese infants fixating the eye region more. This pattern was also observed in young children aged between 1 and 7 years when employing dynamic, emotionally expressive stimuli (Senju, Vernetti, Kikuchi, Akechi, & Hasegawa, 2013). In addition, Kelly, Liu, et al. (2011) used static stimuli depicting faces with neutral expressions and found British and Chinese 7- to 12-year-olds to exhibit the triangular versus central fixation patterns, respectively, that were reported in adults (cf., Blais et al., 2008).

Several additional studies also examined infants' scanning behaviour for faces of their own versus an unfamiliar ethnicity; however, only one cultural group was tested in each of the studies. These studies revealed that 6- to 10-month-old Western Caucasian infants increasingly looked longer at the eyes and less at the mouth of faces from their own ethnicity (Wheeler et al., 2011; Xiao et al., 2013), and Chinese 4- to 9-month-old infants showed fewer fixations to internal facial features (and especially the nose) of unfamiliar Western Caucasian face stimuli (Liu et al., 2011). Unlike the studies by Geangu et al. (2016) and Senju, Vernetti, Kikuchi, Akechi and Hasegawa (2013), these infant studies suggest that the ethnicity of face stimuli also modulates viewing behaviour. However, direct group comparisons will be required to draw conclusions about cultural differences in infants' face scanning. Altogether, the findings from the developmental studies suggest that cultural differences in face scanning likely emerge by 10 months of age, with face ethnicity possibly influencing viewing behaviour.

5.2.2. The role of executive function

Given that cultural differences likely emerge during the second half of the first year of life, it is critical to identify potential mechanisms within this period of development to understand how infants increasingly adapt to their cultural environment. Given that no cross-cultural infant study has previously investigated possible explanatory factors in relation to face scanning, it is difficult to formulate novel predictions. However, three sets of findings from the developmental literature can provide initial insights into some explanatory factors that in turn can cast light on possible mechanisms. First, as outlined above, cultural differences have been found in face scanning, highlighting a role of cultural influences in social development. Secondly, cultural modulations have also been found in the development of non-social, cognitive domains such as executive function, with East Asian children typically showing more advanced development than Western Caucasian children (e.g., Imada, Carlson, & Itakura, 2013; Lan, Legare, Ponitz, Li, & Morrison, 2011; Oh & Lewis, 2008; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). Thirdly, current evidence points to a functional dependency between social development and executive function (see Moriguchi, 2014). For instance, Sabbagh et al. (2006) found advanced executive function in Chinese 3- to 5-year-olds compared to their US counterparts, with individual differences in executive functioning further predicting performance in the theory-of-mind task. Given this suggested relationship between culture, executive function, and social development (summarised in Figure 5.1 on page 157), it is important to understand the interactive nature between these components.

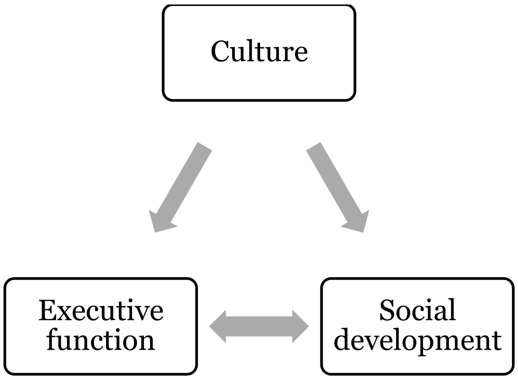


Figure 5.1. Hypothesised relationship between culture, executive function, and social development based on previous empirical findings.

Based on this relationship (Figure 5.1), it is possible that increasingly culture-specific face perception processes involve the development of executive function. For instance, the locations of visually informative regions for caregivers' facial expressions may differ between cultural groups (Fogel et al., 1988). More advanced executive function could allow for greater top-down control in visual attention, thereby helping infants to visually disengage from irrelevant features (e.g., irrelevant salient features) and attend to informative regions, which in turn could give rise to culture-specific scanning patterns. Given that cultural differences in face scanning begin to emerge toward the end of the first year of life, the idea that executive functions may be involved is consistent with findings pointing to the rapid development in key areas including controlled visual attention, working memory, and inhibitory control during the second half of the first year of life (e.g., Diamond, 1985; Diamond, 2002; Rose et al., 2012). Cultural differences in face scanning and increasingly advanced executive functions thus both emerge toward the end of the first year of life. To investigate this further, the current study aimed to map the developmental trajectory of face scanning in British and Japanese infants and adults, examining also whether cultural differences in viewing behaviour involved the development of executive function. The following section will introduce the current study in more detail and will outline the predictions.

5.2.3. The current study

As outlined above, it is necessary to map the developmental trajectory in each cultural group to directly compare face scanning patterns. In the current study, eye movements of 10- and 15- to 17-month-old British and Japanese infants as well as adults were recorded while participants viewed faces on screen. Given that early cultural differences possibly emerge by 10 months (cf., Geangu et al., 2016; Liu et al., 2011; Wheeler et al., 2011; Xiao et al., 2013), it was expected that early group differences should be present at this age. However, evidence is currently limited, and it is possible that cultural differences emerge later in development. An older infant age group consisting of 15- to 17-month-olds was therefore also included. From a developmental perspective, cultural differences should become increasingly evident in 15- to 17-month-olds and even more so in adults as individuals become more adapted to their cultural environment with age.

Stimuli included both White-British and Japanese faces and were presented in three different conditions: (1) *static* with neutral expression, as commonly employed in cross-cultural face scanning studies with adults (cf., Blais et al., 2008); (2) *dynamic-neutral*, showing a dynamic face with neutral expression as in previous cross-cultural infant studies (cf., Liu et al., 2011; Wheeler et al., 2011); (3) *dynamic-social*, a dynamic condition with emotionally expressive faces (cf., Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). The dynamic-social condition was included to take into account findings showing different scanning patterns for emotionally expressive versus neutral faces, and also to consider that infants typically encounter emotionally expressive faces in everyday face-to-face interactions. It was expected that British participants would exhibit greater triangular scanning (eyes and mouth) in the static and dynamic-neutral conditions, whereas Japanese individuals would show more central face looking (cf., Blais et al., 2008). In addition, it was predicted that both cultural groups would show increased mouth looking in the dynamic-neutral condition as a result of greater motion on the mouth. For the dynamic-social condition, it was expected that Japanese participants would look more at the eye region, while the British group would show greater mouth looking (cf., Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013). For all groups, increased eye scanning was also expected in the dynamic-social compared to the dynamic-neutral condition given the greater motion in the eye region of dynamic-social faces.

Previous findings on the role of face ethnicity in modulating scanning behaviour were inconsistent (Fu et al., 2012; Liu et al., 2011; Wheeler et al., 2011; Xiao et al., 2013, versus, e.g., Geangu et al., 2016; Senju, Vernetti, Kikuchi, Akechi, & Hasegawa, 2013). However, given that such inconsistencies in ethnicity effects could be due to differences in analyses approaches rather than result from true cultural effects (Arizpe et al., 2016), it was expected that face ethnicity would not modulate eye movements. The current study used both dwell time and fixation measures to ensure that results would converge to similar conclusions.

Given that research on the development of executive function during the first year of life is still limited, no standardised tasks currently exist for infant populations (Hendry, Jones, & Charman, 2016). Typically, executive function tasks require verbal instructions; however, infants cannot comply with such verbal instructions, restricting research methodologies. More recent studies have been able to study early executive function using entirely non-verbal, gaze-contingent paradigms. For the current study, the gap-overlap paradigm (Hood & Atkinson, 1993) and cognitive control task (cf., Kovács & Mehler, 2009; Wass, Porayska-Pomsta, & Johnson, 2011; see Section 5.3.3.2 for more details) were selected. These gaze-contingent tasks have been used in previous studies with similar age groups to study executive function and attentional control more specifically (e.g., Holmboe, Bonneville-Roussy, Csibra, & Johnson, 2018; Wass et al., 2011). The gap-overlap task measures the ability to visually disengage from a stimulus in order to move overt attention to a new location (Hood & Atkinson, 1993; see Section 5.3.3.2.1 for a more detailed description). The ability to visually disengage significantly improves during the first year of life and has been suggested to represent a precursor or prerequisite for inhibitory control and controlled visual attention more generally

(Hendry et al., 2016; Holmboe et al., 2018). The cognitive control task examines anticipatory looking behaviour, testing rule learning during a pre-switch phase, and rule switching during a post-switch phase (Wass et al., 2011; see Section 5.3.3.2.2 for more details). Although cultural differences have been found in executive function during toddlerhood (cf., Imada et al., 2013; Lan et al., 2011; Oh & Lewis, 2008; Sabbagh et al., 2006), it remains unknown whether similar patterns can already be observed during infancy. For the current study, the preliminary expectation was that early cultural differences in these measures would be present, with Japanese infants exhibiting more advanced executive function compared to their British counterparts. It was also predicted that more advanced executive function would relate to increasingly culture-specific face scanning strategies. In sum, the purpose of this study was to fill gaps in the existing literature to obtain a more in-depth characterisation of cultural differences in face scanning during development, and to gain first insights into possible explanatory factors at the cognitive level.

5.3. Methods

5.3.1. Participants

The study was conducted in the UK and in Japan, with each cultural group consisting of 10-month-olds, 15- to 17-month-olds, and adults. Participants in the UK were tested at Birkbeck, University of London, and Japanese participants were tested at Kyoto University. Additional participant details will be described separately for infants and adults in Sections 5.3.1.1 and 5.3.1.2, respectively. The study was approved locally by the ethics committees of the Department of Psychological Sciences, Birkbeck, University of London, and the Department of Psychology, Kyoto University.

5.3.1.1. *Infants*

The 10-month-old group consisted of 26 infants in the UK (10 females, 16 males) and 22 infants in Japan (11 females, 11 males). The cultural groups did not differ in age (British: M = 307.00 days, SD = 11.07 days, ranging from 288 to 330 days; Japanese: M = 306.32 days, SD = 8.43 days, ranging from 289 to 323 days; $t(46) = 0.24$, $p = 0.814$). In the UK, an additional 12 infants were tested but excluded from analysis due to flickering gaze data (N = 8), equipment failure (N = 1), fussiness (N = 1), failed calibration (N = 1), or not meeting ethnicity requirements for this study (N = 1). In Japan, an additional four infants were tested but excluded from analysis due to fussiness.

The 15- to 17-month-old group consisted of 26 infants in the UK (11 female, 15 males) and 15 infants in Japan (5 females, 10 males). No (cultural) group differences were found for age (British: M = 474.08 days, SD = 18.39 days, ranging from 446 to 507 days; Japanese: M = 481.07 days, SD = 20.82 days, ranging from 449 to 534 days; $t(39) = 1.12$, $p = 0.271$). Two additional infants were tested in the UK but excluded from analysis due to flickering gaze data (N = 1) or fussiness (N = 1). In Japan, the data from one infant was excluded due to fussiness during the study protocol.

All infants were full-term, had normal vision and hearing and no developmental condition as reported by their parents. Demographic information about infants and their caregivers was collected using a questionnaire (see Appendix E). Infants tested in the UK were born and raised in the UK (except for one 10-month-old infant born in Germany), were of White ethnicity, had never lived in a country outside the UK, and their caregiver communicated with them in English. Infants tested in Japan were born and raised in Japan, were of Japanese ethnicity (except for one 15-month-old infant whose secondary caregiver was of White ethnicity), had never lived in a country outside Japan, and their caregiver communicated with them in Japanese.

Compared to Japanese families, infants tested in London may have been more likely to encounter different ethnicities (including Japanese) due to the city's multicultural environment. To ensure that exposure to other ethnicities did not significantly differ between the two cultural groups, the accompanying caregivers were asked to rate their child's amount of contact with Western Caucasian and East Asian people on a scale from 1 (very little) to 7 (very extensive). Descriptive statistics are reported in Table 5.1. Within each age group, independent *t*-tests revealed no significant rating differences between cultures with respect to the amount of contact with same-ethnicity individuals (10 months: $t(45) = -1.84$, $p = 0.072$; 15-17 months: $t(39) = -0.92$, $p = 0.365$) and with other-ethnicity individuals (e.g., East Asian ethnicities for infants tested in London; 10 months: $t(46) = 1.36$, $p = 0.181$; 15-17 months: $t(39) = 2.03$, $p = 0.050$).

Table 5.1. *Means and standard deviations of ratings for amount of contact with same-ethnicity and other-ethnicity individuals.*

	Contact with foreign culture		Contact with own culture	
	UK	Japan	UK	Japan
10 months	1.85 (0.93)	1.41 (1.30)	6.62 (0.64)	6.90 (0.44)
15 to 17 months	1.85 (0.93)	1.27 (0.80)	6.81 (0.49)	6.93 (0.26)

The session lasted up to 45 minutes, allowing also for short breaks and play time to familiarise infants with the testing environment. Families were recruited via internal databases of the local university departments and volunteered their infants to take part. In line with standard departmental protocols, families in the UK were reimbursed travel expenses and received a T-Shirt and certificate of participation, and Japanese families were reimbursed ¥3000 for their time. The accompanying caregiver provided written informed consent prior to the study.

5.3.1.2. *Adults*

Thirty-one adults in the UK (16 females, 15 males; aged $M = 27.35$ years, $SD = 6.72$ years, ranging from 19 to 40 years) and 30 adults in Japan (17 female, 13 males; aged $M = 21.73$

years, *SD* = 2.63 years, ranging from 18 to 31 years) took part in the study. Two additional adults were tested in the UK but excluded from analysis due to flickering gaze data ($N = 1$) or equipment failure ($N = 1$). In Japan, three participants were excluded due to flickering gaze data.

Participants in the UK were born and raised in the UK, were of White ethnicity, had never lived in a country outside Western Europe or North America, and indicated English as their native language. Japanese participants were born and raised in Japan, were of Japanese ethnicity, had never lived in a country outside East Asia, and indicated Japanese as their first language. All participants had normal or corrected-to-normal vision and hearing. The study lasted approximately 30 minutes and, in line with departmental guidelines, each participant received £8 (London) or ¥1000 (Kyoto) for their time. Participants were recruited via internal databases of Birkbeck, University of London, and Kyoto University. Each participant provided written informed consent prior to the study.

5.3.2. Apparatus

Eye movements were recorded using a Tobii TX300 eye tracker (Tobii Technology, Sweden) at a sampling rate of 120 Hz. Although the TX300 eye tracker can run at a maximum sampling rate of 300 Hz, a lower rate was chosen to improve the quality of data collected from infants (cf., Saez de Urabain et al., 2015). All stimuli were presented at 1920 x 1080 resolution on a 23" monitor that was mounted onto the eye tracker. The experimental protocol was monitored and controlled through MATLAB (R2013a 64-bit, MathWorks) and used the Psychophysics toolbox (Version 3; Shukla, Wen, White, & Aslin, 2011). The protocol was implemented in MATLAB to allow for greater control over the experimental tasks, including the possibility for gaze-contingent paradigms. Two external speakers, each located next to one side of the monitor, were used to play sounds. All participants were monitored via the built-in webcam of the monitor, and the DwayneCam app (Version 1.4) was used to live-stream the footage to the experimenter laptop. A small external camera was also used to record the screen of the experimenter laptop to allow for post-hoc data quality checks.

5.3.3. Design

All participants were presented with images and videos of faces, the gap-overlap paradigm, the cognitive control task, and stimuli for post-hoc data quality checks. Each experimental task will be outlined in more detail below.

5.3.3.1. *Face scanning task*

The face scanning task included three different conditions: a *static* condition whereby an image of a British or Japanese face with neutral expression was presented (see Section 5.3.3.1.2); a *dynamic-neutral* condition showing a video of a speaking British or Japanese face with neutral expression (see Section 5.3.3.1.3); and a *dynamic-social* condition whereby a British or Japanese actor performed baby-friendly face actions (see Section 5.3.3.1.4). The following sections will first describe how the stimuli were created before going on to outlining the implementation of the face scanning task within the experimental protocol.

5.3.3.1.1. *Creating the stimuli*

Video recordings of 12 female actors (six British, six Japanese) aged between 25 and 35 were used as stimuli. Each actor provided recordings for all three stimulus conditions, but only one recording was presented to a single participant (i.e., a participant never saw the same actor more than once; for more details see Section 5.3.3.1.5). To minimise distractions, all actors wore a black T-Shirt, removed jewellery, glasses, and visible make-up, the hair was tied back, and actors were recorded against a black background. All actors also had dark brown or black hair to ensure that unfamiliar hair colours would not systematically distract participants (e.g., blonde hair is relatively uncommon in Japan).

Each actor was recorded individually at a camera resolution of 1920 x 1080. The faces were recorded in frontal view, with actors making direct eye contact with the camera. All stimuli were edited and exported using Final Cut Pro X Version 10.0.8. Face stimuli were in colour, controlled for luminance, and videos were exported at a stable frame rate of 25 frames per second. The size and position of the faces were also controlled, with each face roughly measuring 16.5° (height) x 12.0° (width) (excluding hair and ears) to represent faces in life-size (at 65 cm viewing distance), and all faces were aligned at the midpoint between the nose tip and the bridge (region between the eyes). Timings of facial actions between actors were matched with the use of a metronome that played at 60 beats per minute. Given that speech can modulate face scanning in infants (Bahrick, Netto, & Hern, 1998; Dodd, 1979), the original video sounds were muted and instead overlaid with unsynchronised instrumental background music by the folk group *The Chieftains*. Since video sounds were muted during the editing process, it was possible to use a metronome that helped actors to focus on the production of facial actions.

5.3.3.1.2. *Static face stimuli*

In total, 12 static images were created (six Japanese, six British). Each actor was recorded for 3 seconds (at 25 frames per second), and a suitable frame was selected as the face image. Figure 5.2 shows example stimuli for a British and a Japanese face in the static condition. For the full list of static face stimuli go to Appendix D.

Figure 5.2. Example stimuli used in the static condition.

5.3.3.1.3. *Dynamic-neutral face stimuli*

Twelve dynamic-neutral videos were created (six Japanese, six British). Faces always had a neutral expression, and actors were asked to speak the syllables *do re mi fa sol la ti do*. These syllables were chosen since they occur in both the English and Japanese language, and therefore minimise language-specific mouth movements. Using such familiar syllable sequences (as opposed to unfamiliar non-words) also facilitated the filming process by helping actors focus on the production of facial actions.

During the first second of the video, the actor made direct eye contact with the camera and did not move. This was followed by speaking the sequence in which each syllable lasted 1 second (i.e., 8 seconds in total for the sequence). This action was then repeated, resulting in a total video duration of 18 seconds. Figure 5.3 shows randomly selected video frames for the stimuli used in the dynamic-neutral condition. Appendix D provides a screenshot of each dynamic-neutral face stimulus.

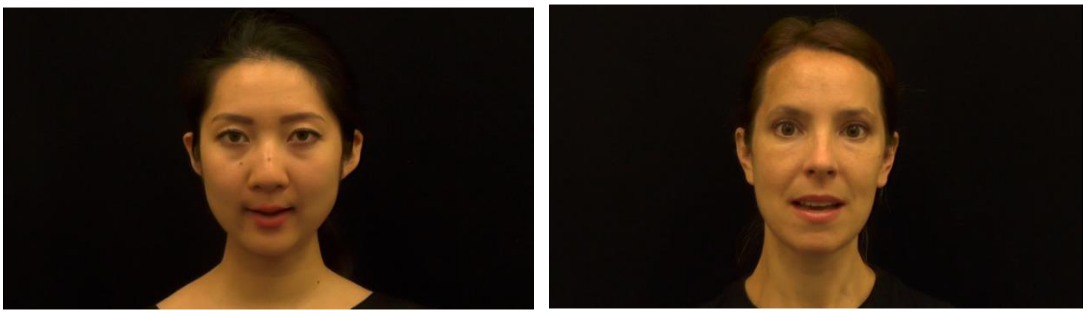


Figure 5.3. Example stimuli used in the dynamic-neutral condition.

5.3.3.1.4. *Dynamic-social face stimuli*

Thirty-six dynamic-social videos were initially created (three facial actions; six Japanese, six British). During the piloting stage, the facial action that elicited least attention in infants was dropped, and the other two facial actions were retained for the experiment.

As with the dynamic-neutral condition, the dynamic-social videos were recorded by asking actors to speak the syllables *do re mi fa sol la ti do*. During the first second of the video, actors did not move, and smiled (without showing teeth) while making direct eye contact with the camera. This was followed by speaking the sequence for which each syllable lasted 1 second (i.e., 8 seconds in total for the sequence). For the *nodding* action, actors were asked to smile while speaking the first two syllables (*do re*), to continue smiling while nodding the head towards their right for the subsequent two syllables (*mi fa*), to nod the head toward their left for the next two syllables (*sol la*), and to return the head to the central position for the last two syllables while continuing to smile (*ti do*). This 9-second sequence was then repeated, giving a total video duration of 18 seconds. Actors were also asked to avoid nodding head movements that were too far left or right to avoid rapidly moving shifts in regions-of-interest positions. This was done since infants may not initiate saccades to new locations as fast as adults (Colombo, 2001), which in turn would lead to systematic age differences in regions-of-interest looking times. For the *peekaboo* action, actors were asked to smile while speaking the first two syllables (*do re*), to cover their face with their hands for the subsequent two syllables (*mi fa*), to uncover their face (such that the hands were out of sight) and look surprised for the next two syllables (*sol la*), and to continue smiling for the final two syllables (*ti do*). This sequence was then repeated to obtain a total video duration of 18 seconds. For all dynamic-social videos, actors were instructed to move their upper face (eyes and eye brows) and lower face regions (mouth) to create salient facial expressions. Given that faces were occasionally occluded in the peekaboo condition, these short periods were excluded from analysis. Figure 5.4 on page 168 shows randomly selected video frames for the stimuli used in the dynamic-social condition. Appendix D provides a screenshot of each dynamic-social face stimulus.

Figure 5.4. Example stimuli used in the dynamic-social condition.

5.3.3.1.5. *Implementation of the face scanning task*

Each face scanning trial was preceded by a gaze-contingent fixation point located in the centre of the screen to control the location of the first fixation and to ensure that the trial could only be triggered when participants were attending to the screen. Once the trial began, a face was presented for 18 seconds.

Every participant was presented with four face identities (two British and two Japanese) for each of the three conditions. In other words, each participant viewed twelve face identities across the experimental protocol: four static faces, four dynamic-neutral faces, and four dynamic-social faces. An identity was never repeated throughout the experiment to minimise boredom in infants and to avoid familiarisation effects (e.g., O'Donnell & Bruce, 2001). The order of stimulus condition was the same for every participant – first static, then dynamic-neutral, and finally dynamic-social – and condition blocks were interleaved with other experimental tasks (see Section 5.3.4). This was done to account for carry-over effects of eye movement behaviour which in turn could mask group differences, and to increase infant attention throughout the testing session. Specifically, since infants become less attentive over the course of testing, it is useful to employ more engaging social stimuli at the end of the protocol. Within each condition, face identities of the same ethnicity were grouped together and presented in sequence. The order of face ethnicity, as well as the order of identities across the paradigm, was counterbalanced across participants and groups. For the dynamic-social condition, the order of facial actions (nodding, peekaboo) was also randomised.

5.3.3.2. *Gaze-contingent paradigms*

5.3.3.2.1. *Gap-overlap paradigm*

Each trial started with a gaze-contingent central stimulus (an animated clock; see Figure 5.5) measuring a maximum of 4.41° x 4.41° visual angle at 65 cm viewing distance. Once the participant fixated the central stimulus, the peripheral stimulus (an animated cloud measuring approximately 2.64° x 2.64° visual angle; see Figure 5.5) was presented randomly on the left or right side of the screen at an eccentricity of 18.35° visual angle.

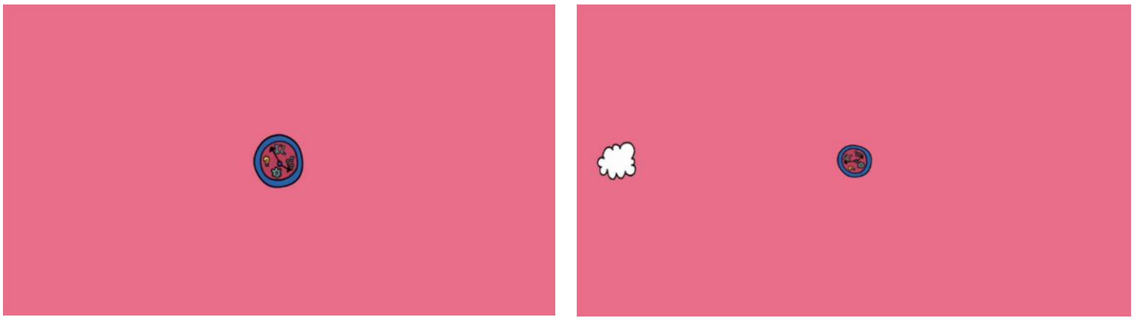


Figure 5.5. Screenshots from the gap-overlap paradigm displaying the central stimulus (left) and also the peripheral stimulus (right).

The timing of the peripheral stimulus onset differed depending on the trial condition. In the *gap* condition, the central stimulus disappeared after a jitter of 600 to 700 milliseconds and the peripheral stimulus was displayed after a 200-millisecond delay. In the *overlap* condition, the peripheral stimulus was presented after the central stimulus was fixated, and the central stimulus remained on screen until the end of the trial. In the *baseline* condition, the peripheral stimulus was also presented immediately, but the central stimulus disappeared upon fixation (after a jitter of 600 to 700 milliseconds). Once the participant fixated the peripheral stimulus, a rewarding infant-friendly animation was played before the next trial was triggered. If the participant did not fixate the peripheral stimulus after 2 seconds, the next trial would be triggered. The background colour remained constant across trials (see Figure 5.5). Each gap-overlap block consisted of 12 trials, with 4 trials in each condition (gap, overlap, and baseline), and trial condition was randomised within each block. In total, four blocks (48 trials) were run, and an additional fifth block was presented toward the end of the protocol only if a participant completed less than 12 valid trials per condition. The blocks were alternated with other experimental tasks in the protocol (see Section 5.3.4).

A single trial was considered valid if all the following conditions were met:

(1) The gaze remained within the region of the central stimulus until the onset of the peripheral stimulus.

(2) The peripheral stimulus was fixated within 1200 milliseconds after onset (cf., Elsabbagh et al., 2013, 2009).

(3) The peripheral stimulus was not fixated within 200 milliseconds after onset (cf., Wass et al., 2011).

(4) No saccade was made toward the opposite side of the peripheral stimulus.

The median latency required to saccade from the central to the peripheral stimulus was calculated separately for each condition. For the current analysis, the disengagement latencies were calculated by subtracting the baseline condition from the overlap condition (e.g., Elsabbagh et al., 2013).

5.3.3.2.2. *Cognitive control task*

The paradigm of the cognitive control task used the same parameter values as in Wass et al. (2011). The task consisted of 18 trials (nine pre-switch and nine post-switch trials). Each trial started with an animated gaze-contingent fixation point in the centre of the screen to control for the location of the first fixation and to ensure the trial only started when the participant was attending to the screen (see Figure 5.6 on page 171). Once the participant looked at the fixation point, the anticipatory period was triggered during which two purple rectangles were displayed and only the audio of the clip was played through the speakers (see Figure 5.6 on page 171). The participant had 2 seconds to saccade toward the correct side of the screen (i.e., toward the correct rectangle) to trigger the rewarding clip in a gaze-contingent manner (see Figure 5.6 on page 171). During the nine pre-switch trials, the clip was always shown on the same side before switching to the opposite side for the nine post-switch trials. Once the participant looked at the correct rectangle, the clip was played for 4 seconds before the next trial was triggered. If the participant did not look at either rectangle after 2 seconds, the rewarding clip was played for 4 seconds and the next trial was initiated.

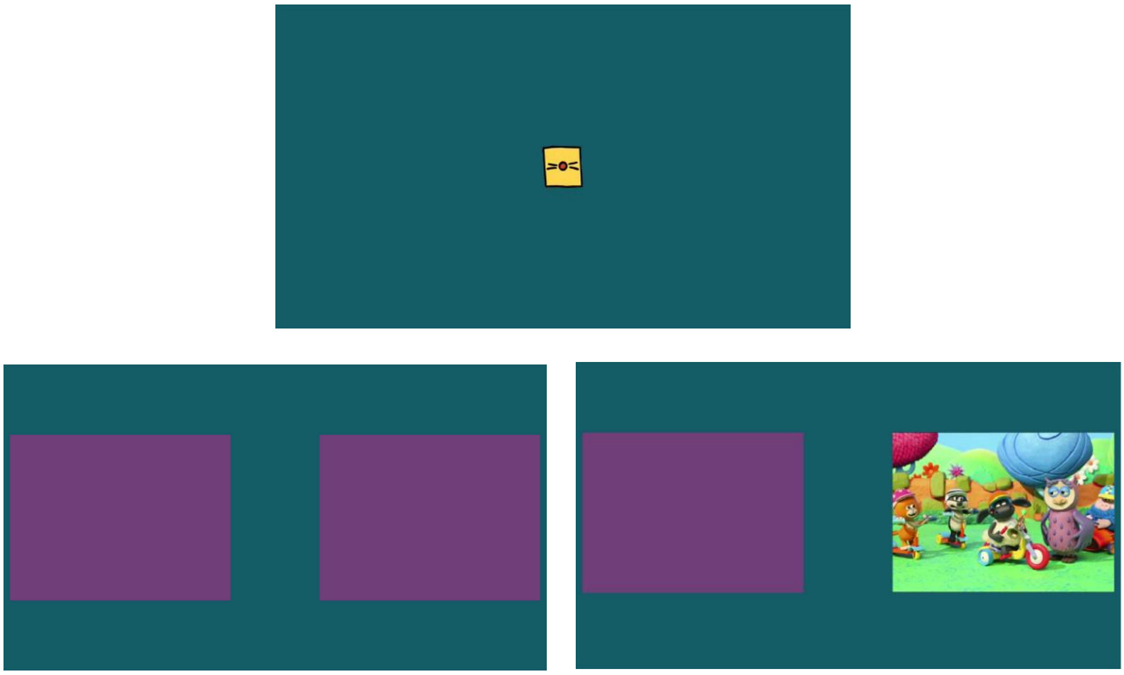


Figure 5.6. Screenshots from the cognitive control task displaying the central fixation point (top), the two rectangles during the anticipatory period (bottom left), and the rewarding clip (bottom right).

The location of the clip for the pre-switch phase was determined in the following way: if the participant looked at the left rectangle in the first trial, the clip would be played on the right side throughout the pre-switch phase. If the participant first looked at the right rectangle, the clip would be played on the left side. This ensured that the clip would not be presented on the participants' preferred side. If, however, the participant did not look at either rectangle after 2 seconds, the location of the clip would be determined randomly. In total, two cognitive control blocks were included in the protocol. The clips were taken from *Pingu* and *Timmy Time*, which do not contain language and therefore represent suitable baby-friendly stimuli for cross-cultural studies. The order of the clips was counterbalanced across participants and the blocks were alternated with other experimental tasks in the protocol (see Section 5.3.4).

A single trial was coded as a hit if the participant looked at the correct side of the screen within 2 seconds; it was coded as a miss otherwise. For each participant, the proportion of hits relative to the number of trials was calculated separately for the pre- and post-switch phase. The first trial of each phase was always excluded from analysis since the participant could not anticipate the location of the clip. For the current study, only the first block was used for analysis. If the participant did not successfully complete the first block, the second block was used instead.

5.3.3.3. *Post-hoc data quality checks*

Given that factors such as age and ethnicity can affect data quality (Blignaut & Wium, 2014; Saez de Urabain et al., 2015), any observed age and cultural differences in face scanning could potentially be confounded by group differences in spatial accuracy. For instance, if eye tracking data is lower in spatial accuracy for infants than for adults, gaze points may fall into different regions-of-interest, resulting in age differences in face scanning. To account for group differences in spatial accuracy, infant-friendly spiralling points were presented throughout the protocol for post-hoc data quality checks.

Each point was presented as an animated, inward-turning spiral (see Figure 5.7 on page 173). Twenty raw gaze data points were collected just before the disappearance of the spiral – i.e., when the spiral reached its smallest size (see Figure 5.7 on page 173) – and the true *x*- and *y*-positions of the spiral were recorded. The mean *x*- and *y*-coordinates of the 20 gaze points was also obtained, and the distance to the spiral was calculated. For each participant, the mean distance measure, in degrees visual angle, was then used for analysis (see Table 5.2 on page 173 for descriptive statistics).

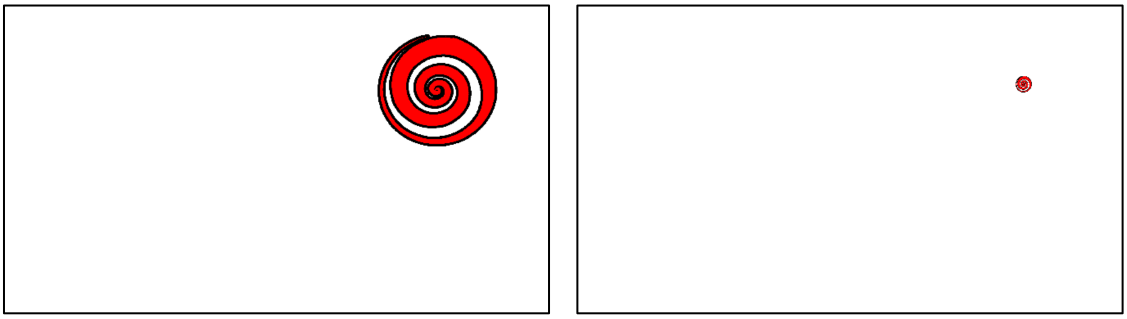


Figure 5.7. Screenshots of a spiral at the start (left) and at the end of the presentation (right).

Table 5.2. *Mean spatial offsets and standard deviations (in degrees visual angle) for each experimental group.*

	UK	Japan
10 months	1.16 (0.47)	0.79 (0.36)
15 to 17 months	0.93 (0.35)	0.86 (0.28)
Adults	1.10 (1.11)	0.93 (0.61)

A 2 (Group: British, Japanese) x 3 (Age: 10 months, 15 to 17 months, adults) ANOVA was conducted on the offset measures (in degrees visual angle). No significant main effects or interactions were found, suggesting that spatial accuracy did not differ between groups (Group: $F(1,136) = 3.15$, $p = 0.078$; Age: $F(2,136) = 0.37$, $p = 0.690$; Group x Age: $F(2,136) = 0.54$, $p = 0.584$).

5.3.4. Procedure

Participants were welcomed in the reception room where the experimenter explained the study, collected written informed consent, and asked caregivers/adult participants to fill out a demographic questionnaire. During this period, infants were able to familiarise themselves with the new environment. The participants were then guided to the testing room and sat either on a chair (adults) or on the caregiver's lap (infants) at a distance of approximately 65 cm from the screen. Once the eye tracker was accurately positioned, participants completed a five-point calibration procedure which involved inward-moving spirals along with an attention-grabbing sound. The calibration spirals were the same as the stimuli used for the points for post-hoc data quality checks (see Section 5.3.3.3). Calibration performance was visualised on the experimenter laptop, and the study began once the experimenter approved the calibration data. After successful five-point calibration, the study protocol was started. Each participant was presented with:

(1) Two static face images (White-British or Japanese)

(2) Twelve gap-overlap trials

(3) Two static face images, with ethnicity not shown in (1)

(4) Twelve gap-overlap trials

(5) One cognitive control task

(6) Two dynamic-neutral face videos (White-British or Japanese)

(7) Twelve gap-overlap trials

(8) Two dynamic-neutral face videos, with ethnicity not shown in (6)

(9) Twelve gap-overlap trials

(10) One cognitive control task

(11) Two dynamic-social face videos (White-British or Japanese)

(12) If less than twelve valid trials per gap-overlap condition were collected, an additional twelve trials were presented

(13) Two dynamic-social face videos, with ethnicity not shown in (11).

A spiral for post-hoc data quality checks was included between every task reported above. When infants became inattentive, an auditory attention grabber was triggered by the experimenter who monitored participants via a webcam. In the case that infants became fussy, the protocol was interrupted to allow for a play break and resumed after an additional five-point calibration procedure, or the study was stopped.

5.3.5. Data pre-processing

5.3.5.1. *Fixation detection and coding*

The raw eye tracking data was processed using GraFIX (Saez de Urabain et al., 2015) for data smoothing, interpolation, and subsequent fixation detection and coding. As outlined in Chapter 2, GraFIX was chosen due to its two-step approach involving rapid automatic pre-processing of gaze data followed by an optional moderation stage for fixation coding. In other words, GraFIX can take into account any variability in data quality across participants and, importantly, across age and cultural groups by allowing the user to manually override automatically coded fixations in an efficient manner.

For the present analysis, the automatic parameter values for smoothing, interpolation, and the criteria for fixation detection remained the same across participants (see Table 5.3), and values were chosen based on Saez de Urabain et al. (2015) and Wass et al. (2013). After automatic processing, GraFIX also provides the smoothed and interpolated gaze data, which was used for the dwell time analysis.

Table 5.3. *Parameter settings for automatic processing of gaze data in GraFIX.*

Parameter	Value
Fill missing data with opposite eye	Yes
Smoothing time (samples)	30
Smoothing space (samples)	21
Velocity threshold (°/second)	20
Maximum interpolation latency (milliseconds)	150
Maximum interpolation displacement (°)	0.25
Merge consecutive fixations with similar location (°)	0.25
Maximum root mean square per fixation (°)	0.30
Minimum fixation duration (milliseconds)	100

Manual fixation coding was also conducted using the following guidelines:

(1) Fixations without a clear beginning and end are deleted.

(2) A fixation will be coded if a clear beginning and end velocity spike can be observed and gaze data between the spikes remains relatively stable. A velocity spike exceeding the threshold – usually representing noisy data and not a saccade – typically prevents the fixation to be detected automatically. For highly noisy data, a fixation will also be coded if the smoothed data forms separate gaze lines (as opposed to dot clusters due to highly variable data).

(3) Two or more automatically coded fixations separated by very short periods of missing data are merged if the fixation positions remain constant.

(4) Fixations will be merged if they are separated by microsaccades within the velocity threshold.

(5) Very long fixations (> 5 seconds) characterised by a small, progressive drift in eye position are deleted since they likely indicate tiredness or boredom.

To ensure reliability in the number of detected fixations as well as the mean fixation durations, an external coder without eye tracking experience and who was naïve to the study aims processed 20% of the data using the above parameters and guidelines. Datasets for second coding were chosen randomly and were distributed evenly across age and cultural groups. Following procedures reported in Saez de Urabain et al. (2015), the Intraclass Correlation Coefficient (ICC; Hallgren, 2012) was used, revealing an excellent agreement for the number of detected fixations (ICC of 0.93, $p < 0.001$) and a good agreement for the mean durations of fixations (ICC of 0.75, $p < 0.001$).

5.3.5.2. *Regions-of-interest coding*

Based on previous developmental studies on cultural differences in face scanning (e.g., Senju, Vernetti, Kikuchi, Akechi, & Hasegawa, 2013; Wheeler et al., 2011), the following four non-overlapping regions-of-interest were coded to examine face scanning behaviour: eyes, bridge (between the eyes), nose, mouth (see Figure 5.8). Using MATLAB, the outline of each region was manually drawn for the first frame of a stimulus, and the outlines were then automatically superimposed on all subsequent frames, with manual coding only being performed when regions-of-interest positions had changed in size, position, or angle. The inner face outline (excluding hair and ears) was additionally coded to calculate dwell time and fixation time in each region-of-interest as a proportion of overall inner face looking time. Gaze data for the left and right eye was collapsed to form a single region-of-interest for the eyes. Dependent measures were then analysed using mixed ANOVAs (see Section 5.4.1).

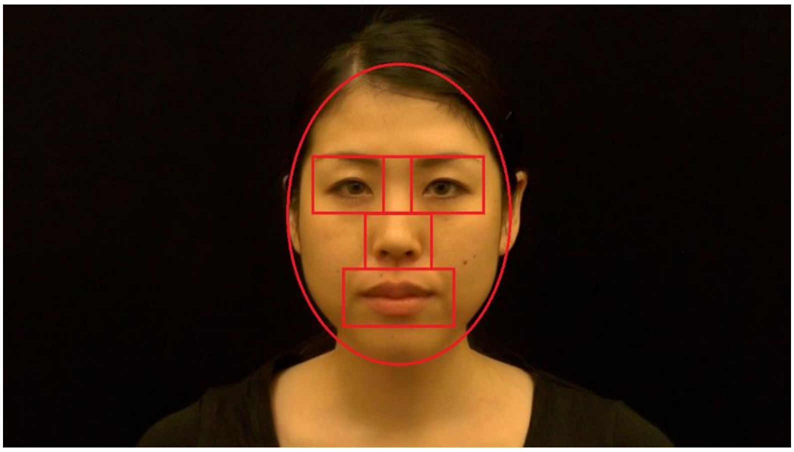


Figure 5.8. Outlines of regions-of-interest superimposed onto a face.

5.4. Results

5.4.1. Regions-of-interest (ROI) analysis

Dwell time in each ROI was calculated as a proportion of inner face looking time. Dwell time means and standard deviations for each ROI are presented from Table 13.1 to Table 13.12 in Appendix F. Fixation time in each ROI was computed relative to the total fixation time directed to the inner face region, with corresponding descriptive statistics presented from Table 13.16 to Table 13.27 in Appendix F. Overall, the descriptive data indicated similar proportions for face stimuli of Japanese and White-British ethnicity, with participants mostly scanning the eyes and the mouth. Eye scanning was most prominent in the static condition, whereas moving faces gave rise to increased mouth looking. British participants also scanned the mouth more than Japanese individuals, with mouth looking peaking at 15-17 months, particularly for moving faces. A reversed pattern could be observed for eye scanning.

A five-way mixed ANOVA was conducted, separately for dwell time and fixation time. The following factors were included in the model: Group (British, Japanese), Age (10 months, 15-17 months, adults), Stimulus Condition (static, dynamic-neutral, dynamic-social), Face Ethnicity (White-British, Japanese), and ROI (eyes, bridge, nose, mouth). When the assumption of sphericity was violated, degrees of freedom were corrected using Greenhouse-Geisser estimates.

5.4.1.1. *Five-way ANOVA: Group x Age x Stimulus Condition x Face Ethnicity x ROI*

5.4.1.1.1. *Dwell time analysis*

Full results of the five-way ANOVA on dwell time data are displayed in Table 13.13 in Appendix F, indicating also all significant main effects and interactions. Crucially, a significant five-way interaction was found (Group x Age x Stimulus Condition x Face Ethnicity x ROI: $F(8.02,577.71) = 2.29$, $p = 0.020$, $\eta_p^2 = 0.031$). To break down this interaction, two four-way ANOVAs (Group x Age x Stimulus Condition x ROI) were conducted separately at each level of the Face Ethnicity factor. Detailed results of these four-way ANOVAs are given in Table 13.14 and Table 13.15 in Appendix F for face stimuli of Japanese and White-British ethnicity, respectively. The significant main effects and interactions were consistent across both four-way ANOVAs, except for an additional significant main effect of Stimulus Condition $F(1.82,261.85) = 9.37$, $p < 0.001$, $\eta_p^2 = 0.061$) and ROI x Age x Group interaction for White-British faces ($F(3.66,263.55) = 2.72$, $p = 0.035$, $\eta_p^2 = 0.036$). The highest-order significant effects (also common to both analyses) were the three-way interactions Stimulus Condition x

ROI x Age (Japanese face ethnicity: $F(8.25,594.28) = 1.93$, $p = 0.051$, $\eta_p^2 = 0.026$; White-British face ethnicity: $F(7.86,565.95) = 2.11$, $p = 0.034$, $\eta_p^2 = 0.028$), and Stimulus Condition x ROI x Group (Japanese face ethnicity: $F(4.13,594.28) = 3.11$, $p = 0.014$, $\eta_p^2 = 0.021$; White-British face ethnicity: $F(3.93, 565.95) = 2.45$, $p = 0.047$, $\eta_p^2 = 0.017$). Before examining these significant three-way interactions further, the following section will first present the initial findings of the five-way ANOVA using fixation time data.

5.4.1.1.2. *Fixation time analysis*

Detailed results of the five-way ANOVA are displayed in Table 13.28 in Appendix F. In contrast to the dwell time analysis (Section 5.4.1.1.1), the five-way interaction for fixation data did not reach significance ($F(9.29,668.65) = 1.84$, $p = 0.057$). The ROI x Age x Group interaction that was significant for dwell time data also did not reach significance ($F(3.71,267.43) = 2.37$, $p = 0.057$), suggesting that cultural differences in face scanning did not significantly change across age groups.

The highest-order significant interactions were Stimulus Condition x ROI x Group ($F(3.99,575.12) = 2.62$, $p = 0.034$, $\eta_p^2 = 0.018$), and Stimulus Condition x ROI x Age ($F(7.99,668.65) = 2.52$, $p = 0.011$, $\eta_p^2 = 0.034$), which were also found for the dwell time analysis. These three-way interactions were further examined below. Altogether, the fixation analysis found that face ethnicity did not differentially modulate scanning patterns for British or Japanese participants. Finally, although the Stimulus Condition x Face Ethnicity x ROI interaction was also found to be significant for the fixation analysis ($F(4.64,668.65) = 2.37$, $p = 0.042$, $\eta_p^2 = 0.016$), this was not investigated further given that this study was primarily concerned with age and cultural differences in scanning behaviour.

5.4.1.2. *Cultural differences in face scanning: Examining the Stimulus Condition x ROI x Group interaction*

The significant three-way interaction Stimulus Condition x ROI x Group found for both the dwell time (Section 5.4.1.1.1) and fixation time analysis (Section 5.4.1.1.2) was further broken down by conducting ROI x Group ANOVAs separately at each level of Stimulus Condition (static, dynamic-neutral, dynamic-social). Given that a central research question was concerned with cultural differences in scanning behaviour, the Group factor was retained within the model to examine group differences directly. Since the question regarding stimulus-dependent scanning strategies was secondary to this study, the follow-up ANOVAs were conducted at each level of Stimulus Condition and the ROI factor was retained. For the dwell time data, the follow-up ANOVAs were also conducted at each level of Face Ethnicity.

5.4.1.2.1. *Dwell time analysis*

The descriptive data is visualised in Figure 5.9 on page 182 and Figure 5.10 on page 183. Results for the ROI x Group analysis on dwell time data are provided in Table 13.29 to Table 13.34 in Appendix F. A consistent finding was observed in all ANOVAs across the different levels of Face Ethnicity and Stimulus Condition, revealing a significant main effect of ROI (static White-British: $F(1.64,242.96) = 251.18$, $p < 0.001$, $\eta_p^2 = 0.629$; dynamic-neutral White-British: $F(1.71,253.44) = 104.10$, $p < 0.001$, $\eta_p^2 = 0.413$; dynamic-social White-British: $F(1.92,284.77) = 45.01$, $p < 0.001$, $\eta_p^2 = 0.233$; static Japanese: $F(1.90,280.45) = 185.99$, $p < 0.001$, $\eta_p^2 = 0.557$; dynamic-neutral Japanese: $F(1.69,250.42) = 110.42$, $p < 0.001$, $\eta_p^2 = 0.427$; dynamic-social Japanese: $F(1.86,275.19) = 39.71$, $p < 0.001$, $\eta_p^2 = 0.212$).

A significant interaction of ROI x Group was also found (static White-British: $F(1.64,242.96) = 3.27$, $p = 0.049$, $\eta_p^2 = 0.022$; dynamic-neutral White-British: $F(1.71,253.44) = 6.27$, $p = 0.004$, $\eta_p^2 = 0.041$; dynamic-social White-British: $F(1.92,284.77) = 10.54$, $p < 0.001$, $\eta_p^2 = 0.066$; static Japanese: $F(1.90,280.45) = 6.37$,

$p = 0.002$, $\eta_p^2 = 0.041$; dynamic-neutral Japanese: $F(1.69,250.42) = 4.55$, $p = 0.016$, $\eta_p^2 = 0.030$; dynamic-social Japanese: $F(1.86,275.19) = 9.91$, $p < 0.001$, $\eta_p^2 = 0.063$).

However, no main effect of Group was revealed (for statistical values see Table 13.29 to Table 13.34 in Appendix F). In line with predictions, the significant ROI x Group interaction implies underlying cultural differences in face scanning. To compare dwell time for each face region between the two cultural groups, the ROI x Group interaction was broken down further by conducting independent *t*-tests at each level of ROI. A Bonferroni-adjusted alpha-level of 0.0125 was used (=0.05/4).

Detailed results of the independent *t*-tests are reported in Table 13.35 in Appendix F. No significant cultural differences were found for the eyes or bridge in any condition (see Table 13.35 in Appendix F), such that the prediction for Japanese participants to show increased looking at the eye region for dynamic-social faces could not be supported. In addition, the lack of a group difference for the eyes suggests that British participants did not exhibit greater triangular scanning of the eyes and mouth for static and dynamic-neutral faces. However, British participants were found to exhibit significantly more mouth looking across all conditions, supporting the prediction that the British group would engage in greater mouth looking for dynamic-social faces (all $p \leq 0.008$; for dynamic-neutral Japanese faces, this effect was marginally significant after Bonferroni correction at $p = 0.022$).

An additional consistent significant effect indicated that Japanese participants engaged in greater scanning of the nose in dynamic-neutral and dynamic-social conditions (all $p \leq 0.011$). This is in line with the prediction that Japanese participants would engage in more central face looking for dynamic-neutral faces. However, no significant group difference in nose scanning was found for the static condition, such that the prediction could not be fully supported.

In sum, findings from the dwell time analysis only partly support the predictions on the manifestation of cultural differences in face scanning. Overall, British participants scanned the mouth significantly more than Japanese individuals, and Japanese participants, meanwhile, tended to show greater central face looking for dynamic-neutral and social faces.

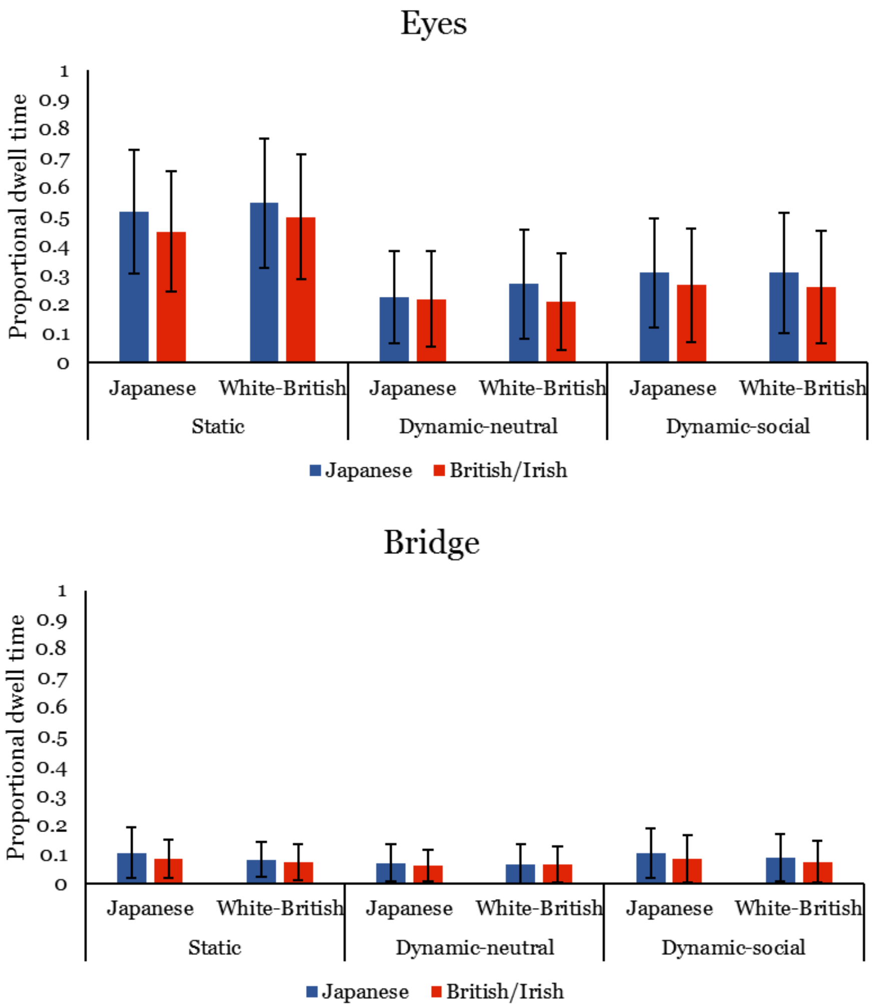


Figure 5.9. Proportional dwell times for the eye and bridge region, separately for stimulus condition, face ethnicity, and cultural group. Error bars represent +/- 1 SD.

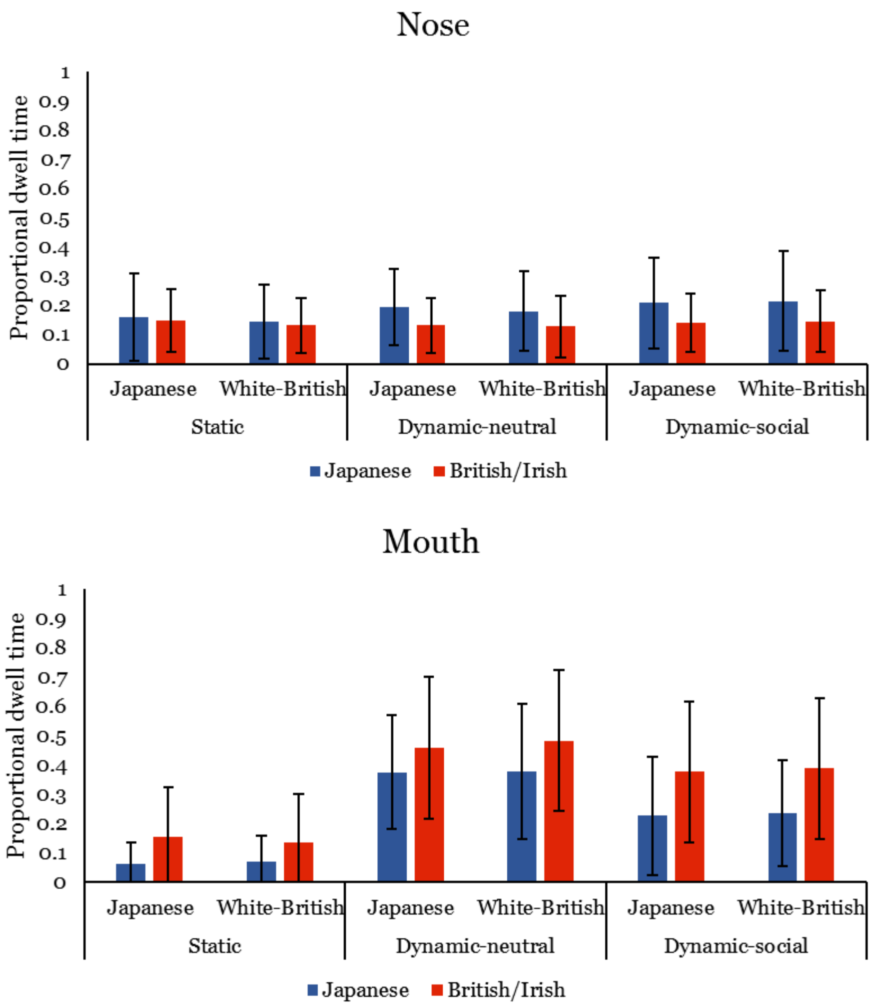


Figure 5.10. Proportional dwell times for the nose and mouth region, separately for stimulus condition, face ethnicity, and cultural group. Error bars represent +/- 1 SD.

5.4.1.2.2. *Fixation time analysis*

The descriptive data is visualised in Figure 5.11 on page 184. Results for the ROI x Group analysis on fixation data are provided in Table 13.36 to Table 13.38 in Appendix F. As with the dwell time analysis, a significant main effect of ROI was found (static: $F(1.71,252.32) = 244.00$, $p < 0.001$ $\eta_p^2 = 0.622$; dynamic-neutral: $F(1.75,258.86) = 115.97$, $p < 0.001$, $\eta_p^2 = 0.439$; dynamic-social: $F(1.89,279.17) = 48.91$, $p < 0.001$, $\eta_p^2 = 0.248$). A significant ROI x Group interaction was also revealed across all stimulus conditions (static: $F(1.71,252.32) = 5.51$, $p = 0.007$ $\eta_p^2 = 0.036$; dynamic-neutral: $F(1.75,258.86) = 6.13$, $p = 0.004$, $\eta_p^2 = 0.040$; dynamic-social: $F(1.89,279.17) = 10.72$,

$p < 0.001$, $\eta_p{}^2 = 0.068$). As with the dwell time analysis, no main effect of Group was found (for statistical values see Table 13.36 to Table 13.38 in Appendix F). Results for the independent *t*-tests, using a Bonferroni-adjusted alpha-level of 0.0125, are reported in Table 13.39 in Appendix F. Significant effects were consistent with those of the dwell time analysis (all $p \leq 0.005$) with the exception that increased mouth looking in British participants for dynamic-neutral faces was only marginally significant after Bonferroni correction ($p = 0.015$).

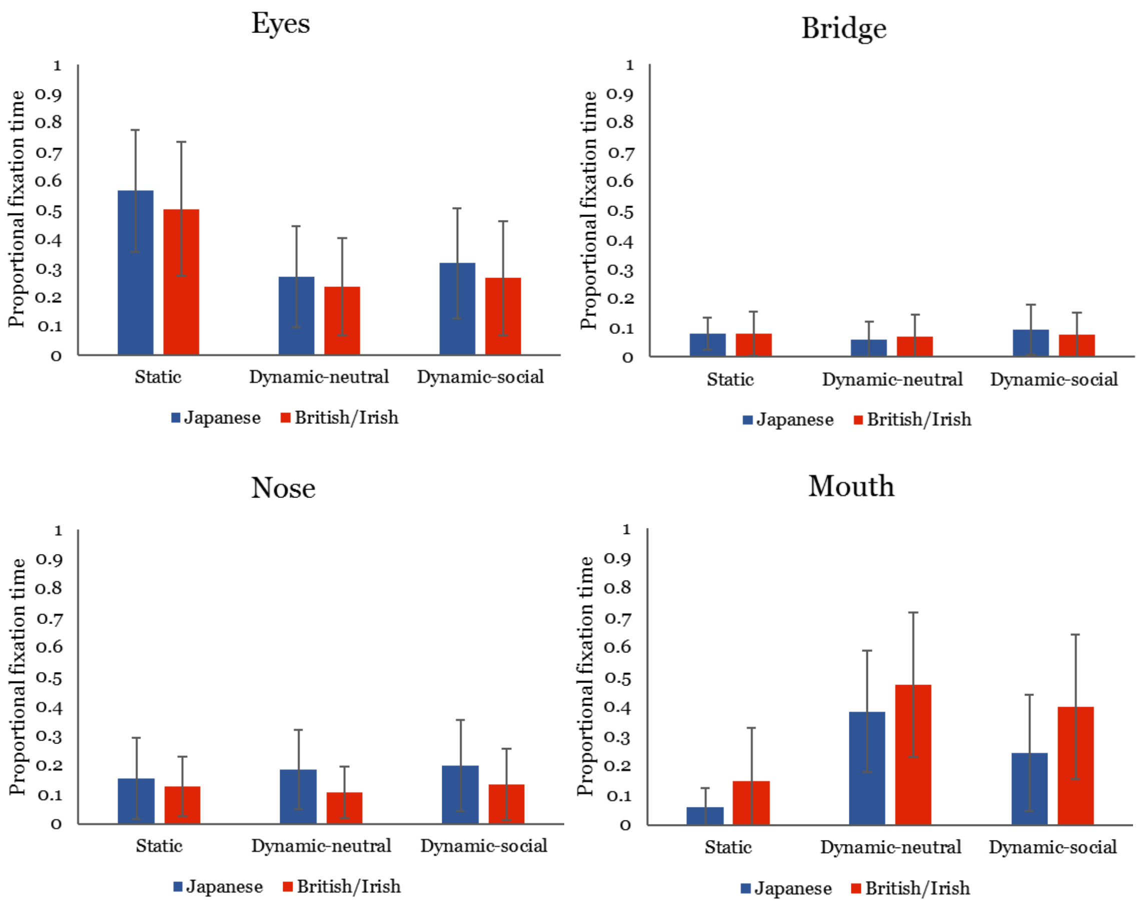


Figure 5.11. Proportional fixation times for each region-of-interest, stimulus condition, and cultural group. Error bars represent +/- 1 SD.

5.4.1.3. *Age differences in face scanning: Examining the Stimulus Condition x ROI x Age interaction*

The significant three-way interaction Stimulus Condition x ROI x Age found for both the dwell time (Section 5.4.1.2.1) and fixation time analysis (Section 5.4.1.1.2) was further broken down by conducting two-way ANOVAs (ROI x Age) separately for static, dynamic-neutral, and dynamic-social face conditions. Given that a central research question was of developmental nature, the Age factor was retained in the model to examine age differences.

5.4.1.3.1. *Dwell time analysis*

The descriptive data is visualised in Figure 5.12 on page 188 and Figure 5.13 on page 189. Findings for the ROI x Age ANOVA for dwell time data are provided in Table 13.40 to Table 13.45 in Appendix F. Across all levels of Face Ethnicity and Stimulus Condition, a significant main effect of ROI was found (static White-British: $F(1.67,245.37) = 252.99$, $p < 0.001$, $\eta_p^2 = 0.632$; dynamic-neutral White-British: $F(1.73,254.09) = 113.74$, $p < 0.001$, $\eta_p^2 = 0.436$; dynamic-social White-British: $F(1.94,285.49) = 51.66$, $p < 0.001$, $\eta_p^2 = 0.260$; static Japanese: $F(1.92,282.51) = 184.01$, $p < 0.001$, $\eta_p^2 = 0.556$; dynamic-neutral Japanese: $F(1.75,257.45) = 123.47$, $p < 0.001$, $\eta_p^2 = 0.456$; dynamic-social Japanese: $F(1.90,278.83) = 45.94$, $p < 0.001$, $\eta_p^2 = 0.238$).

A significant interaction of ROI x Age was also consistently found (static White-British: $F(3.34,245.37) = 5.14$, $p = 0.001$, $\eta_p^2 = 0.065$; dynamic-neutral White-British: $F(3.46,254.09) = 5.67$, $p < 0.001$, $\eta_p^2 = 0.072$; dynamic-social White-British: $F(3.88,285.49) = 6.80$, $p < 0.001$, $\eta_p^2 = 0.085$; static Japanese: $F(3.84,282.51) = 5.95$, $p < 0.001$, $\eta_p^2 = 0.075$; dynamic-neutral Japanese: $F(3.50,257.45) = 7.75$, $p < 0.001$, $\eta_p^2 = 0.095$; dynamic-social Japanese: $F(3.79,278.83) = 6.73$, $p < 0.001$, $\eta_p^2 = 0.084$).

However, the main effect of Age was less consistent (see Table 13.40 to Table 13.45 in Appendix F; significant effects were found for dynamic-neutral White-British: $F(2,147) = 3.97$, $p = 0.021$, $\eta_p^2 = 0.051$; dynamic-social White-British: $F(2,147) = 3.77$,

$p = 0.025$, $\eta_p^2 = 0.049$; dynamic-neutral Japanese: $F(2,147) = 3.27$, $p = 0.041$, $\eta_p^2 = 0.043$).

The ROI x Age interaction indicates that face scanning differed between at least two age groups. A one-way ANOVA with factor Age was conducted at each level of ROI to identify the face regions which gave rise to these age differences (findings are reported in Table 13.46 in the Appendix F). Consistent age differences in looking time were found for the eyes of static faces (White-British: $F(2,147) = 4.89$, $p = 0.009$, $\eta_p^2 = 0.062$; Japanese: $F(2,147) = 5.97$, $p = 0.003$, $\eta_p^2 = 0.075$), for the mouth of static faces (White-British: $F(2,147) = 7.36$, $p < 0.001$, $\eta_p^2 = 0.091$; Japanese: $F(2,147) = 8.58$, $p < 0.001$, $\eta_p^2 = 0.104$), for the eyes of dynamic-neutral faces (White-British: $F(2,147) = 4.08$, $p = 0.019$, $\eta_p^2 = 0.053$; Japanese: $F(2,147) = 7.28$, $p < 0.001$, $\eta_p^2 = 0.090$), for the nose of dynamic-neutral faces (White-British: $F(2,147) = 4.56$, $p = 0.012$, $\eta_p^2 = 0.058$; Japanese: $F(2,147) = 4.93$, $p = 0.008$, $\eta_p^2 = 0.063$), for the mouth of dynamic-neutral faces (White-British: $F(2,147) = 7.12$, $p = 0.001$, $\eta_p^2 = 0.088$; Japanese: $F(2,147) = 8.99$, $p < 0.001$, $\eta_p^2 = 0.109$), for the bridge of dynamic-social faces (White-British: $F(2,147) = 3.14$, $p = 0.046$, $\eta_p^2 = 0.041$; Japanese: $F(2,147) = 6.42$, $p = 0.002$, $\eta_p^2 = 0.080$), and for the mouth of dynamic-social faces (White-British: $F(2,147) = 10.64$, $p < 0.001$, $\eta_p^2 = 0.126$; Japanese: $F(2,147) = 8.55$, $p < 0.001$, $\eta_p^2 = 0.104$).

Less consistent patterns were also found for the eye region (Japanese: $F(2,147) = 5.80$, $p = 0.004$, $\eta_p^2 = 0.073$) and the nose of dynamic-social faces (White-British: $F(2,147) = 6.49$, $p = 0.002$, $\eta_p^2 = 0.081$), indicating age differences only for stimuli of Japanese ethnicity and White-British ethnicity, respectively. To understand at which age these scanning differences occurred, each significant effect was followed up in a final step using post-hoc comparisons at a Bonferroni-adjusted alpha-level of 0.0167 (=0.05/3).

Results are reported in Table 13.47 and Table 13.48 in Appendix F. For static and dynamic-neutral faces, eye looking significantly decreased from 10 months to 15-17 months (all $p \leq 0.012$). No additional decrease was found from 15-17 months to adulthood (for statistical values see Table 13.47 and Table 13.48 in Appendix F). This was also found for dynamic-social faces of Japanese ethnicity ($p = 0.001$), but not White-British ethnicity, although a similar trend was observed.

For all stimulus conditions, a significant increase in mouth looking was observed from 10 months to 15-17 months (all $p \leq 0.003$). For the static and dynamic-social condition, this pattern significantly reversed as indicated by a decrease in mouth looking between 15-17 months and adulthood (all $p \leq 0.008$), with the exception for static White-British faces although a similar pattern was observed ($p = 0.047$). Mouth looking in adults was then no longer significantly different from the 10-month group. For the dynamic-neutral condition, mouth looking remained stable in adults after an increase in mouth looking from 10 to 15-17 months.

ROI effects that were less consistent across stimulus or face ethnicity conditions included increased nose looking of dynamic-neutral faces in adults compared to 15- to 17-month-olds (all $p \leq 0.002$), who did not differ from the 10-month-group (for statistical values see Table 13.47 and Table 13.48 in Appendix F). A similar pattern was observed for dynamic-social faces, but only for White-British face stimuli ($p = 0.002$). For both Japanese and White-British faces, the bridge of the face in the dynamic-social condition was scanned significantly more by adults than 15- to 17-month-olds (all $p \leq 0.011$), but no other age effects were found. This suggests a decreasing trend in bridge looking from 10 to 15-17 months, which increases again thereafter.

In sum, significant age differences in face scanning were identified for dwell time data, with findings consistently pointing to a peak in eye looking at 10 months, which decreased for 15- to 17-month-olds and adults. The decrease in eye looking was also accompanied by an increase in mouth looking which peaked at 15-17 months. Adults also scanned facial regions in a more distributed manner than infants, as indicated by increased nose or bridge scanning, although stimulus condition and ethnicity of the face stimulus modulated this pattern.

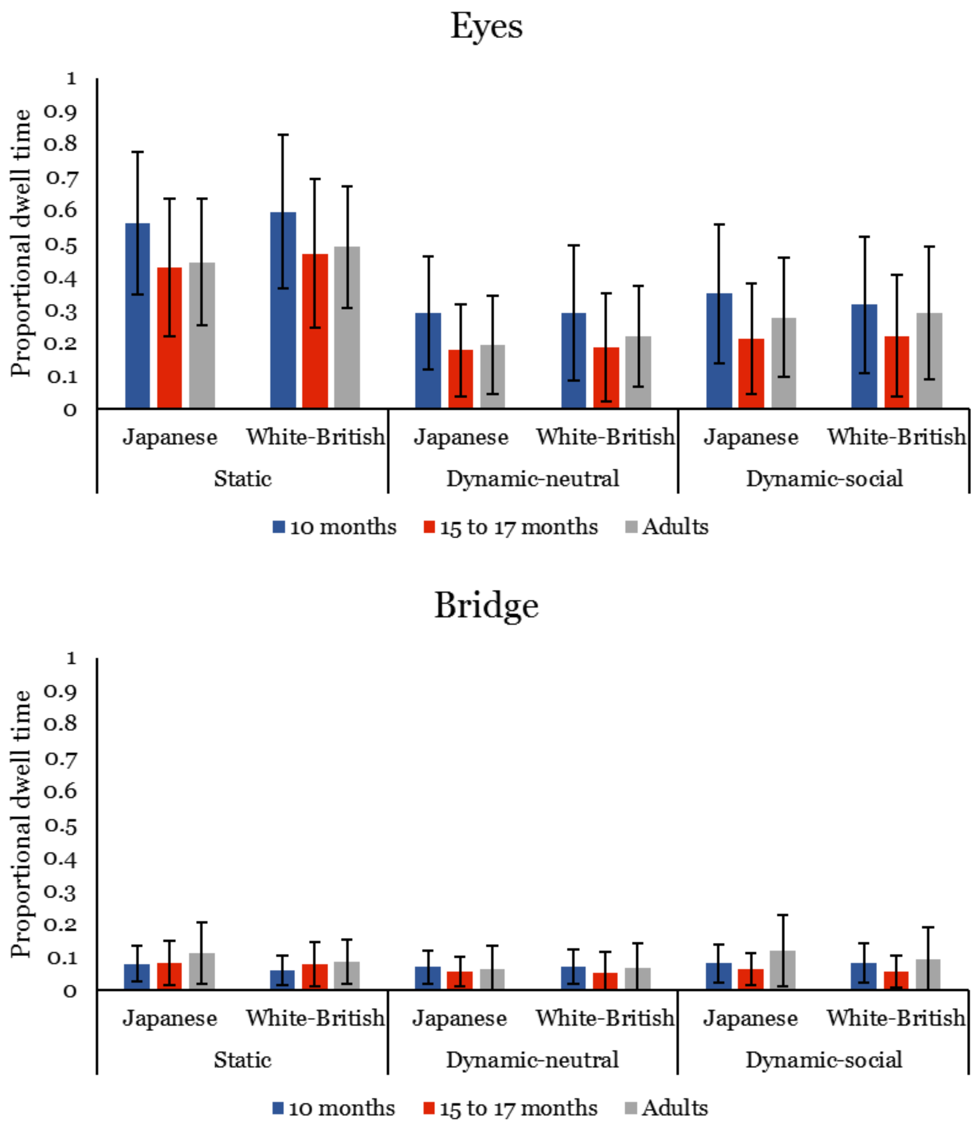


Figure 5.12. Proportional dwell times for the eyes and bridge region, separately for stimulus condition, face ethnicity, and age group. Error bars represent +/- 1 SD.

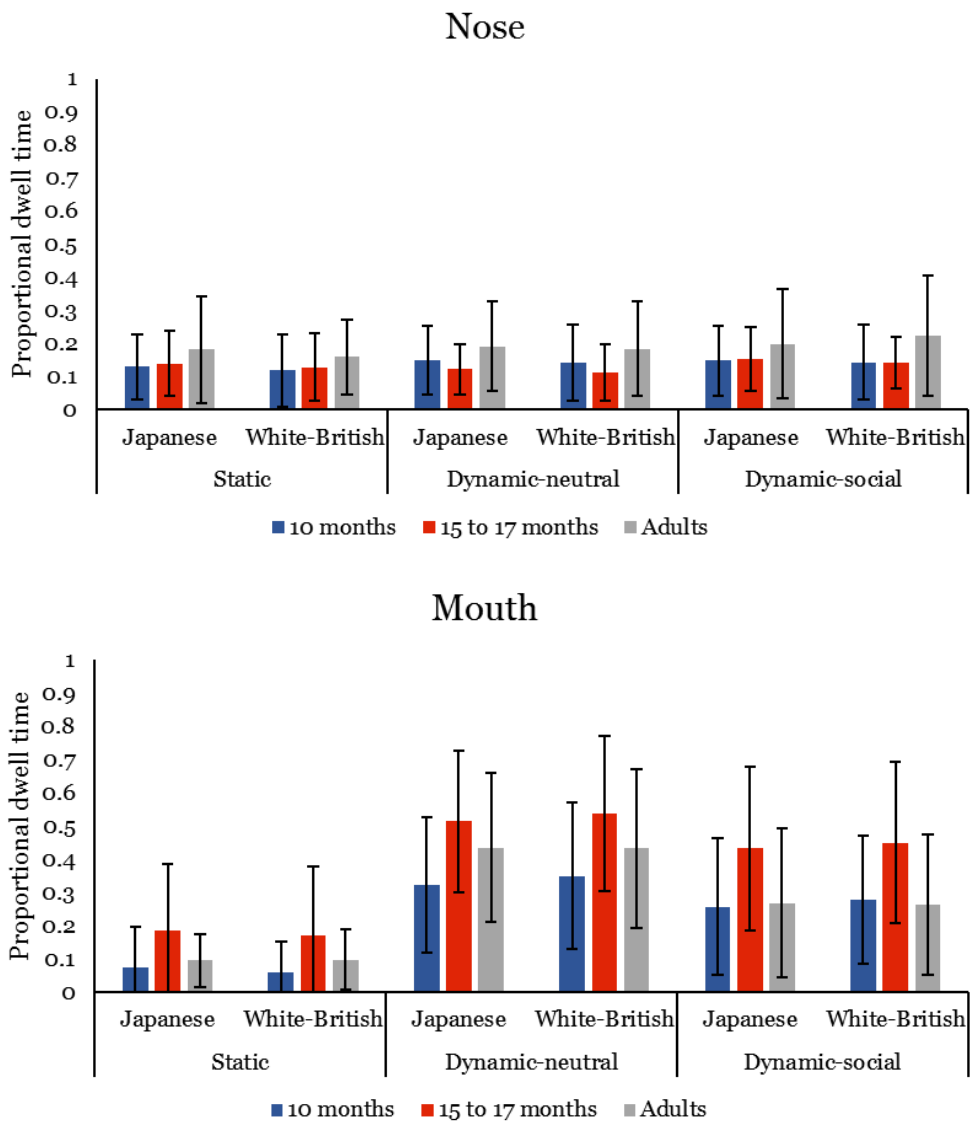


Figure 5.13. Proportional dwell times for the nose and mouth region, separately for stimulus condition, face ethnicity, and age group. Error bars represent +/- 1 SD.

5.4.1.3.2. *Fixation time analysis*

The descriptive data is visualised in Figure 5.14 on page 191. Findings for the ROI x Age analysis are provided in Table 13.49 to Table 13.51 in Appendix F. Significant effects of the two-way ANOVA were consistent with those of the dwell time analysis. Specifically, the ROI x Age interaction was significant across all stimulus conditions (static: $F(3.51,258.03) = 6.89$, $p < 0.001$, $\eta_p^2 = 0.086$; dynamic-neutral: $F(3.63,266.48) = 8.31$, $p < 0.001$, $\eta_p^2 = 0.102$; dynamic-social: $F(3.81,266.48) = 8.31$, $p < 0.001$, $\eta_p^2 = 0.102$).

The main effect of ROI was also significant (static: $F(1.76,258.03) = 247.98$, $p < 0.001$, $\eta_p^2 = 0.628$; dynamic-neutral: $F(1.81,266.48) = 129.83$, $p < 0.001$, $\eta_p^2 = 0.469$; dynamic-social: $F(1.91,280.24) = 55.48$, $p < 0.001$, $\eta_p^2 = 0.274$). As before, the Age effect was inconsistent and was only found for dynamic-social faces ($F(2,147) = 3.86$, $p = 0.023$, $\eta_p^2 = 0.050$).

As above (Section 5.4.1.3.1), a one-way ANOVA with factor Age was conducted at each level of ROI, and the same significant main effects as in the dwell time analysis were found (static faces – eyes: $F(2,147) = 8.10$, $p < 0.001$, $\eta_p^2 = 0.099$, mouth: $F(2,147) = 6.91$, $p = 0.001$, $\eta_p^2 = 0.086$; dynamic-neutral faces – eyes: $F(2,147) = 7.80$, $p = 0.001$, $\eta_p^2 = 0.096$, nose: $F(2,147) = 6.45$, $p = 0.002$, $\eta_p^2 = 0.081$, mouth: $F(2,147) = 9.50$, $p < 0.001$, $\eta_p^2 = 0.115$; dynamic-social faces – eyes: $F(2,147) = 3.56$, $p = 0.031$, $\eta_p^2 = 0.046$, bridge: $F(2,147) = 3.26$, $p = 0.041$, $\eta_p^2 = 0.042$, mouth: $F(2,147) = 9.24$, $p < 0.001$, $\eta_p^2 = 0.112$; see Table 13.52 in Appendix F for all statistical values). However, the present fixation data analysis also revealed that the significant age difference in nose scanning was not only present for the dynamic-neutral condition, but also in the static ($F(2,147) = 3.28$, $p = 0.040$, $\eta^2 = 0.043$) and dynamic-social condition ($F(2,147) = 4.42$, $p = 0.014$, $\eta^2 = 0.057$).

Post-hoc comparisons (see Table 13.53 in Appendix F) showed that adults exhibited increased nose looking compared to 10-month-olds (all $p \leq 0.011$); however, this effect did not survive multiple comparison correction for the static condition, although similar trends were observed ($p = 0.025$). Other effects converged with those of the dwell time analysis (Section 5.4.1.3.1), demonstrating a peak in eye looking at 10 months (all $p \leq 0.013$). This decreased thereafter with a simultaneous increase in mouth looking, peaking at 15-17 months (all $p \leq 0.014$). Adults showed greater gaze distribution across the face, as indicated by increased nose ($p = 0.001$) or bridge scanning ($p = 0.010$), although these effects were less consistent across stimulus conditions.

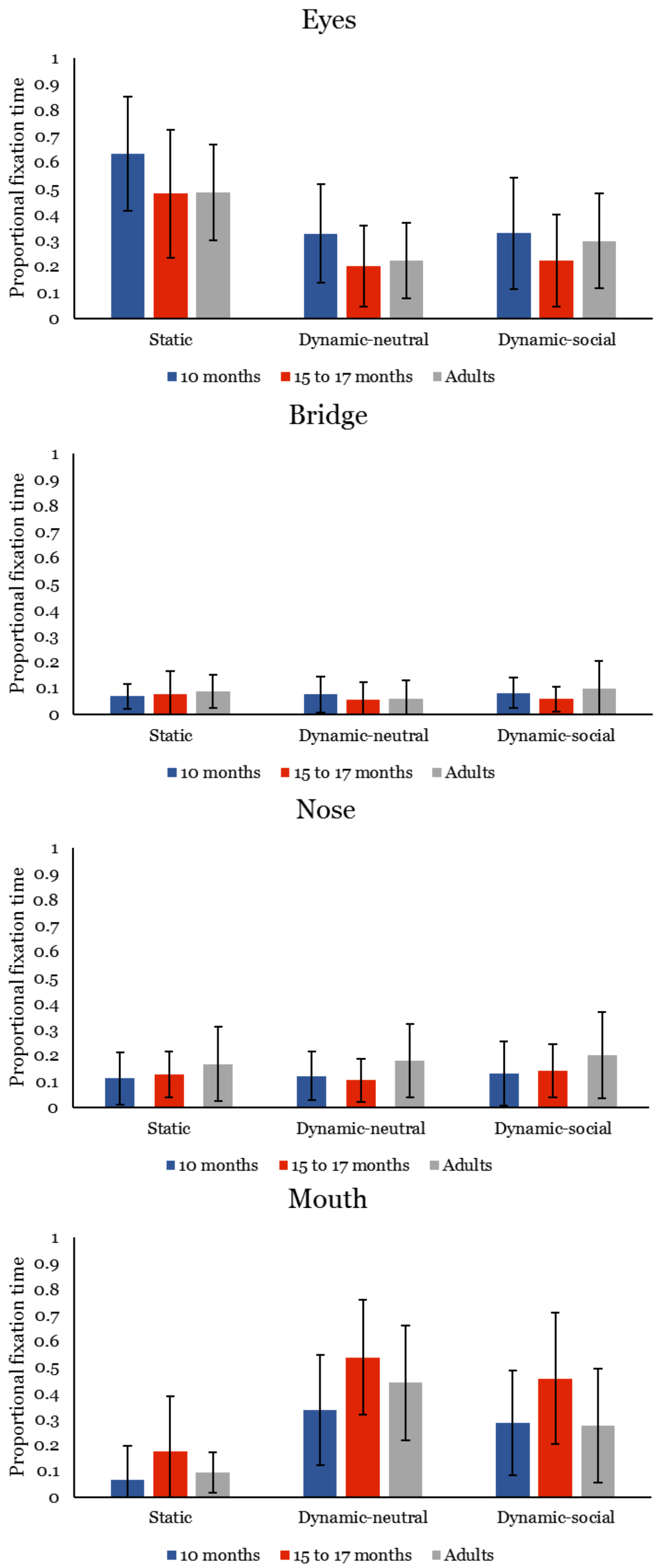


Figure 5.14. Proportional fixation times for each region-of-interest, stimulus condition, and age group. Error bars represent +/- 1 SD.

5.4.1.3.3. *Age-related development of cultural differences in face scanning: Examining the ROI x Age x Group interaction*

Sections 5.4.1.2 and 5.4.1.3 examined cultural and age differences in face scanning, respectively. As reported in Section 5.4.1.1.1, a significant ROI x Age x Group interaction was also found for dwell time data. The interaction was only significant for face stimuli of White-British ethnicity, implying that cultural differences in face scanning did not significantly change across age groups for face stimuli of Japanese ethnicity. To address the prediction that cultural differences become more evident with increasing age, Age x Group ANOVAs were conducted for White-British face stimuli at each level of ROI (collapsed across Stimulus Condition). Given that the primary research question was concerned with age-related differences in face scanning between cultures, Age and Group were retained within the same model.

Detailed findings are presented from Table 13.54 to Table 13.57 in Appendix F. A significant Age x Group interaction was only found for the eye region ($F(2,144) = 4.22$, $p = 0.017$, $\eta^2 = 0.055$); however, post-hoc comparisons revealed that none of the age contrasts survived a Bonferroni-corrected alpha-level of 0.0167 (=0.05/3; 10 months: $t(46) = 2.09$, $p = 0.042$; 15-17 months: $t(39) = 2.23$, $p = 0.032$; adults: $t(59) = 1.23$, $p = 0.225$). Altogether, the findings did not support the prediction that cultural differences would become increasingly more evident with age.

5.4.1.4. ***Examining stimulus effects***

The significant Stimulus Condition x ROI x Group and Stimulus Condition x ROI x Age interactions were re-analysed by conducting two-way ANOVAs for each level of cultural or age group. This allowed direct comparison of the different stimulus conditions.

5.4.1.4.1. *Dwell time analysis*

Findings of the two-way ANOVAs (ROI x Stimulus Condition) are presented from Table 13.58 to Table 13.61 in Appendix F, separately for British and Japanese participants. The

ROI x Stimulus Condition interaction was significant for each cultural group and across face ethnicities (Japanese group, Japanese ethnicity: $F(3.98,262.88) = 80.65$, $p < 0.001$, $\eta^2 = 0.550$; British group, Japanese ethnicity: $F(4.07,333.73) = 88.84$, $p < 0.001$, $\eta^2 = 0.520$; Japanese group, White-British ethnicity: $F(3.99,262.26) = 73.25$, $p < 0.001$, $\eta^2 = 0.526$; British group, White-British ethnicity: $F(3.66,300.49) = 115.99$, $p < 0.001$, $\eta^2 = 0.586$).

A follow-up one-way ANOVA with factor Stimulus Condition (for statistical values see Table 13.62 in Appendix F) showed an effect for the eyes and mouth. Given that no stimulus-dependent effects were predicted for the nose and bridge, and given that effects were inconsistent for these regions, only the eye and mouth effects were examined further. Post-hoc pairwise comparisons at a Bonferroni-adjusted alpha-level of 0.0167 (=0.05/3; see Table 13.63 and Table 13.64 in Appendix F) revealed that both British and Japanese participants looked significantly more at the eyes and less at the mouth in the static compared to dynamic-neutral (all $p < 0.001$) and dynamic-social condition (all $p < 0.001$). In addition, both cultural groups looked more at the eyes and less at the mouth in the dynamic-social compared to dynamic-neutral condition (all $p \leq 0.001$), except for Japanese participants viewing White-British faces, for which the significant difference in eye looking did not survive Bonferroni correction, but a similar pattern was observed ($p = 0.037$).

Similar findings were obtained for the ROI x Stimulus Condition analysis for each age group (see Table 13.65 to Table 13.70 in Appendix F). Given significant ROI x Stimulus Condition interactions for each age group, and a one-way ANOVA revealing effects for the eyes and mouth (see Table 13.71 in Appendix F), post-hoc pairwise comparisons were conducted to examine stimulus effects (see Table 13.72 to Table 13.74 in Appendix F). All age groups looked more at the eyes and less at the mouth in the static compared to dynamic-neutral (all $p < 0.001$) and dynamic-social condition (all $p < 0.001$). Adults also looked significantly less at the eyes and more at the mouth in the dynamic-neutral compared to dynamic-social condition (all $p \leq 0.001$). The 15- to 17-

month-old group looked more at the mouth in the dynamic-neutral than dynamic-social condition (all $p \leq 0.008$), but no significant difference were found with respect to the eye region. Finally, the 10-month group looked less at the eyes of dynamic-neutral compared to dynamic-social Japanese faces ($p = 0.008$), and more at the mouth of dynamic-neutral compared to dynamic-social White-British faces ($p = 0.009$).

5.4.1.4.2. *Fixation time analysis*

The above analysis (Section 5.4.1.4.1) was repeated for the fixation data. The findings for each cultural group converged with the dwell time analysis, except that significant differences in eye looking of dynamic-social faces could not be found in British participants (in Appendix F, see Table 13.75 and Table 13.76 for the two-way ANOVA ROI x Stimulus Condition separately for British and Japanese participants; see Table 13.77 for the one-way ANOVA with factor Stimulus Condition; see Table 13.78 and Table 13.79 for post-hoc pairwise comparisons).

Similarly, the results of the fixation analysis converged with those of the dwell time analysis (in Appendix F, see Table 13.80 to Table 13.82 for the two-way ANOVA ROI x Stimulus Condition separately for each age group; see Table 13.83 for the one-way ANOVA with factor Stimulus Condition; see Table 13.84 to Table 13.86 for post-hoc pairwise comparisons). However, 10-month-old infants did not show significant differences in eye or mouth looking between the dynamic-neutral and dynamic-social condition.

5.4.2. Gap-overlap task

Table 5.4 on page 195 summarises descriptive statistics of saccadic latencies for the baseline, overlap, and gap conditions, separately for each age and cultural group. As expected, saccadic latencies were numerically longest for the overlap condition, and shortest in the gap condition. Saccadic latencies also decreased with older age groups.

Table 5.4. *Means and standard deviations (in brackets) of saccadic latencies (in milliseconds) for the baseline, overlap, and gap conditions.*

	Baseline		Overlap		Gap	
	UK	Japan	UK	Japan	UK	Japan
10 months	438.05 (55.42)	443.95 (57.59)	533.06 (135.88)	576.78 (89.33)	345.31 (37.98)	358.52 (47.62)
15-17 months	376.71 (39.43)	388.30 (38.94)	433.31 (66.46)	481.23 (82.17)	304.00 (45.03)	319.79 (39.84)
Adults	271.14 (65.51)	273.46 (33.16)	291.72 (74.97)	293.51 (66.39)	245.97 (61.68)	243.54 (40.84)

To calculate disengagement latencies, baseline measures were subtracted from the overlap condition (see Section 5.3.3.2.1). Findings are summarised below in Table 5.5.

Table 5.5. *Medians and interquartile ranges (in brackets) of disengagement latencies (in milliseconds) by group and age.*

	UK	Japan
10 months	82.59 (98.61)	133.45 (96.65)
15 to 17 months	52.90 (40.78)	87.58 (73.56)
Adults	8.59 (35.61)	1.22 (54.28)

A 2 (Group: British, Japanese) x 3 (Age: 10 months, 15 to 17 months, adults) between-subjects ANOVA was conducted on disengagement latencies, revealing a significant main effect of Age ($F(2,144) = 24.30$, $p < 0.001$, $\eta_p^2 = 0.252$) and Group ($F(1,144) = 4.37$, $p = 0.038$, $\eta_p^2 = 0.029$). The Group x Age interaction was not significant ($F(2,144) = 1.30$, $p = 0.277$). Given that disengagement latencies were not normally distributed, the Group effect was examined again using the non-parametric Mann Whitney U test, but could not be supported ($U = 2381.50$, $p = 0.131$). Although the Group effect did not reach significance, the descriptive statistics (see Table 5.5 on page 195) numerically suggest that latencies for Japanese infants were longer than those of British infants; in other words, the data indicated a tendency into the opposite direction of the preliminary prediction that Japanese infants would outperform their British counterparts. The Age effect (with three levels) was investigated more closely with post-hoc comparisons using the non-parametric Mann Whitney U test with an adjusted Bonferroni-corrected alpha-level of 0.0167 (=0.05/3). Disengagement latencies significantly decreased with age (10 versus 15-17 months: $U = 672.00$, $p = 0.010$, $r = 0.272$; 10 months versus adults: $U = 511.50.$, $p < 0.001$, $r = 0.557$; 15-17 months versus adults: $U = 516.00$, $p < 0.001$, $r = 0.496$).

5.4.3. Cognitive control task

Table 5.6 summarises descriptive statistics for the mean proportion of correct anticipatory looks during the pre- and post-switch phase. For this analysis, five 10-month-old British infants, three 10-month-old Japanese infants, and two 15-month-old British infants were excluded from analysis due to fussiness during data collection.

Table 5.6. *Means and standard deviations (in brackets) of proportion of correct anticipatory looks.*

	Pre-switch		Post-switch	
	UK	Japan	UK	Japan
10 months	0.48 (0.34)	0.64 (0.29)	0.37 (0.29)	0.57 (0.33)
15 to 17 months	0.54 (0.33)	0.58 (0.29)	0.54 (0.34)	0.52 (0.36)
Adults	0.66 (0.31)	0.67 (0.30)	0.79 (0.21)	0.74 (0.25)

A 2 (Group: British, Japanese) x 2 (Condition: pre-switch, post-switch) x 3 (Age: 10 months, 15 to 17 months, adults) mixed ANOVA was conducted on the proportion of correct anticipatory looks. A main effect of Age ($F(2,134) = 14.00$, $p < 0.001$, $\eta_p^2 = 0.173$) was found, suggesting superior anticipatory looking performance in older age groups (see Table 5.6 on page 196). The Age x Group interaction ($F(2,134) = 3.06$, $p = 0.049$, $\eta_p^2 = 0.044$) was also significant. No other significant main effects or interactions were found (Group: $F(1,134) = 2.33$, $p = 0.129$; Condition: $F(1,134) = 0.02$, $p = 0.897$; Condition x Age: $F(2,134) = 2.52$, $p = 0.084$; Condition x Group: $F(1,134) = 0.089$, $p = 0.765$; Condition x Age x Group: $F(2,134) = 0.21$, $p = 0.810$). To follow up the Age x Group interaction, three independent *t*-tests (separately for each level of Age) were conducted on the proportion of correct anticipatory looks, which were collapsed across the pre- and post-switch phase for this analysis. Japanese participants had a significantly higher proportion of correct anticipatory looks than the British group at 10 months ($t(38) = 3.13$, $p = 0.004$, $d = 1.002$), but not at 15-17 months ($t(37) = -0.10$, $p = 0.919$) or in adults ($t(59) = 0.37$, $p = 0.712$).

When following up the Age x Group interaction separately for each cultural group, no age-related changes in performance were found for the Japanese group ($F(2,61) = 2.90$, $p = 0.063$). In contrast, British participants showed age-related differences ($F(2,73) = 15.64$, $p < 0.001$, $\eta_p^2 = 0.300$), with an increase in performance from 10 months to adulthood ($t(50) = 5.84$, $p < 0.001$, $d = 1.572$) and from 15-17 months to adulthood ($t(53) = 3.21$, $p = 0.002$, $d = 0.801$), but not from 10 to 15-17 months ($t(43) = 2.31$, $p = 0.026$; Bonferroni-adjusted at alpha-level 0.0167).

5.4.4. Relationship between executive function and face scanning

A relationship between more advanced executive function and increasingly culture-specific face scanning was predicted. Given the lack of an Age x Group interaction for face scanning behaviours (see Section 5.4.1.3.3), however, no developmental changes in cultural differences could be observed such that this prediction could not be supported. The following will present an exploratory analysis that examined the possibility that higher executive function scores at an individual level could relate to more culture-specific face scanning. In particular, correlational analyses between proportional mouth scanning and disengagement latencies *or* cognitive control task performance were conducted at each age and for each cultural group. The analyses were performed separately for British and Japanese participants given that opposite effects were expected for each cultural group (see below for details). Since disengagement latencies significantly differed between age groups (see Section 5.4.2), correlational analyses were additionally conducted separately at each age. Similarly, age effects were found for the performance on the cognitive control task in the British group (see Section 5.4.3), and correlational analyses were therefore performed separately at each age. Mouth scanning was selected as an index given that cultural differences in mouth looking have been highly consistent across studies.

Based on the findings demonstrating greater mouth scanning in the British compared to the Japanese group (see Section 5.4.1.2), a negative relationship between disengagement latencies and proportional mouth looking was expected for British participants. In other words, more advanced performance on the gap-overlap task (shorter disengagement latencies) would be associated with higher proportional mouth looking, and vice versa. For Japanese participants, a reverse pattern was expected. Specifically, shorter disengagement latencies would be associated with lower proportional mouth looking. In addition, a positive relationship between proportional mouth looking and performance on the cognitive control task was expected for the British group, i.e. more advanced performance (higher proportion of correct anticipatory looks) would be associated with increased mouth looking. Conversely, a negative relationship was predicted for the Japanese group, with more advanced performance on the cognitive control task being associated with lower proportional mouth scanning. For each correlational analysis, the alpha-level was adjusted using Bonferroni correction. Specifically, the alpha-level for the analysis using dwell time measures was set to 0.008 (=0.05/6) to consider six levels for mouth looking including three stimulus conditions (static, dynamic-neutral, dynamic-social) and two face ethnicity levels (Japanese and

White-British). When using fixation time measures, the alpha-level was set to 0.017 (=0.05/3) to consider mouth looking in each stimulus condition (static, dynamic-neutral, dynamic-social).

Altogether, the findings indicated no significant correlations (all $p > 0.05$) between disengagement latencies and proportional mouth looking using dwell time measures (see Table 14.1 in Appendix G for detailed statistical values) or fixation time measures (see Table 14.2 in Appendix G). Furthermore, the correlation coefficients for proportional mouth looking and performance on the cognitive control task indicated a numerical tendency into the opposite direction of the predicted effect. Specifically, correlation coefficients were consistently positive for Japanese infants (see Table 14.3 and Table 14.4 in Appendix G), which indicated a relationship between more advanced cognitive control task performance and increased mouth looking. For British infants, correlation coefficients were negative (see Table 14.3 and Table 14.4 in Appendix G), which suggested a relationship between better performance and decreased mouth looking. However, none of the correlations survived the Bonferroni-adjusted alpha-level when using dwell time (see Table 14.3 in Appendix G) or fixation time measures (see Table 14.4 in Appendix G).

5.5. Discussion

This study aimed to map the developmental trajectory of face scanning strategies in British and Japanese 10- and 15- to 17-month-olds as well as adults. To gain a comprehensive understanding of how stimulus characteristics can modulate scanning strategies, both White-British and Japanese faces were presented in static, dynamic-neutral, and dynamic-social conditions. The relationship between the development of more advanced executive function and the emergence of cultural differences in face scanning was also examined. The following will first discuss the findings on cultural and age differences in face scanning, before considering the role of executive function.

5.5.1. Cultural differences in face scanning

It was predicted that British participants would exhibit greater triangular scanning of the eyes and mouth when viewing neutral facial expressions in the static and dynamic-neutral condition, and greater mouth looking in the dynamic-social condition. Across all stimulus conditions and face ethnicities, the British group showed more mouth looking than Japanese participants. This cultural difference in mouth scanning therefore replicated previous studies using static faces with neutral expressions (e.g., Blais et al., 2008), static and dynamic emotionally expressive stimuli (Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013), and also converged with the findings from the live dyadic interaction study in Chapter 4. This therefore points to a highly consistent marker for cultural differences between Western Caucasian and East Asian individuals. Contrary to predictions, however, British participants did not exhibit greater eye looking in the static and dynamic-neutral condition, so that the triangular scanning pattern could not be replicated. Similarly, Japanese participants were expected to engage in greater eye looking in the dynamic-social condition and in increased central face scanning for faces with neutral expressions, but this prediction could not be fully supported. Instead, central face scanning patterns were found for both the dynamic-neutral and dynamic-social condition. A tendency for increased eye scanning was observed across all conditions, but significant group differences could not be found. The current predictions on the manifestation of cultural differences in face scanning were based on the possibility that the diverging scanning patterns reported in previous studies resulted from using neutral versus emotionally expressive face stimuli (e.g., Blais et al., 2008 versus Jack et al., 2009). The present findings, however, cannot support this as a single explanation, suggesting that other factors also modulated scanning behaviour.

It is possible that the experimental tasks in previous studies influenced eye movement behaviour in a top-down manner (Yarbus, 1967). The triangular scanning pattern (Western Caucasians) and central bias (East Asians) were observed when participants were asked to learn and recognise face identities (Blais et al., 2008), requiring visual attention to diagnostic facial features. Increased eye scanning in East Asian participants was found when asked to categorise faces by emotional expression (Jack et al., 2009). Given that group differences in eye scanning could not be found in the present free-viewing paradigm, an increased focus on the eye region could reflect a beneficial strategy for British individuals during face recognition tasks, and for Japanese individuals during emotion categorisation. However, Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al. (2013) also adopted a free-viewing paradigm and found increased eye (and not nose) scanning for emotionally expressive faces in Japanese participants, suggesting that task differences cannot fully account for the observed scanning differences between current and previous findings.

An additional methodological factor that may have influenced eye movement behaviour concerns the stimulus differences in mouth movements. Whereas Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al. (2013) presented participants with dynamic faces of actors producing a gaze shift and a smile (without teeth), the current dynamic-neutral and dynamic-social stimuli consisted of actors speaking syllables that were muted and accompanied by background music (see Section 5.3.3.1.1). It has been shown that increasing noise levels during speech result in greater attention to the mouth region (Vatikiotis-Bateson et al., 1998), likely as a compensatory strategy for language understanding. Further supportive evidence comes from Thompson and Malloy (2004) who found that older adults (with a mean age of 71.5 years) scanned the mouth region significantly more than younger adults (with a mean age of 23.4 years) at the expense of the eye region. A proposed explanation for this change in adulthood involved an increased reliance on the mouth as a source for visible speech that may result from, e.g., hearing loss and slower information processing in older adults (Thompson, 1995). It is therefore possible that the unintelligible speech in the dynamic-neutral and dynamic-social condition differentially modulated scanning behaviour in the two cultural groups. British participants may have engaged in greater mouth looking to decode unintelligible speech, whereas Japanese participants could have increasingly focused on the nose region to extract visual information from the mouth parafoveally. This account would be consistent with findings based on the Spotlight technique (with static images), showing that both Western Caucasians and East Asians fixate the eyes and mouth when visual information from the face was highly constrained (2° or 5°), though a shift toward a central fixation bias was observed only for East Asians when both the eyes and mouth were visible (at 8°; Caldara et al., 2010). Given that the dynamic-neutral and dynamic-social faces in this study were not only characterised by unintelligible speech but also by increased low-level motion, future studies would need to examine this further.

Altogether, cultural differences were observed across all stimuli, though the manifestation of viewing behaviour differed between stimulus conditions, emphasising also the importance of examining face scanning in different contexts. Given that empirical findings on cultural differences in face scanning are limited and existing studies greatly differ in methodology and analysis, further studies will be required to draw firm conclusions about the driving factors that modulate scanning strategies.

5.5.2. Age differences in face scanning

In addition to cultural effects, age-related influences on face scanning were found. Ten-month-old infants showed increased scanning of the eyes compared to the 15- to 17-month group and compared to adults, with the latter two age groups not showing significant differences in eye looking. Conversely, an increase in mouth scanning was observed after 10 months and peaked in the 15- to 17-month-group. Such a change from eye to mouth looking toward the end of the first year of life has also been observed in previous studies, especially for talking faces (e.g., Frank et al., 2012; Król, 2018). This pattern could be reflective of an adaptive mechanism for the learning requirements at each age. Infants may benefit from looking at the eyes for social learning and early non-verbal communication (Csibra & Gergely, 2006; Kleinke, 1986). For instance, eye contact can allow infants to engage in subsequent gaze following and joint attention (Scaife & Bruner, 1975; Senju & Csibra, 2008). By 15 to 17 months of age, infants have typically entered the word acquisition stage (Oller, 2000), and an increased focus on the moving mouth may provide a source for language learning (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2018). This is consistent with findings demonstrating increased mouth scanning in bilingual compared to monolingual infants (Pons, Bosch, & Lewkowicz, 2015), and in infants who were presented with faces speaking a non-native compared to a native language (Lewkowicz & Hansen-Tift, 2012). In addition, an association between amount of mouth scanning and expressive language skills has been found in infants (Tenenbaum, Sobel, Sheinkopf, Malle, & Morgan, 2015; Tsang, Atagi, & Johnson, 2018), supporting the idea that a looking bias toward the mouth may reflect an adaptive mechanism for language learning.

Less consistent findings were found for the bridge and nose, with adults typically scanning these regions more than 15- to 17-month-olds in the dynamic-neutral and dynamic-social condition. Given that adults tended to show an intermediary level of eye and mouth looking relative to the two infant groups, it is possible that adults distributed their looking behaviour more flexibly across the face to extract both social and language cues. Greater face exploration may therefore have elicited an increase in nose and bridge scanning, though not manifested as a consistent effect across conditions.

5.5.3. Age-related changes for cultural differences in face scanning

The current study was the first to investigate the development of face scanning by contrasting two cultural groups at different ages during infancy. Previous infant studies only examined a single age group (Geangu et al., 2016) or a single cultural group (Liu et al., 2011; Wheeler et al., 2011; Xiao et al., 2013), and it remained unclear how developmental changes in face scanning manifest. The present findings point to *independent* effects of culture and age on face scanning, suggesting that differences in face scanning were largely established by 10 months. The high within-group variance observed in the current study may have also masked between-group variance, again pointing to the possibility that cultural differences may be small in nature.

Nevertheless, given that studies have highlighted a role of postnatal social experience in shaping face scanning behaviour (Caldara, Richoz, Liu, & Lao, 2016; Kelly, Jack, et al., 2011), it is likely that cultural differences largely emerged prior to 10 months of age, consistent with findings from Geangu et al. (2016) who revealed cultural differences in the scanning of static, emotionally expressive faces at 7 months.

5.5.4. Role of face ethnicity and stimulus condition

Given that several infant and adult studies found the ethnicity of face stimuli to modulate scanning strategies (e.g., Fu et al., 2012; Liu et al., 2011; Wheeler et al., 2011; Xiao et al., 2013) while others did not replicate this pattern (e.g., Blais et al., 2008; Geangu et al., 2016; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013), this study included both White-British and Japanese faces. Interestingly, an effect of face ethnicity for cultural or age differences was only observed in the analysis using dwell time data but could not be replicated in the fixation time analysis. This effect of face ethnicity manifested in an inconsistent manner across the different experimental conditions; for instance, the British group engaged in more mouth looking than Japanese participants across all conditions except for the dynamic-neutral Japanese faces, which did not survive multiple comparison correction. Several explanations could account for this difference in face ethnicity effect. First, face ethnicity effects were small and may therefore be statistically difficult to detect, such that differences in analysis approaches could produce inconsistent findings. Secondly, unlike fixations, dwell time measures additionally included all data points during which gaze was not stable (e.g., saccades). Consequently, more eye movement data was available for the spatial pooling of the regions-of-interest approach, making it potentially more statistically sensitive to detect scanning differences. Thirdly, fixations in the current study were coded in a semi-automatic fashion, whereas the dwell measures were obtained in a fully automatic manner (see Section 5.3.4). In light of the good to excellent inter-coder reliability for fixations (see Section 5.3.4), it is possible that the dwell measures additionally included more noisy data, which may have affected study findings. Altogether, the inconsistent face ethnicity effect in the present study also converges with findings from Arizpe et al. (2016) who demonstrated that discrepancies in the manifestation of cultural differences can be in part explained by differences in the statistical approach and dependent measures used for analysis. The findings of the current study therefore emphasise the necessity to consider such methodological differences in the interpretation of cultural (or age) differences in face scanning.

This study also examined the effect of stimulus characteristics on face scanning and found consistent age- but not culture-related differences. Across all ages and cultures, participants looked significantly more at the eyes and less at the mouth in the static compared to dynamic-neutral and dynamic-social condition. The increased focus on the mouth may have resulted from greater low-level motion in this region or a compensatory strategy to decode language (Vatikiotis-Bateson et al., 1998), coming at the expense of eye scanning. Both British and Japanese participants looked more at the eyes and less at the mouth in the dynamic-social than dynamic-neutral condition, although the eye effect was less consistent across face ethnicities and dependent measures. This could have resulted from the increased low-level salience in the eye region of faces in the dynamic-social condition (e.g., raised eyebrows), or a greater source of social cues in this area. The same tendency was also observed for adults and infants when examining age-related differences, although significant differences for both the eyes and mouth were only found for adults. Whereas the 15- to 17-month-olds also looked more at the mouth in the dynamic-neutral versus dynamic-social condition, no significant differences were found for the eye region. The 10-month group exhibited less consistent findings, with the fixation analysis suggesting that no significant differences between the dynamic-neutral and dynamic-social condition existed for the eye or mouth region. Overall, the results suggest that a moving mouth may be a driving factor in face scanning for all age groups. In the presence of increased salience or social cues in the eye region, a trade-off may occur in favour of eye scanning. However, this effect was not as pronounced for infants, who continue to show a preference for the eye scanning at 10 months and mouth scanning at 15 to 17 months.

5.5.5. Executive function and its role in the development of face scanning

It was also expected that performance on executive function measures would become increasingly more advanced with age, and that Japanese participants would additionally outperform their British counterparts. Although disengagement latencies became shorter with age – replicating previous studies on developmental changes in visual disengagement abilities (Hood & Atkinson, 1993) – no significant cultural differences were observed. Additionally, descriptive values indicated that disengagement latencies were higher in Japanese compared to British infants, meaning that a numerical trend was found into the opposite direction of the predicted effect (that the Japanese group would outperform their British counterparts). For the cognitive control task, Japanese participants showed higher performance scores only at 10 months relative to the British group. No developmental changes were observed in task performance for the Japanese group, while a significant improvement was found from infancy to adulthood in British participants. This could indicate that British infants developed the capacity to perform the cognitive control task more slowly than the Japanese group. However, this explanation only applied to infants at 10 months of age; by 15 to 17 months, cultural differences could no longer be observed, with British infants reaching similar performance levels as their Japanese counterparts. The more established gap-overlap paradigm therefore revealed age-related increases in task performance for both groups, whereas the cognitive control task could not consistently capture this effect across cultures, questioning its role as a general executive function measure across wide age ranges. Indeed, the cognitive control task has not previously been adopted for participants beyond the first year of life (cf., Ballieux et al., 2016; Kovacs & Mehler, 2009; Wass et al., 2011; Wass & Smith, 2014). Furthermore, the performance on the pre- versus post-switch phases did not significantly differ, such that the use of average measures for the cognitive control task could not distinguish between performance in rule learning (pre-switch) and rule switching (post-switch). It is possible that more trials would be needed to establish an effect in performance for both cultural groups; however, this would give rise to practical challenges as longer task durations may increase fussiness or boredom in infants, and therefore lead to significant drop-out rates. Although a more continuous measure could be derived, e.g., the latency to saccade toward the correct side, future studies should ideally also adopt different executive function measures, such as the freeze frame task (Holmboe et al., 2018; Holmboe, Pasco Fearon, Csibra, Tucker, & Johnson, 2008).

Furthermore, several explanations could account for the absence of significant cultural differences in disengagement abilities. First, it is possible that cultural differences cannot yet be observed in infancy; no cross-cultural infant studies on executive function currently exist. However, differences in disengagement latencies were also not observed in adults. An additional possibility concerns the nature of the executive function measures. Specifically, previous studies demonstrating underlying cultural differences included measures such as the Dimensional Change Card Sort (DCCS) task, gift delay task, or Day-Night Stroop task (e.g., Imada et al., 2013; Sabbagh et al., 2006; see Moriguchi, 2014). However, given that these measures required verbal instructions, performance was inherently confounded by potential cultural differences in the understanding and interpretation of these tasks. In East Asian cultures, for instance, a greater emphasis is placed on conformity, and self-control is valued and encouraged in educational settings (see Oh & Lewis, 2008). As such, tasks may not only tap onto executive function but also the ability or willingness to conform to societal expectations in East Asian (but not Western Caucasian) participants. The gap-overlap paradigm used in this study was free from verbal instructions and cultural interpretations and found no differences in performance between cultural groups. However, it is also possible that only subcomponents of executive function give rise to cross-cultural differences. Studies (including the current experiment) typically employed a narrow range of tasks, but general underlying processes cannot be assumed based on limited observations. Future studies would need to employ a wide range of tasks to better understand the nature of cultural modulations on executive function.

It was also predicted that increasingly advanced executive function would emerge concurrently with more culture-specific face scanning. However, this prediction did not hold given that no age-related changes in cultural differences in face scanning were found. A relationship between mouth scanning and disengagement latencies or performance on the cognitive control task could also not be found in an exploratory analysis. Indeed, for the correlational analysis using the cognitive control task performance, a numerical tendency in the opposite direction of the predicted effect was found. In particular, more advanced task performance was associated with greater mouth scanning in Japanese participants, and decreased mouth looking in the British group. However, correlations were significant only at an uncorrected *p*-value and not after Bonferroni correction, and these effects were additionally inconsistent across stimulus and face ethnicity conditions. Nevertheless, it cannot necessarily be concluded that a relationship between executive function and culture-specific face scanning does not exist generally. Greater sample sizes would be required for correlational analyses, and additional measures should be used to capture executive function in a more comprehensive manner. This would then allow a more in-depth examination of the relationship between face scanning and executive function.

5.5.6. Conclusion

The current study revealed that eye movements during face viewing were modulated by various factors, including culture, age, and stimulus characteristics. Each of these factors independently influenced eye movement behaviour during face perception, but their relative contribution changed depending on the specific context. Altogether, individuals adopted different strategies for extracting visual information from faces, in line with their cultural background and stage in development. The lack of age-related changes for cultural differences in face scanning further points to similar developmental trajectories in face scanning in British and Japanese individuals. Future studies will be required to examine the role of executive function in face scanning in more depth. After having established that cultural differences exist by 10 months, the next chapter will present a study that aimed to describe face scanning in British and Japanese 10-month-olds during face-to-face interactions in order to examine whether cultural differences in infancy extend to naturalistic social settings.

Chapter 6

Cultural Differences in Naturalistic Face Scanning of 10-Month-Old Infants

6.1. Chapter Overview

Chapter 5 introduced a screen-based eye tracking paradigm, demonstrating that cultural differences in face scanning can be observed already at 10 months of age, with the British group exhibiting more mouth scanning across all stimulus conditions (static, dynamic-neutral, dynamic-social) and the Japanese group showing a central face bias for dynamic faces. The present study sought to examine whether scanning behaviour of British and Japanese 10-month-old infants also differed within more naturalistic face-to-face interactions. The following sections will briefly contextualise this work before going on to introducing the current study.

6.2. Introduction

As outlined in Chapter 1, cross-cultural studies on face scanning strategies of adults were largely restricted to screen-based studies. Chapters 3 and 4 presented dyadic face-to-face interaction paradigms to address this gap in the literature, finding that scanning behaviours also extend to naturalistic social contexts. Furthermore, since evidence on cultural differences in scanning strategies of infants was limited, Chapter 5 examined infants' eye movement behaviour when presented with a range of face stimuli on screen. Findings showed that cultural differences already manifested at 10 months of age, with British infants engaging in greater mouth scanning than their Japanese counterparts. However, whether this cultural difference can also be observed in infants who are engaged in naturalistic face-to-face interactions remains unclear. Although face-to-face interaction studies have been conducted with infant populations, these studies were typically restricted to parent-child interactions and adopted observational paradigms to code cultural differences in factors such as maternal responsiveness (Bornstein et al., 1992), infant-directed speech (Toda, Fogel, & Kawai, 1990), parental behaviours (Fogel et al., 1988), or the development of social behaviours such as infant smiling (Wörmann, Holodynski, Kärtner, & Keller, 2012). No study to date, however, has applied eye tracking techniques to examine infants' face scanning strategies when interacting with another person.

6.2.1. The current study

The current study established a face-to-face interaction paradigm to examine face scanning behaviour of British and Japanese 10-month-old infants. Infants' eye movements were recorded while they interacted with a local research assistant in either the UK or in Japan. Using the detection and tracker tool presented in Chapter 2, face regions were coded semi-automatically and divided further into upper and lower face regions. Given that there is no evidence to suggest cultural differences in the development of face orienting, no group differences were predicted. Based on screen-based eye tracking findings (see Chapter 5), it was expected that British infants would exhibit greater mouth (lower face) looking than the Japanese group. As discussed in Chapter 2, selecting regions-of-interest based on findings from screen-based eye tracking paradigms could limit novel insights into cultural differences in naturalistic face scanning since eye movements could manifest differently within live social interactions. For this reason, an additional Monte Carlo permutation test was conducted to investigate scanning behaviour in a data-driven and spatially sensitive manner (see Chapter 2 for details). As in the previous studies of this thesis (Chapter 4 and Chapter 5) both dwell time and fixation time were examined.

As mentioned in Chapter 1 (see Section 1.2.3), Geangu et al. (2016) suggested that visual experience with the caregiver's facial expression and associated cultural differences may provide an early source for diverging face scanning patterns between British and Japanese infants. To examine this further, the caregivers' eye movements were also recorded using head-mounted eye tracking techniques while they interacted with their child (after the infant interacted with the research assistant). However, the parent-child interaction study was characterised by a high drop-out rate in the Japanese group due to fussiness or insufficient data (69.57%, or 16 out of 23; compared to 20%, or

4 out of 20, for the British group). Specifically, the caregiver was required to be seated at a distance of 155 cm from the infant due to the experimental set-up (for more details see below in Section 6.3.2), and Japanese infants did not tolerate the physical separation at such distance. Due to the high drop-out rate, the data was not analysed. This point will be briefly revisited in Chapter 7.

6.3. Methods

6.3.1. Participants

Seventeen British 10-month-olds (7 females, 10 males) and 17 Japanese 10-month-olds (6 females, 11 males) took part in this study. British infants were tested at Birkbeck, University of London (UK), and Japanese infants were tested at Kyoto University (Japan). The two cultural groups did not significantly differ in age (British: $M = 305.94$ days, $SD = 10.18$ days, ranging from 291 to 326 days; Japanese: $M = 303.12$ days, $SD = 9.10$ days, ranging from 289 to 317 days; $t(32) = 0.85$, $p = 0.400$). In the UK, an additional three infants were tested but not included in the final analysis due to fussiness ($N = 2$) or not meeting ethnicity requirements ($N = 1$). In Japan, an additional six infants were tested but excluded from analysis due to fussiness ($N = 5$) or failure to track eyes and calibrate ($N = 1$). Twelve British and 14 Japanese infants tested in this study were also included in the analysis reported in Chapter 5. These infants were tested either in the same session after the screen-based study presented in Chapter 5, or as part of a separate visit if the family agreed to return.

As in Chapter 5 (see Section 5.3.1.1), demographic information was collected using a questionnaire (see Appendix E). For this study, infants tested in the UK were born and raised in the UK, were of White ethnicity, had never lived in a country outside the UK, and their caregiver communicated with them in English (except for two infants whose primary caregiver spoke in Italian, but English was also spoken at home). Infants tested in Japan were born and raised in Japan, were of Japanese ethnicity (except for one 10-month-old infant whose secondary caregiver was of White ethnicity), had never lived in a country outside Japan, and their caregiver communicated with them in Japanese.

Caregivers also rated their child's amount of contact with Western Caucasian and East Asian people on a scale from 1 (very little) to 7 (very extensive) to ensure that exposure to other ethnicities did not significantly differ between the two groups. No significant differences were found between British and Japanese infants with respect to the amount of contact with same-ethnicity individuals (British: $M = 6.65$, $SD = 0.61$; Japanese: $M = 6.94$, $SD = 0.24$; $t(32) = -1.86$; $p = 0.077$) and with other-ethnicity individuals (e.g., exposure to East Asian ethnicities for British infants; British: $M = 2.00$, $SD = 1.17$; Japanese: $M = 1.41$, $SD = 1.46$; $t(32) = 1.30$; $p = 0.205$).

All infants were full-term, had normal vision and hearing and no developmental condition as reported by their parents. The session lasted up to one hour including breaks and play time. Families were recruited via internal databases of the local university departments and volunteered their infants to take part. In line with standard departmental protocols, families in the UK were reimbursed travel expenses and received a T-Shirt and certificate of participation, and Japanese families were reimbursed ¥3000 for their time. The accompanying caregiver provided written informed consent prior to the study. The study was approved locally by the ethics committees of the Department of Psychological Sciences, Birkbeck, University of London, and the Department of Psychology, Kyoto University.

6.3.2. Apparatus

The experimental set-up is visualised in Figure 6.1 on page 215 and in Figure 6.2 on page 216. Infants sat in a high-chair and a small cushion was used to minimise movements. Eye movements were recorded using a Tobii TX300 eye tracker (Tobii Technology, Sweden) at a sampling rate of 120 Hz. The eye tracker was located on a small box in front of the high-chair. A Dell laptop was connected to the eye tracker and ran the experiment using *Tobii Studio* in standalone mode (Version 3.1.6; Tobii Technology, Sweden). A Logitech C920 webcam fitted with a normal lens was mounted onto a tripod and connected to the Dell laptop for scene and audio recordings. The webcam was located behind the high-chair and above the infant's head to capture her or his point of view (37° horizontally and 21° vertically). Scene recordings were captured at 20 frames per second and at 1280 x 720 resolution. The research assistant sat on a small stool behind the eye tracker at eye level with the infant. Two standing lamps were located behind the research assistant to ensure appropriate lighting levels for eye tracking. An external room camera recorded the session in case post-hoc data quality checks or behavioural coding were required. Finally, a foam board with five holes and a small squeaky toy were used for calibration (see Figure 6.1, and Figure 6.2 on page 216).

Figure 6.1. Eye tracking an infant participant for post-hoc data quality checks.

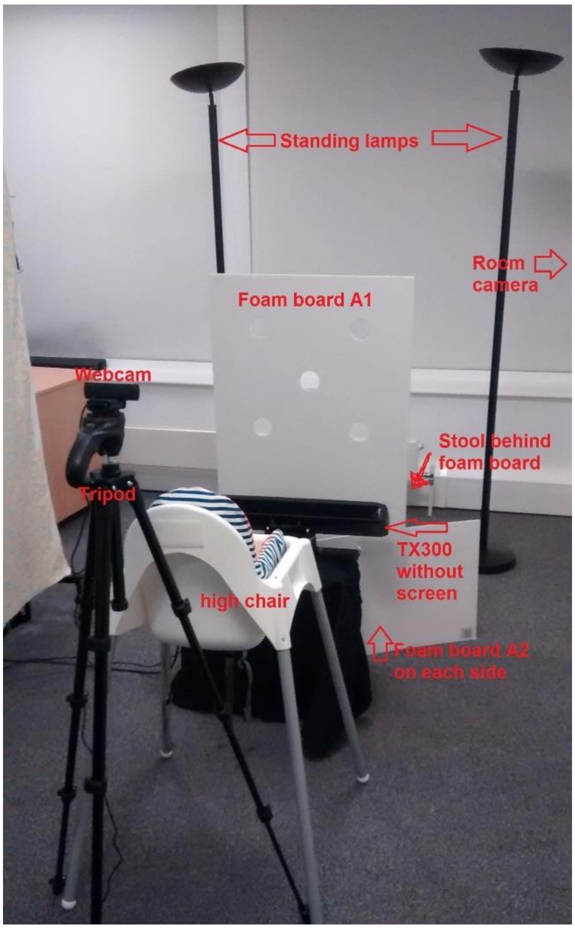


Figure 6.2. Experimental set-up. The foam boards were removed after successful calibration.

6.3.3. Procedure

Families were welcomed in the reception room where the experimenter explained the study and collected written informed consent. Caregivers were then asked to fill out a demographic questionnaire while infants were given time to familiarise themselves with the new environment. The caregiver and infant were then guided to the testing room where the experimenter started the room camera recording. Infants were placed in the high-chair and the research assistant sat on a stool at approximately 155-centimetre distance from the infant. A five-point calibration procedure was then conducted whereby the research assistant moved a small squeaky toy across five holes of the calibration foam board (see also Figure 6.2), and the experimenter controlled the calibration using *Tobii Studio*. Calibration performance was visualised in *Tobii Studio* and the procedure was repeated if necessary. After successful calibration, the foam board was removed, and the research assistant started interacting with the infant by singing nursery rhymes, playing peekaboo, or engaging in infant-directed speech. The research assistant in the UK was of White ethnicity and interacted with infants in English, whereas the research assistant in Japan was of Japanese ethnicity and spoke in Japanese. Both research assistants were female and in their early/mid-20s (see Figure 6.3). After two minutes of face-to-face interaction, the calibration toy was used again for post-hoc gaze data quality checks (see Section 6.3.4.1), and the recording was stopped.

Figure 6.3. The infant's point of view: research assistants interacting with infants in Japan (left) and in the UK (right).

6.3.4. Data pre-processing

6.3.4.1. *Data quality*

As in Chapter 4, the current study only included data sets with sufficient spatial accuracy to ensure that any observed cultural differences were not driven by systematic differences in data quality (cf., Blignaut & Wium, 2014). The gaze data was superimposed onto the scene recordings; if the gaze points fell onto the calibration toy after the face-to-face interaction, the data set was retained for analysis. For this study, no data set was excluded based on gaze offset.

6.3.4.2. *Coding of video start and end times for analysis*

The start and end times of periods used for analysis were coded manually. The start time was defined as the first frame (after successful calibration) in which the infant was not fussy (e.g., crying) and looked in the direction of the research assistant (i.e., when infants visually oriented to the research assistant; face looking was not required). The end time was coded to be 60 seconds after the start time. If an infant engaged in face-to-face interaction play for longer than 60 seconds, the scene recording for analysis was cropped to the 60-second mark to ensure all infants contributed a similar amount of data. If an infant turned around for 5 seconds or more such that no gaze data could be collected (e.g., turning toward the parent sitting behind her or him), the end time was counted as the instance just preceding the interruption. If an interruption was coded and the infant did not yet contribute 60 seconds of data, a second start and end time was used. Specifically, the second start time was coded as the instance after the infant looked back into the direction of the research assistant (as above, face looking was not required). If necessary, a third start and end time was used; this was required for data from three British and three Japanese infants. No further start/end times were used. Overall, each period of face-to-face interaction was only included in the analysis if at least 10 seconds of data were available (e.g., if a valid interaction period lasted only 2 seconds, this was not included in the analysis). Scene segments were then exported using *Tobii Studio*. Two British infants and five Japanese infants contributed less than 60 seconds of recording time; however, recording times did not differ between cultural groups (British: $M = 55.71$ seconds, $SD = 10.75$ seconds; Japanese: $M = 54.29$ seconds, $SD = 8.83$ seconds; $t(32) = -0.42$, $p = 0.730$).

6.3.4.3. *Eye tracking data pre-processing*

In addition to the scene segments, the corresponding gaze data was also exported. *Tobii Studio* allows the user to set individual parameters, and the following values were selected: for the dwell time analysis, the maximum interpolation latency was set to 150 milliseconds and the gaze data was smoothed based on a window size of 5 data points before and after the current sample. For the fixation time analysis, the Velocity-Threshold Identification (I-VT) filter was selected with a maximum interpolation latency of 150 milliseconds, a 5-sample smoothing space, a velocity threshold of 20°/second, a minimum fixation duration of 100 milliseconds, and adjacent fixations were merged if they were no more than 0.5° (space) and 75 milliseconds (time) apart. Given that both the infant participant and the research assistant produced head movements, it was possible that smooth pursuit movements would occasionally be observed. In other words, fixations on the research assistant's face would be characterised in a displacement over time if the face moved (cf., Smith & Mital, 2013). However, no parsing algorithm currently exists that can adequately extract fixations during dynamic scene viewing. Given that the proportion of affected gaze data would be very small (Smith & Mital, 2013), and given that no systematic differences between cultural groups should be present, no further pre-processing was conducted on the present gaze data.

6.3.4.4. *Regions-of-interest coding*

For all scene recordings, face regions were coded semi-automatically using the detection and tracker tool presented in Chapter 2. Face regions were further subdivided into upper and lower face areas. Face regions were not coded when the face was fully covered, such as during periods of peekaboo play.

6.4. Results

6.4.1. Regions-of-interest analysis

To compare data from British and Japanese infants, an independent *t*-test was conducted separately for proportional dwell time or fixation time on the face, and proportional dwell time or fixation time on the upper face. Unlike the data extraction procedures in Chapter 3 (see Section 3.4.1) and Chapter 4 (see Section 4.4.1), which computed face looking time proportional to the valid recording time (i.e., excluding periods of blinking or data loss), the present analysis computed dwell time and fixation time as a proportion to total recording time with a cut-off at 60 seconds (see Section 6.3.4.2). This was done given that the eye tracker in this study was fixed in position and therefore highly dependent on head and eye positions for data collection. In particular, while a head turn of a participant wearing a head-mounted eye tracker still typically allows for gaze data collection, such a head turn would result in data loss when using an eye tracker fixed in position. Given that the eye tracker in this study was positioned to optimise gaze data recording for face looking, most periods of valid recording would reflect face orienting behaviour. To ensure that face looking time was not inflated, the present study examined dwell or fixation time measures proportional to overall recording time. No significant group differences were found for amount of data loss proportional to the recording time included for analysis (British: $M = 13.66\%$, $SD = 10.92\%$; Japanese: $M = 16.47\%$, $SD = 9.71\%$; $t(32) = 0.79$, $p = 0.433$).

6.4.1.1. *Face looking*

6.4.1.1.1. *Dwell time analysis*

Dwell time measures showed that both cultural groups spent a similar amount of time looking at the face of the local research assistant (British: $M = 46.84\%$, $SD = 13.27\%$; Japanese: $M = 49.71\%$, $SD = 11.52\%$; see Figure 6.4 on page 221). An independent *t*-test revealed no significant group difference in face looking time ($t(32) = -0.67$, $p = 0.506$).

6.4.1.1.2. *Fixation time analysis*

Fixation time measures supported the findings from the dwell time analysis (Section 6.4.1.1.1), showing no significant difference in face looking time between British infants ($M = 41.75\%$, $SD = 14.26\%$) and Japanese infants ($M = 42.36\%$, $SD = 11.94\%$; $t(32) = -0.13$, $p = 0.894$; see Figure 6.4 on page 221).

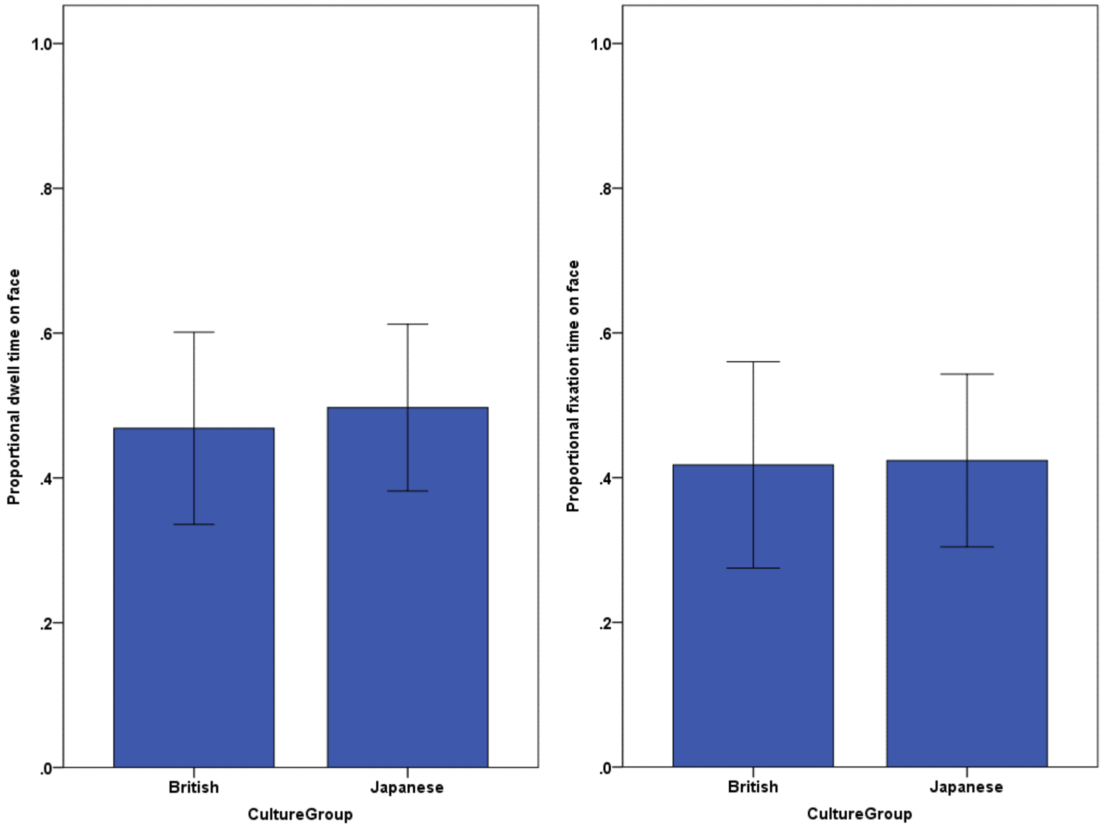


Figure 6.4. Proportional dwell time (left) and fixation time (right) spent looking at the face. Error bars represent +/-1 SD.

6.4.1.2. *Upper face looking*

6.4.1.2.1. *Dwell time analysis*

In contrast to predictions, dwell time measures did not reveal any significant cultural differences in upper face scanning (British: $M = 28.35\%$, $SD = 27.76\%$; Japanese: $M = 32.79\%$, $SD = 22.01\%$; $t(32) = -0.52$, $p = 0.608$; see Figure 6.5 on page 222).

6.4.1.2.2. *Fixation time analysis*

Converging with the dwell time analysis (Section 6.4.1.2.1), no significant group differences in upper face scanning could be found using fixation time measures (British: $M = 28.60\%$, $SD = 29.28\%$; Japanese: $M = 32.70\%$, $SD = 22.95\%$; $t(32) = -0.45$, $p = 0.653$; see Figure 6.5 on page 222).

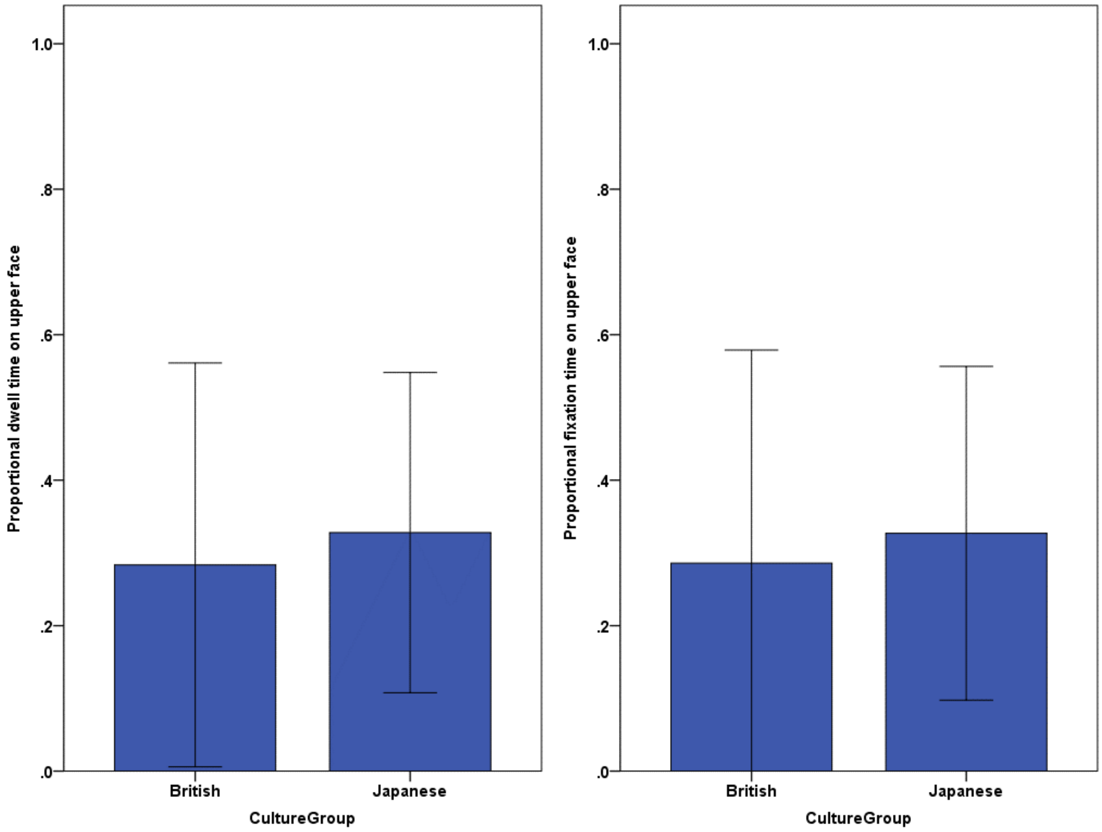


Figure 6.5. Proportional dwell time (left) and fixation time (right) spent looking at the upper face. Error bars represent +/-1 SD.

6.4.2. Monte Carlo permutation test

The findings from the screen-based study presented in Chapter 5 suggested that the Japanese group exhibited a central face bias for dynamic faces, which cannot be sufficiently captured using a regions-of-interest approach. To examine face scanning in a more spatially sensitive manner – particularly given the findings, data-driven Monte Carlo permutation tests (see Chapter 2) were conducted using the *CoSMoMVPA* toolbox (Oosterhof et al., 2016) and *FieldTrip* toolbox (Oostenveld et al., 2011). Analysis was performed separately for dwell time and fixation time measures. As in the previous chapters, parameters were set to the following:

- an uncorrected *p*-value threshold of 0.01, and
- 10,000 iterations.

Figure 6.6 illustrates group differences in gaze density for dwell time and fixation time measures, showing more looking at the mouth region in British infants, and central and lower face scanning in Japanese infants.

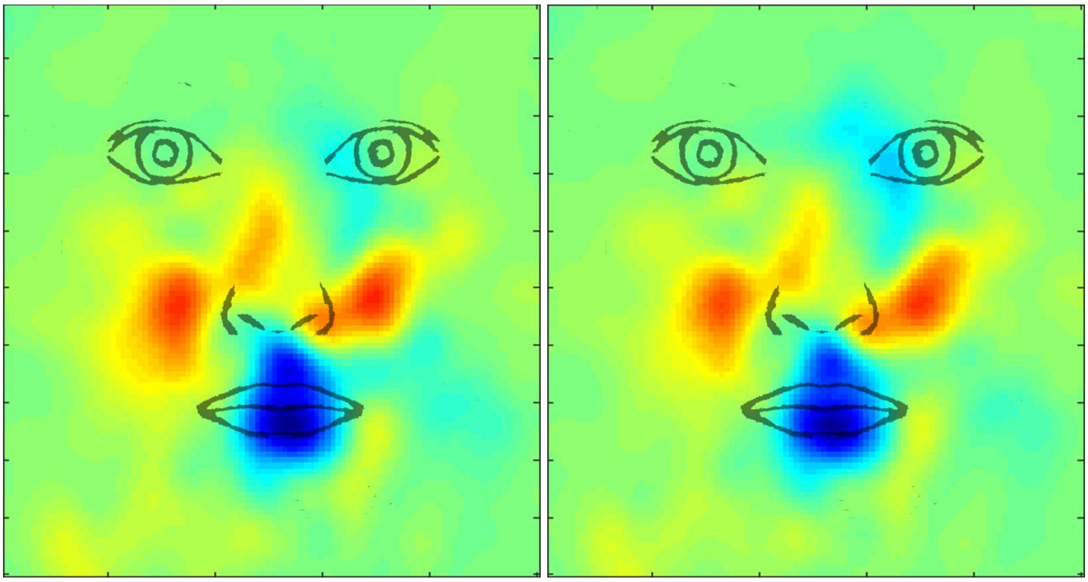


Figure 6.6. Heat maps of group differences in gaze density for dwell data (left) and fixation data (right), illustrating regions that were scanned significantly more by British (blue) and Japanese infants (red).

The statistical analysis, however, did not reveal any significant clusters for either dwell or fixation measures (Figure 6.7 on page 224).

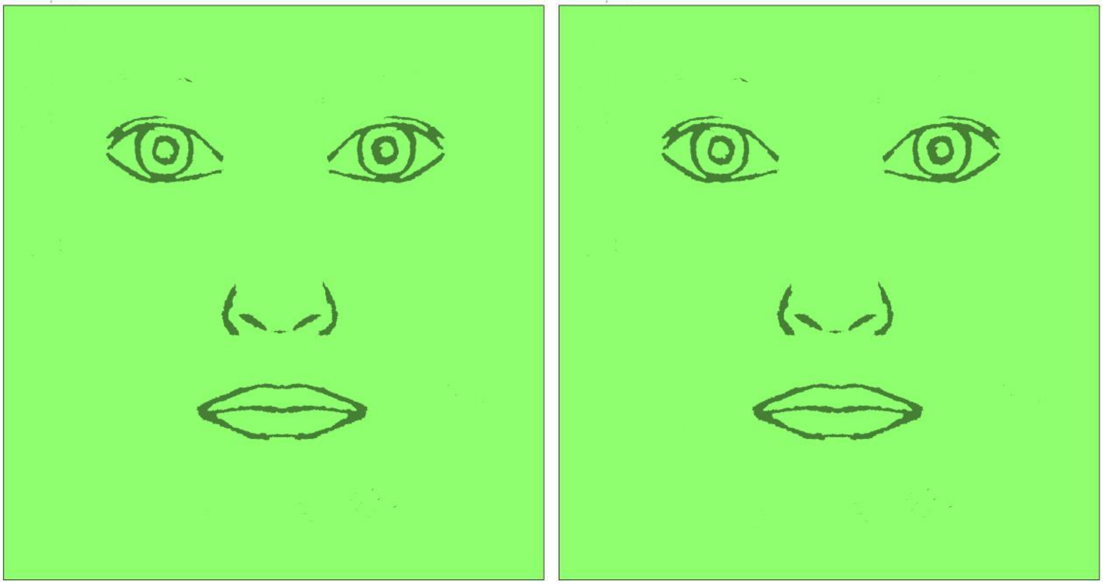


Figure 6.7. Gaze cluster map illustrating the findings from the Monte Carlo permutation test. No significant gaze clusters could be found for dwell (left) or fixation (right) measures.

6.5. Discussion

The current study examined British and Japanese infants' face orienting and scanning behaviour during live dyadic interactions with another adult to determine the extent to which cultural differences manifest in naturalistic social contexts. As expected, the regions-of-interest analysis suggested no significant differences in face looking time between British and Japanese infants. Both groups oriented toward the face roughly half of the time during the 60-second period for analysis. This corresponds also with findings from Frank et al. (2012) who applied screen-based eye tracking techniques to show that infants aged 10 months visually orient to the face approximately half of the time when a whole person (and not only an isolated face) was in view. In contrast to predictions, however, the regions-of-interest analysis did not reveal significant cultural differences in upper or lower face looking, with both groups scanning the lower region approximately two-thirds of total face looking time. As already discussed in Chapter 5, a gradual increase in mouth looking toward the end of the first year of life has also been observed in previous screen-based eye tracking studies, particularly for talking faces (Frank et al., 2012; Król, 2018). This could reflect an adaptive mechanism for language development, with an increased focus on the mouth providing a source for language learning (Hillairet de Boisferon et al., 2018).

To examine face scanning in a more spatially sensitive manner, Monte Carlo permutation tests were also conducted but – contrary to predictions – did not reveal any significant differences in gaze clustering for either the dwell time or fixation time data. This could therefore suggest that cultural differences in face scanning are restricted to screen-based eye tracking paradigms (see Chapter 5) and do not yet generalise to naturalistic social contexts; future studies will be required to also include older infants to map the developmental trajectory. However, the descriptive heat maps that visualised the group differences in gaze density (see Figure 6.6 on page 223) indicated a non-significant numerical trend that British infants looked more at the mouth region, whereas Japanese infants exhibited greater scanning of the face centre and its lower surrounding regions, highlighting that the regions-of-interest coding may not have been spatially sensitive enough. The scanning patterns resembled the screen-based findings in Chapter 5 for social-dynamic faces, i.e. the stimulus category that most closely resembled the faces encountered in the current dyadic interactions. Specifically, the screen-based study (Chapter 5) showed that British participants engaged in significantly more mouth looking, whereas the Japanese group exhibited greater scanning of the nose (i.e., the face centre). It is possible – given also the high within-group variability and relatively small cultural effects found in Chapters 3, 4, and 5 – that a larger amount of gaze data would be required to detect significant cultural differences in infants' naturalistic face scanning.

Alternatively, infants' scanning strategies may indeed manifest differently in naturalistic social contexts. As mentioned above, the infants tested in the current study directed their gaze toward the lower half of the face region for two-thirds of overall face looking time. In Chapter 5, mouth scanning was observed for a third and a fifth of face looking time in British and Japanese participants, respectively (see Table 13.12 and Table 13.27 in Appendix F). Although this suggests that scanning of the lower face may be increased in naturalistic face-to-face interactions than in screen-based paradigms, a range of methodological differences could also account for this finding. For instance, the current study and the experiment in Chapter 5 differed in the location and number of regions-of-interest, which could have potentially resulted in different proportional looking times. The actors who performed the baby-friendly facial actions for the social-dynamic condition in the screen-based study were also explicitly instructed to create movements in the eye region (e.g., raising eyebrows). Given that the research assistants in the current study were not given the same instructions, the mouth typically remained the most salient region which likely resulted in increased mouth scanning. Crucially, however, infants in the screen-based study were presented with the same face stimuli, whereas the current study implemented a dyadic interaction of bi-directional nature. Specifically, research assistants directly responded to the individual infant's behaviour, and it is therefore possible that the observed findings in this study were person-specific, i.e. unique to the local research assistant. This could be due to individual or cultural differences in interaction styles; for instance, increased smiling behaviours could potentially result in greater mouth looking. As already discussed in Chapter 4, directly matching dyadic interaction partners across cultures and ethnicities is virtually impossible, and any attempt to experimentally manipulate natural culture-specific behaviours may mask dynamic characteristics of social interactions that would otherwise give rise to true cultural differences in face scanning. Several research assistants could alternatively be employed for each cultural group to minimise the possibility that findings were unique to the interacting individual; however, this comes with practical and resource challenges, and would additionally increase data variability.

In sum, this chapter presented a cross-cultural face-to-face interaction study with infants for the first time, revealing no significant cultural differences in face orienting or scanning within naturalistic social interactions at 10 months. Both cultural groups largely directed their gaze toward the lower half of the face, with the data-driven visualisations of gaze density showing that British infants particularly focused on the mouth and Japanese infants scanned the face centre and its lower surroundings, but these differences did not reach significance. The visualisations, however, indicate that the regions-of-interest lacked the spatial sensitivity to capture group differences, emphasising the strength of the current data-driven approach. In the final chapter, the findings from all empirical studies of this thesis will be discussed along with general limitations and future directions.

Chapter 7

General Discussion

7.1. Thesis Overview

The overall aim of this thesis was to examine cultural influences in the development of face processing. Specifically, the studies in this thesis investigated cultural differences in infants' and adults' face scanning during dyadic social interactions and within screen-based paradigms. The goals of the empirical studies were to explore face scanning behaviour within naturalistic social settings, to map the developmental trajectory of cultural differences in face scanning, and to raise possible mechanisms to generate new predictions for how the postnatal environment shapes face perception. An additional aim of this thesis was to develop computational eye tracking methodologies that can solve the practical limitations associated with gaze coding of 'real-world' eye tracking data and also enable a novel data-driven analysis. The following will briefly recap each chapter presented in this thesis and summarise key empirical findings.

Chapter 1 illustrated the importance of considering the influence of postnatal social experience on the development of specialised face processing by outlining existing studies that examined cultural modulations on face scanning behaviour. Gaps in the literature were also identified; specifically, Chapter 1 emphasised the need for more naturalistic face-to-face interaction paradigms as well as a developmental framework to cast light on possible mechanisms that can explain how social experience modulates face processing. Chapter 2 then provided a brief overview of eye tracking techniques, discussing in particular the challenges associated with pre-processing and analysing 'real-world' eye tracking data. To overcome these challenges and make a series of face-to-face interaction studies possible for this thesis, Chapter 2 introduced a computational solution for regions-of-interest coding and subsequent gaze annotation. In addition, the limitations of applying a regions-of-interest approach for exploratory purposes was discussed, and a novel data-driven method was developed for 'real-world' eye tracking data (Monte Carlo permutation test). These computational methods were then adopted for the empirical studies presented in Chapters 3, 4, and 6. Chapter 3 applied dual eye tracking techniques in Western Caucasian and East Asian adults to examine cultural differences in face scanning during dyadic social interactions. The findings demonstrated greater face orienting in East Asians compared to Western Caucasians, and this was discussed with respect to possible cultural differences in social signalling. In addition, whereas the regions-of-interest analysis did not identify cultural differences in face scanning, the data-driven method suggested increased looking between the eyes in East Asians and the left side of the face (from the observer's perspective) in Western Caucasians. This scanning pattern was discussed with respect to a possible systematic group difference in the extent to which participants were distracted by the eye tracking hardware that was located on the left side of the face. Given this methodological limitation as well as several other factors that could have influenced study findings (e.g., testing immigrants for the East Asian sample, or the relatively small face size), some interpretations required further clarification. Chapter 4 therefore presented a follow-up study with several modifications to the paradigm and revealed cultural differences in face scanning but not face orienting. Japanese participants exhibited greater scanning of the region between the eyes, whereas British/Irish individuals looked more at the mouth. This was further discussed with respect to cultural differences in social signalling and in the locations of visually informative features of emotionally expressive faces. Given that the findings demonstrated that cultural differences manifested during social interactions, Chapter 5 sought to map the developmental trajectory of cultural differences and cast light on possible cognitive explanations. In particular, Chapter 5 presented a screen-based, cross-sectional study with British and Japanese infants (aged 10 months and 15-17 months) and adults. The findings revealed that several factors independently modulated face scanning, including culture, age, and stimulus characteristics. Against predictions, no age-related changes could be identified for cultural differences, suggesting that cultural differences already largely manifested in the youngest age group. Finally, no relationship could be found with executive function performance. Given that cultural differences in face scanning were present at 10 months, Chapter 6 examined to what extent this generalised to more naturalistic face-to-face interaction play. In contrast to the screen-based study, however, no cultural differences could be found between British and Japanese 10-month-olds, with both groups largely scanning the lower face region. However, a tendency was observed that showed a more central face scanning pattern in Japanese infants, and more direct mouth looking in the British group.

In the following, the thesis findings will be critically discussed from a broader perspective. Specifically, four themes will be considered for the critical discussion, namely cultural differences in naturalistic face scanning, the development of cultural differences, underlying explanations, and eye tracking methodology. Methodological and theoretical limitations of the experiments and possible avenues for future research will also be outlined.

7.2. Critical Discussion

7.2.1. Cultural differences in adults' face scanning

The studies in this thesis adopted the semi-automatic gaze coding tool presented in Chapter 2 to conduct both regions-of-interest analysis and novel data-driven methods. The findings thereby revealed, for the first time, cultural differences in face scanning during naturalistic social interactions (Chapters 3 and 4). In addition, a range of new face stimuli was created for this thesis to directly compare face scanning patterns across different stimulus conditions and to examine how this manifests cross-culturally within a screen-based paradigm (Chapter 5).

Several consistent patterns in face scanning behaviours of adult participants were found across the studies presented in this thesis (Chapters 3, 4, and 5). For the social interaction paradigms (Chapters 3 and 4), both cultural groups showed more visual orienting to the face of the conversational partner during periods of listening compared to speaking. This speech effect was not only replicated within this thesis but is also consistent with previous studies (e.g., Freeth, Foulsham, et al., 2013). In addition, both cultural groups exhibited increased face orienting for the introductory task relative to the story-telling game (Chapters 3 and 4). This task effect could be explained by increased social signalling during the early stages of meeting a new person, or by a greater need to visually orient away from the face during the more effortful story-telling game in order to reduce cognitive load. Traditionally, an attentional bias to faces has been revealed across many screen-based studies (e.g., Bindemann et al., 2005; Ro et al., 2001; Theeuwes & Van der Stigchel, 2006); however, the present findings demonstrate that situational factors, such as moments of speech or the nature of a task, also need to be taken into account to establish the *degree* of face orienting in naturalistic settings.

As mentioned above, Chapter 3 revealed the possibility that methodological limitations potentially affected the study findings and interpretations, emphasising in particular that dual eye tracking techniques may not be optimal to study face looking behaviour. The social interaction paradigm was therefore modified for Chapter 4, which presented a follow-up study that addressed relevant limitations. Specifically, Chapter 4 demonstrated that Japanese adults scanned the upper face more, and specifically the region between the eyes, whereas British/Irish individuals showed a preference for the mouth region. However, the Japanese group did not exhibit such patterns in the screen-based study (Chapter 5), and instead showed central face scanning strategies for the dynamic-neutral and dynamic-social conditions. A possible explanation that accounts for this inconsistency between the findings from Chapters 4 and 5 could be that Japanese participants engaged in more eye looking during the face-to-face interaction study as a social signal for the conversational partner (Richardson & Gobel, 2015; Risko et al., 2016). In contrast, given that no social partner was presented for the screen-based paradigm, Japanese participants may instead have focused more on extracting relevant visual information for face processing; by increasingly focusing on the central face region, visual information could be extracted both from the eyes and the mouth using extrafoveal vision. This account would therefore also be in line with the proposal that East Asians, compared to Western Caucasians, can use their extrafoveal vision more efficiently for face processing (Caldara et al., 2010; Miellet et al., 2012).

A more consistent pattern in scanning behaviours was observed for Western Caucasian participants. As with the social interaction study (Chapter 4), the screen-based study (Chapter 5) also found that British/Irish participants focused more on the mouth than the Japanese group across all stimulus conditions (static, dynamic-neutral, dynamic-social). The increased mouth looking in Western Caucasian participants therefore points to a highly consistent marker for cultural differences that can be observed across different experimental stimuli and paradigms within and beyond this thesis. In particular, cross-cultural studies employing face stimuli with neutral (e.g., Blais et al., 2008) or emotional expressions (Jack et al., 2009; Senju, Vernetti, Kikuchi, Akechi, Hasegawa, et al., 2013) also identified increased mouth scanning in Western Caucasian compared to East Asian participants. However, the increased eye scanning that has previously been reported (cf., Blais et al., 2008; Kelly, Liu, et al., 2011; Kelly et al., 2010; Kita et al., 2010; Rodger et al., 2010) could not be found in any of the studies in this thesis. This discrepancy is likely due to various methodological differences, most notably the lack of an explicit face processing task in the present thesis. In other words, increased eye scanning may have been beneficial for face recognition tasks but not for social interaction or free-viewing paradigms. Although future studies will need to disentangle the different explanations underlying the observed face scanning patterns, the findings highlight the need to study cultural differences across a range of experimental settings and paradigms.

The studies in this thesis only investigated cultural modulations in adults with intact face processing abilities. Nevertheless, the present findings challenge, to an extent, our current understanding of both typical and atypical face processing. Atypical face perception has previously been identified as a symptomatic marker of certain neurological and mental health conditions, such as schizophrenia (e.g., Hall et al., 2004; McCleery et al., 2015; for a review see Marwick & Hall, 2008), autism spectrum conditions (ASC; e.g., Dalton et al., 2005; Pierce, Müller, Ambrose, Allen, & Courchesne, 2001; see Weigelt, Koldewyn, & Kanwisher, 2012), or William's syndrome (e.g., Riby &

Hancock, 2008; Riby, Hancock, Jones, & Hanley, 2013). For instance, ASC has been shown to manifest in lower performance for face recognition and face memory tasks, compared to individuals without ASC (e.g., Chawarska & Shic, 2009; Hauck, Fein, Maltby, Waterhouse, & Feinstein, 1998; for a review see Weigelt et al., 2012). This has additionally been linked to inefficient visual strategies in face scanning patterns, with studies traditionally revealing decreased attentional orienting to faces and eyes, and a preference for the mouth region (Chawarska & Shic, 2009; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Riby & Hancock, 2008; Riby et al., 2013). The findings in this thesis, however, showed that face orienting was significantly decreased during periods of speech in social interactions, particularly for the story-telling game (Chapters 3 and 4). In addition, the findings reported in this thesis demonstrated a preference for the mouth in the Western Caucasian group (Chapters 4, 5, and 6). Altogether, the findings revealed typical scanning behaviours that have traditionally been considered symptomatic of ASC. Although it is possible that previously observed behavioural markers of ASC, such as a visual preference for the mouth, may indeed represent inefficient strategies to accomplish certain face processing tasks, future studies will need to consider also more naturalistic paradigms to examine face scanning – and social attention more generally – to better understand the nature of (a)typical face processing. Future research would also need to examine cultural differences in face processing of ASC participants to establish whether or not, and to what extent, the postnatal social environment modulates the development of specialised face perception in ASC.

The thesis findings also revealed high variability in face orienting and scanning, even within a single cultural group (Chapters 3, 4, and 5), suggesting that individual differences vary considerably. A possible explanation that could account for this high variability concerns the lack of an explicit face processing task such as face recognition or emotion categorisation, which in turn meant that participants were not restricted by similar task-dependent top-down influences on face processing. Various factors modulated face orienting and scanning (e.g.., speech, social context, stimulus characteristics) and individual differences in the extent to which each factor influences face processing may exist. The relative contribution of each factor may also dynamically change with time, particularly in naturalistic social interactions. Consequently, it is possible that the extent of cultural modulations on eye movement behaviour changes depending on the relevance of other contextual factors, giving rise to the high data variability. This has also been illustrated in a study showing that cultural differences in analytic versus holistic viewing strategies for scene perception in Western Caucasian and East Asian adults (cf., Chua et al., 2005) could be replicated (Rayner, Castelhano, & Yang, 2009), but only when scenes did not contain salient objects (Rayner, Li, Williams, Cave, & Well, 2007). This, along with the thesis findings, thus highlight the context-dependent nature of cultural modulations on viewing behaviour.

7.2.2. Mapping the developmental trajectory

The present thesis aimed to map the developmental trajectory for cultural differences in face scanning. A key prediction was that cultural differences should become more evident with age, thereby supporting an experience-dependent account of specialised face processing (Nelson, 2001). Previously, cultural influences on face scanning during infancy were only investigated in a single cultural group (Liu et al., 2011; Wheeler et al., 2011; Xiao et al., 2013) or in a single age group (Geangu et al., 2016). To address this gap in the literature, Chapter 5 presented the first cross-cultural infant study investigating face scanning strategies in two infant age groups as well as in adults, thereby providing insight into the developmental trajectory. In contrast to the predictions, however, cultural differences did not become increasingly evident with age; instead, culture and age independently modulated face scanning. Although this suggests the possibility that cultural influences did not modulate the development of face processing, studies have previously observed cultural differences in infants younger than 10 months (e.g., at 7 months; Geangu et al., 2016), which in turn raises the question of whether or not cultural modulations may be involved in face processing beyond the first year of life. Future cross-cultural studies could address these outstanding questions by examining face scanning strategies in a much younger infant age group, e.g. at 5 months when cultural differences have not yet been reported to be present. Additionally, longitudinal study designs could be adopted to track the developmental trajectory of face scanning behaviour in each cultural group more thoroughly.

Given that cultural differences already emerged by 10 months (Chapter 5), it was expected that this would also extend to face-to-face interaction play. To date, no study had examined face scanning during infancy within a naturalistic social context. Chapter 6 therefore presented the first study to examine naturalistic face scanning in infants, casting light also on cultural differences. A face-to-face eye tracking paradigm for developmental populations, which previously had only been adopted for 6-year-old children (Falck-Ytter, 2015; Falck-Ytter, Carlström, & Johansson, 2015), was also established. Only more recently has such a face-to-face eye tracking paradigm also been adopted for 10-month-old infants (by the same research group; Nyström et al., 2017; Thorup et al., 2016, 2018), although the distance between the infant and the interacting adult in those studies was much larger (200 cm) than the distance required for the experiment in Chapter 6 (155 cm). Although initially head-mounted eye tracking techniques were adopted during the piloting stage, this led to a high drop-out rate due to infants not tolerating the equipment (see Chapter 2 for more details). In contrast, infants were able to tolerate the eye tracking paradigm presented in Chapter 6, which could therefore also be applied in future face-to-face interaction studies.

Against predictions, Chapter 6 did not reveal significant cultural differences, with both cultural groups scanning the lower face region approximately two thirds of face looking time, though Japanese infants showed a tendency to scan the region slightly above the mouth. Given that cultural differences were observed in Chapter 5 but not Chapter 6, the question remains as to whether this was the result of insufficient data (and therefore lack of statistical sensitivity) for the face-to-face interaction study or the high data variability that may have masked true cultural differences. Taking into account the findings across all empirical studies, however, an alternative explanation is that other contextual factors were more relevant for infants during the social interaction. For instance, it may have been more useful to look at the mouth. Unlike the screen-based study (Chapter 5), infants were presented with language, which audio-visually matched the lip movements of the research assistant, during the social interaction (Bahrick et al., 1998; Dodd, 1979). This was particularly relevant for the present age group given that infants were entering the language learning stage (Oller, 2000). Altogether, this suggests that the extent to which cultural differences in face scanning can be observed may depend also on contextual factors, mirroring the interpretation proposed above for adults (Section 7.2.1; cf., Rayner et al., 2009, 2007).

7.2.3. Developmental mechanisms for the expression of cultural differences

One aim of the thesis was to raise potential mechanisms that can explain how postnatal social experiences can influence the development of specialised face processing. Chapter 5 also directly examined the role of executive function in the emergence of cultural differences. For instance, as outlined in Chapter 1, increasingly greater abilities in attentional control could help infants to visually disengage from irrelevant features (e.g., irrelevant salient features) and focus on more informative face regions, thereby giving rise to culturally-relevant attentional strategies. However, a relationship between more advanced executive function and increasingly culture-specific face scanning was not observed. Although more research will be required to understand the emergence and manifestation of cultural differences in face scanning, the study findings from this thesis have cast light on several potential driving factors, which will be outlined in the following.

Chapter 4 found that British/Irish participants scanned the mouth region more and the eye region less than Japanese individuals. This was discussed with respect to differences in visual information contained in the mouth region as a result of linguistic differences between English and Japanese languages. This could be examined in more detail by, e.g., investigating the extent to which bilingual participants in each cultural group flexibly shift their gaze across the face. In addition, it was discussed that cultural differences in social signalling may have resulted in increased eye looking in the Japanese group. This could be further investigated by manipulating the social presence during face-to-face interactions; for instance, by using a half-silvered mirror and instructing participants that they can or cannot be seen by their conversation partner (cf., Myllyneva & Hietanen, 2015). Chapter 5 further suggested that different learning requirements at each age – specifically, social learning at 10 months, language learning at 15-17 months, and a more flexible system for adults – resulted in the observed differences in face scanning. However, this will need to be investigated more closely by, for instance, taking relevant developmental measures such as the Communicative Development Inventory (CDI) that can assess the size and growth of infants' vocabulary (Fenson et al., 1994; Fenson, Marchman, Thal, Dale, & Reznick, 2007). An effect of stimulus condition was also observed, showing an overall pattern of increased mouth and decreased eye scanning in the dynamic conditions compared to the static condition, which was particularly evident for dynamic-neutral relative to dynamic-social face stimuli. This finding was discussed with respect to low-level saliency as a driving factor. However, this will need to be quantified and directly investigated in future work using saliency models (e.g., Frank et al., 2009).

Examining the potential origins of cultural differences will also allow a better understanding of the development of face processing. Chapter 5 revealed increased mouth scanning for the dynamic-neutral condition and more eye and mouth looking for the dynamic-social condition in both cultural groups, but the extent of this finding significantly differed between cultures. It is possible that those regions contained more visual information for face processing and therefore elicited more scanning. Alternatively, low-level motion could have driven scanning behaviour in a bottom-up fashion, raising the question whether low-level saliency could represent a very early source for cultural differences in face scanning. Studies have demonstrated that young infants tend to orient to and fixate highly salient regions (Frank et al., 2009), and it is possible that differences in the degree of articulation between languages (e.g., English versus Japanese) may result in differences in saliency of the mouth region, which in turn differentially affects visual orienting. This could also apply to movements of individual facial features, such as increased smiling or the use of the eye region during infant-directed actions.

As mentioned in Chapter 1, Geangu et al. (2016) proposed the possibility that the caregiver's facial expressivity may provide an early source of cultural learning. Specifically, as infants acquire more visual experience with the caregiver's face, they may learn to attend to the visually informative regions. The locations of those regions may in part overlap with areas characterised by increased low-level saliency; for instance, mouth movements can be informative with respect to language and are simultaneously associated with higher low-level motion compared to other face regions. More socially-relevant areas, however, may also be important; for instance, the eye region could be more informative for emotional expressions, particularly in East Asians (cf., Jack et al., 2012). Such cultural learning via the caregiver suggests a developmental mechanism that would also be consistent with two studies that highlight the significant role of early familial experience in social development. For instance, Senju et al. (2015) found that infants raised by blind parents attended less to dynamic eye gaze despite scoring typically on social communication measures. With respect to culture, Kelly, Jack, et al. (2011) showed that only 25-30% of British Born Chinese (BBC) adults employed "Western" triangular scanning patterns, and informal interviews revealed that most BBCs were not exposed to Western cultures until they started school. The role of the familial environment could therefore significantly impact the development of cultural differences in face scanning. As mentioned in Chapter 6, caregivers' eye movements were also recorded using head-mounted eye tracking techniques while they interacted with their child. The aim was to examine the caregivers' eye movements as a source for infants' learning of face scanning. However, the parent-child interaction paradigm was characterised by a high drop-out rate in the Japanese group due to fussiness or insufficient data (69.57%, or 16 out of 23; compared to 20%, or 4 out of 20, for the British group), which likely resulted from Japanese infants not tolerating the physical separation from their caregiver (at a distance of 155 cm). This issue could be solved once a head-mounted eye tracker has been developed that infants can tolerate (for a more detailed discussion see Section 7.3.1 below). An additional possibility could involve examining caregivers' facial expressions during parent-child interactions and examine how the role of, e.g., low-level saliency corresponds to infants' eye movements during face scanning throughout development. Altogether, extensive future work will be required to disentangle the numerous driving factors that lead to highly specialised face processing.

7.2.4. Eye tracking methodology

Chapter 2 presented novel computational methods for semi-automatic regions-of-interest coding and a more spatially sensitive data-driven analysis. These methods were applied to enable a series of face-to-face interaction paradigms (Chapters 3, 4, and 6), and provided more detailed insight into the manifestation of cultural differences in face scanning when adopting the data-driven analysis. As mentioned in Chapter 2, previous studies have typically divided the face area into an upper and lower half for regions-of-interest analysis to study scanning strategies during social interactions. Vabalas and Freeth (2016), for instance, conducted a face-to-face interaction paradigm to examine face scanning (in individuals with low and high autistic traits). A traditional regions-of-interest analysis was conducted on proportional fixation times on the upper face, lower face, body, and background. Although the findings demonstrated significant differences in fixation time between some of the regions-of-interest, fixation times between the upper and lower face did not significantly differ. Although this may reflect the true scanning behaviour of participants, the binary nature of the regions-of-interest approach (i.e., upper versus lower face) may have restricted insight into more subtle differences in face looking behaviour. Although the spatial pooling of gaze data increases statistical sensitivity for detecting differences – making the regions-of-interest analysis a useful approach when testing specific predictions – the findings in this thesis have demonstrated the need for more spatially sensitive methods. For instance, although the regions-of-interest analysis in Chapter 4 revealed that Japanese adults spent a greater proportion of face looking time on the upper half (eye region) than British/Irish participants, the data-driven analysis showed that this was restricted to the area between the eyes. In addition, Chapter 3 found no cultural differences in upper/lower face looking when using the regions-of-interest analysis. However, when adopting data-driven analysis, increased scanning of the region between the eyes in East Asian participants and of the left side of the face (from the observer's perspective) in Western Caucasians was revealed. This in turn revealed the possibility that the eye tracking equipment may have served as a visual distractor and differentially affected cultural groups – a finding that otherwise would have been missed using the traditional regions-of-interest approach. Although the methods were developed to examine cultural differences, the tools provide a computational solution that can also be applied to examine orienting behaviour and scanning strategies in other populations (e.g., ASC individuals; see Section 7.2.1) and across other paradigms. The data-driven analysis in particular can cast light on more subtle differences in eye movement patterns within naturalistic settings.

7.3. Limitations

After having discussed study findings, the following sections will outline some limitations associated with the empirical work presented in this thesis.

7.3.1. What is 'naturalistic'?

Throughout this thesis, the term *naturalistic* was adopted to refer to paradigms or settings that, compared to screen-based experiments, more closely resemble the everyday life that we experience. Chapter 1 discussed several key limitations of screen-based paradigms when studying face scanning, or social attention more generally, including restrained viewing conditions, and a lack of dynamic features, environmental distractors, and social presence. Chapters 3, 4, and 6 aimed to address these limitations by adopting live dyadic social interaction paradigms. For each study, procedures were in place to ensure participants felt comfortable, which in turn allowed for more natural social interactions. For instance, adult participants were given time to briefly familiarise themselves with their conversational partner (i.e., another participant in Chapter 3 or a research assistant in Chapter 4). They were also informed that the content of their speech would not be analysed, thereby minimising the possibility that participants become self-conscious during periods of speaking. The first task of the testing session was additionally designed to be of introductory nature and participants were provided with specific examples of what could be mentioned, such as their name, work, home country, or hobby, and they were also free to have a conversation afterwards.

However, despite these measures, the paradigms could, to an extent, still be considered artificial. Given that adults were wearing a head-mounted eye tracker throughout the social interaction, they were aware that their eye movements were recorded which in turn could have influenced their scanning behaviour. Risko and Kingstone (2011) found that individuals wearing an eye tracker oriented less to a provocative stimulus (a swimsuit calendar) than control participants, suggesting that the eye tracker served as an implied social presence that in turn affected orienting and scanning behaviours. Nasiopoulos, Risko, Foulsham and Kingstone (2015) further found that increased prosocial behaviours that resulted from wearing an eye tracker diminished after 10 minutes, indicating that the implied social presence of eye trackers gradually disappeared as participants adapted to the equipment. In Chapters 3 and 4, adults were therefore mounted with the tracker as soon as they signed the consent forms, and study details and demographic information were collected only afterwards to provide sufficient time to adapt to the eye tracking glasses. After the testing session, participants also typically stated that they forgot about the hardware as soon as they were engaged in the conversation. However, although measures were taken to minimise the social presence effect of eye trackers, it is still possible that participants were influenced by the equipment.

Adult participants (Chapters 3 and 4) were also asked to introduce themselves and tell their conversational partner about a personal event or experience. This was done to establish some consistency and structure between dyads and to facilitate a conversation. However, as the findings from Chapters 3 and 4 demonstrated, face orienting was higher for the introductory task compared to the story-telling game, suggesting the possibility that scanning behaviour may manifest differently in other task-free scenarios. Given that face looking behaviour is highly contextual, future work will need to examine cultural differences in face scanning across a range of experimental settings and tasks.

In addition, infants (Chapter 6) were given enough time to familiarise themselves with the testing environment. During the piloting stage, it also became apparent that infants became less fussy if they spent some time interacting with the research assistant before the recording started, and this familiarisation period was introduced for the testing stage. As reported in Chapter 6, infants sat in a high-chair at a distance of approximately 155 cm from the research assistant. However, this experimental set-up may not necessarily resemble a typical play environment since infants are usually free to move and may be positioned closer to the person who is interacting with them. To record eye movements with the TX300 eye tracker, however, a stable viewing distance of 65 cm was required, and the research assistant needed to be positioned behind the equipment, making the large distance and use of a high-chair inevitable. A solution for future studies could involve head-mounted eye tracking techniques. Currently, Positive Science is the only manufacturer that offers an infant-friendly model. As mentioned in Chapter 2, this eye tracker was initially used during the piloting stage of the study in Chapter 6 but revealed a 72.73% (8 of 11) drop-out rate since infants did not tolerate the beanie or eye camera arm. Unlike previous studies that successfully used the Positive Science eye tracker (e.g., Kretch et al., 2014; Yu & Smith, 2013), infants in the pilot study were younger and therefore less likely to tolerate the equipment. Crucially, infants became fussy since no toys were given as distractors – which would have kept their hands away from the equipment – to avoid significantly reduced face looking and scanning time (cf., Yu & Smith, 2013). To conduct face-to-face paradigms that more closely resemble infants' everyday situations, it will therefore be necessary to first develop a head-mounted eye tracker that infants can tolerate.

7.3.2. Systematic differences in social interaction partner

The cross-cultural studies that adopted face-to-face interaction paradigms (Chapters 3, 4, and 6) inherently contained systematic differences in social interactions that are problematic to solve but need to be acknowledged. Adults in Chapter 4 and infants in Chapter 6 interacted with a local research assistant, raising the possibility that the observed face scanning patterns were specific to the social interaction partner. Given the bidirectional nature of social interactions, it is possible that individual or cultural differences in the research assistants' interaction styles modulated participants' scanning behaviour; for instance, increased smiling or articulation could potentially result in increased mouth scanning due to greater low-level saliency or more visual information being located in that region (Jack et al., 2009). Differences in facial feature characteristics also could have modulated scanning behaviour. The facial physiognomy hypothesis has previously been proposed to provide a bottom-up account of cultural differences in face scanning. Specifically, given that ethnicities differ in their physiognomy (e.g., Le, Farkas, Ngim, Levin, & Forrest, 2002), certain facial features may be more informative or salient depending on the ethnicity (Brielmann, Bülthoff, & Armann, 2014; Wang et al., 2015). In their screen-based study, for instance, Wang et al. (2015) presented Chinese adults with Chinese, Caucasian, and ambiguous faces (morphed 50% Chinese and 50% Caucasian) and found different scanning patterns between the stimulus categories. Specifically, participants looked more at the nose of Chinese (own ethnicity) faces and the eyes of Caucasian (other ethnicity) faces, but they looked at the eyes, nose, and mouth more equally for ambiguous faces even when participants believed that these faces were of Chinese ethnicity. This therefore suggests the possibility that the visual properties of facial features may have driven face scanning patterns. For the face-to-face interaction studies in this thesis, it is difficult to disentangle the extent to which facial features specific to the individual or ethnicity may have driven scanning behaviours. Given that the matching of social interaction partners across cultures and ethnicities is virtually impossible, a solution could involve employing several research assistants within each cultural group to minimise the possibility that findings were unique to the social interaction partner. However, this would come with practical and resource challenges, and would additionally increase data variability. An alternative approach could therefore involve virtual reality (VR) designs whereby participants interact with realistic avatars whose appearance or behaviours can be experimentally manipulated. However, there are three limitations with the VR approach that need to be considered with respect to face-to-face interaction paradigms. First, differences in interaction style or physiognomy may mask informative, true cultural differences if the VR paradigm is experimentally manipulated or too highly controlled. Secondly, infant-friendly VR designs have not yet been developed. Thirdly, it is still unclear to what extent an avatar within an artificial VR environment can induce the social presence inherent to everyday life social interactions. Indeed, VR paradigms of this nature still need to be validated.

7.3.3. Parameter setting for analysis

The analysis of every study presented in this thesis involved choosing appropriate values for a range of parameters, and the initial settings can therefore directly impact the experimental findings. In the following, relevant parameters will be highlighted to acknowledge their influence on the study findings and interpretations.

To examine scanning behaviour during face-to-face interactions in Chapters 3, 4, and 6, a period of 30 seconds (for adults; in line with Freeth, Foulsham, et al., 2013) or

60 seconds (for infants) was considered for analysis. The screen-based study in Chapter 5 also presented each face stimulus for 18 seconds to compute proportional dwell and fixation times. However, cultural differences in the temporal dynamics of face scanning are not yet understood, and it is therefore possible that different durations may result in different study findings. Or, Peterson, and Eckstein (2015), for instance, presented face stimuli to Western Caucasian and East Asian adults and found that the first fixation locations during a face recognition task did not differ between cultural groups. Cultural modulations on eye movement patterns may therefore only become evident during later stages of face scanning, suggesting that the period chosen for analysis can directly affect study findings. With respect to comparisons between stimulus conditions (static, dynamic-neutral, dynamic-social) in the screen-based study in Chapter 5, it is possible that participants required less time to visually encode static faces given that still images contain less information than dynamic faces (Itti, 2005), raising the question whether or not the same duration for stimulus presentation should be selected. In terms of age differences, adults may have also been more efficient in extracting visual information compared to infants, making a uniform stimulus presentation duration a potential limitation for study interpretations. Future work will need to explore the temporal dynamics of face scanning in more detail to understand how viewing patterns unfold over time (see also Section 7.4.2 for a discussion on temporal dynamics of face processing).

Parameter settings were also required to process the raw eye tracking data for Chapters 4, 5, and 6, particularly for the detection of fixations. The choice of parameter values had a direct impact on the number and duration of gaze measures including fixations (Saez de Urabain et al., 2015). Currently, no gold standard approach exists for processing different types of eye tracking data, which poses a significant issue for eye movement research. To ensure a degree of consistency for the studies in this thesis, parameters were selected in line with previous studies wherever possible (Saez de Urabain et al., 2015; Wass et al., 2013). All parameter values were also fully reported in the relevant empirical chapters to facilitate study comparisons and to clarify and define eye tracking measures such as fixations (cf., Hessels, Niehorster, et al., 2018). However, given that the selection of parameter values introduced a degree of subjectivity, this needs to be acknowledged as a potential limitation.

In addition, the uncorrected *p*-value threshold required for the Monte Carlo permutation test (see Chapter 2 for more details) represented an additional parameter that needed to be set manually. The chosen *p*-value directly affected the location, size, and numbers of clusters that were identified for permutation testing (see Chapter 2). For the relevant studies in this thesis (Chapters 3, 4, and 6), a common, moderately strict threshold of 0.01 was selected. However, it is possible that more conservative thresholds would eliminate significant clusters (in Chapters 3 and 4) or that more lenient thresholds would give rise to significant cultural differences (in Chapter 6). A solution to this limitation could involve the application of the Threshold-Free Cluster Enhancement method (TFCE; Smith & Nichols, 2009). The TFCE approach was developed for neuroimaging analysis and adopts a data-driven approach to determine an appropriate threshold based on map surface characteristics (e.g., peaks, troughs; Smith & Nichols, 2009). However, given that the Monte Carlo permutation test was developed for 'real-world' eye tracking data in this thesis, it is unclear whether threshold settings used in neuroimaging are also applicable to gaze data, and future work will need to examine the suitability of different parameter values in more detail.

With respect to the studies in this thesis, the primary aim was to investigate cultural differences in face scanning. Given that the same parameter values were used for each cultural group, no systematic differences were introduced. However, the extent to which cultural differences were observed may have been modulated by the parameters mentioned above. To facilitate data integration and interpretations, all parameter values were fully reported in this thesis, but caution is required when comparing findings from studies that applied different parameter settings.

7.3.4. Cultural differences and the role of genetics

Human social behaviour has been suggested to arise from a gene-culture coevolution that involves an interaction between culture and genes (Gintis, 2011). Way and Lieberman (2010), for instance, reviewed evidence on the association between increased sensitivity to social experiences and the alleles of certain neurotransmitter transporter polymorphisms (e.g., serotonin: 5-HTTLPR; opioid: OPRM1 A118G). It was concluded that collectivist cultures potentially emerged and persisted in regions characterised by higher proportions of the aforementioned alleles, illustrating the importance of considering genetic factors for cross-cultural comparisons.

An issue inherent to the interpretation of the findings in this thesis concerns the genetic differences between cultural groups: Western Caucasians and East Asians differ not only in their cultural experience, but also in their genetic make-up (e.g., Chiao & Blizinsky, 2010; Way & Lieberman, 2010). To disentangle genetic factors from cultural effects, Kelly, Jack, et al. (2011) studied British Born Chinese and found that only 25-30% of participants employed a face scanning strategy typical for Western Caucasian participants (cf., Blais et al., 2008), suggesting significant biological contributions to face perception processes. However, informal interviews revealed that participants were not particularly exposed to a Western cultural environment during infancy and early childhood, pointing to early socialisation and familial experience as potentially important factors in shaping face perception. Moreover, Caldara et al. (2016) showed that East Asian adoptees raised in Western cultures typically employed scanning strategies consistent with those of Western Caucasian participants.

Although these findings highlight a significant role of postnatal social experience in shaping face perception processes, these studies did not consider genetic variations within a single cultural group and how this could potentially interact with the cultural environment. For instance, Kitayama et al. (2014) found that cultural differences in social orientation – Western Caucasians reported higher levels of independence and East Asians scored higher on interdependence – could only be observed for carriers of the 7-

or 2-repeat alleles of the dopamine receptor D_4, which has previously been suggested to be involved in cultural learning (Chen, Burton, Greenberger, & Dmitrieva, 1999). When considering non-carriers only, cultural differences could not be found. Kim et al. (2010) also found cultural differences in perceptual styles – East Asians exhibited a more holistic pattern than Western Caucasians – but again, this pattern was only found in G allele carriers of the receptor HTR1A. Consequently, although the importance of considering postnatal social experiences in the development of specialised face perception was highlighted throughout this thesis, the above studies illustrate the possibility that genetic factors moderate the *extent* to which the postnatal social environment influences the development of face perception. Genetic factors will therefore need to be considered in future studies to establish the relative importance of postnatal influences for face processing and how this may differ between individuals.

7.4. Future Directions

The eye tracking methodologies presented in Chapter 2 and the studies in Chapters 3, 4, 5, and 6 have addressed several research aims relating to the manifestation of cultural differences in face scanning for different age groups and within different experimental settings. The findings in turn raised novel questions that could be addressed in future work. In the following sections, several potential avenues for future work will be outlined.

7.4.1. Improving the regions-of-interest (face) tracker

Chapter 2 presented a novel computational method to detect and track faces in a semi-automatic fashion, allowing also for rapid gaze classification. To expand the applicability for other research areas, there are several possibilities to improve the semi-automatic tool. The scripts can be refined and fully documented to share them publicly. Other classifiers could be included to allow the detection and tracking of non-face objects. Although the semi-automatic tool can already *track* non-face objects – the user can manually mark up any region-of-interest and the KLT algorithm then extracts relevant visual properties within the bounding box for tracking (see Chapter 2 for details) – the tool cannot currently *detect* non-face objects. In addition, non-rectangular shapes (e.g., an ellipse) could be incorporated for more precise regions-of-interest detection and tracking.

7.4.2. The temporal dynamics of face scanning

Given that the thesis aimed to describe face scanning strategies in more detail, the findings of all studies (Chapters 3, 4, 5, and 6) examined gaze locations across a range of experimental stimuli and settings. Considering the *temporal* domain was beyond the scope of this thesis; however, future research will need to consider both spatial and temporal information to examine cultural differences in face scanning – especially with regard to highly dynamic social interactions. As mentioned above, Or et al. (2015) have previously found that cultural differences in fixation locations cannot be observed during the very early stages of face viewing, suggesting that cultural differences only emerge during later processing stages. Future studies will therefore need to examine the dynamic time-course of scanning strategies, which in turn can provide insight into the factors that modulate gaze behaviours (e.g., speech patterns, social cues of the interaction partner). A particular challenge for future research will involve the disentangling of mechanisms when the temporal dynamics are also considered; for instance, given that social interactions are highly dynamic, scanning of the eye region in one moment could be associated with a different underlying mechanism than eye scanning at a later time (e.g., visual information extraction for face recognition versus social signalling the end of a speaking period to the conversational partner; cf., Ho et al., 2015).

The methodology presented in Chapter 2 retained the temporal information on gaze locations during the pre-processing stages for eye tracking data. Given that the gaze data is collapsed across the temporal domain only in the final steps of pre-processing, the computational methods in Chapter 2 would require minimal modifications to the scripts to include temporal analysis of eye movement data. For instance, scan path analysis or gaze transitions between regions-of-interest could be considered as dependent measures for spatiotemporal analysis of face scanning.

7.4.3. What about covert attention?

All studies in this thesis examined participants' visual attention to faces by recording eye movements. However, it is also possible to engage in *covert orienting,* which refers to a shift in attention without a corresponding shift in eye movements (Posner, 1980). In other words, covert attention cannot be studied using eye tracking measures such as dwell or fixation time. As outlined in Chapter 1, several studies have (implicitly) examined covert attention more closely in the context of cultural differences in face scanning. In their screen-based study, for instance, Miellet et al. (2012) masked central vision and found that Western Caucasians shifted their gaze to the face centre to achieve face recognition, whereas no changes were observed for East Asians. Given that face recognition performance was not negatively affected, this pointed to a greater use of extrafoveal vision for face recognition in East Asians. However, covert attention could also occur within more naturalistic settings. It is possible for individuals to covertly attend to the face of another person while foveating elsewhere. Such covert orienting could even be modulated by social norms and thereby affect cultural groups differently. For instance, if excessive gazing is considered more inappropriate in one compared to another culture, the extent to which individuals will engage in covert attention to faces may differ cross-culturally. Future work will need to examine the role of covert attention in naturalistic face orienting and scanning behaviour more closely.

7.4.4. The neural basis of cultural differences in face processing

As outlined in Chapter 1, neuroimaging studies have identified several neural substrates involved in face processing, including the FFA, OFA, and STS (Haxby et al., 2000; Hoffman & Haxby, 2000; Kanwisher et al., 1997). Based on such evidence, several neuroanatomical models have been proposed that describe and explain the different stages involved in face processing (Bernstein & Yovel, 2015; Gobbini & Haxby, 2007; Haxby et al., 2000; O'Toole, Roark, & Abdi, 2002). In their influential neural model, for instance, Haxby et al. (2000) proposed a core network involving the FFA, which processes invariant aspects such as identity, the STS, which processes variant factors including emotional expressions, and the OFA, which inputs to the STS and FFA. The model also proposed an extended system to account for various secondary cognitive functions involved in face processing; for instance, the auditory cortex is included for speech perception and the amygdala for emotion processing. Although such models highlight the importance of considering specialised face processing as a result of an extensive neural network involving a wide range of functional components, the models were largely based on evidence from neuroimaging studies that employed relatively simple stimuli such as static images of faces (e.g., Haxby et al., 2000).

The study findings of this thesis, however, revealed that different experimental stimuli and paradigms gave rise to different eye movement patterns, and factors such as low-level motion or social presence were proposed to modulate face scanning behaviour. More recently, neural models of face processing have been updated to consider also evidence on *dynamic* face perception; for instance, O'Toole et al. (2002) extended the model proposed by Haxby et al. (2000) to include information extraction processes for dynamic faces, and Bernstein and Yovel (2015) further differentiated the processing of facial forms and motions. However, existing models implicitly assumed that face processing is primarily concerned with the extraction of relevant visual information to achieve a specific task such as face recognition. In other words, factors common to everyday situations such as social presence were not explicitly considered. The findings from the social interaction studies (Chapters 3 and 4), however, highlight that social signalling is likely involved in face processing, pointing to a far more extensive network than previously suggested. However, there is currently a lack of evidence on the neural mechanisms underlying naturalistic social interactions (the "dark matter" of social neuroscience; Schilbach et al., 2013, p.394), likely due to the methodological challenges associated with neuroimaging techniques. For instance, whereas participants are required to remain completely still during measurements, live social interactions typically involve movement in form of, e.g., a speaking mouth or facial expressions. While very few neuroimaging studies have studied social interactions by, for instance, presenting live video recordings in an fMRI scanner, participants were not actively engaged in a social interaction (Redcay et al., 2010). More advanced methods will therefore be required to study the neural basis of face-to-face social interactions, which in turn could inform neural models of face processing.

Another factor highlighted in this thesis concerns the cultural modulations on face processing. Existing neuroimaging studies on face perception have typically been conducted in Western societies. However, studies have previously identified cultural differences in the neural basis of theory-of-mind (Kobayashi, Glover, & Temple, 2007), processing of fearful faces (Chiao et al., 2008), or object processing (Gutchess, Welsh, Boduroĝlu, & Park, 2006), suggesting cultural influences also at a neural level. By examining the neural basis of cultural differences in face perception, future studies could identify the extent to which neural substrates or mechanisms involved in specialised face processing are culture-sensitive (i.e., susceptible to postnatal influences) and culture-invariant (i.e., universal; Han & Northoff, 2008).

7.5. Conclusion

For the first time, the findings of this thesis revealed cultural influences on adults' face scanning strategies in both naturalistic and screen-based paradigms. The cross-sectional screen-based study (Chapter 5) also suggested that cultural modulations on face processing likely emerge during the first year of life. Although cultural differences were not observed in the naturalistic infant face-to-face interaction study (Chapter 6), it is possible that other contextual factors were significantly more relevant for face processing during the specific paradigm such that cultural influences were not observed. The precise manifestation of cultural differences and the extent to which these are evident depend on various other situational aspects including social presence, speech, experimental task, or stimulus characteristics – and individual differences likely exist with respect to the relative contribution of these factors. Altogether, this points to a highly adaptive face processing system that is shaped by postnatal social experience and modulated by contextual factors.

References

Adler, R. J. (1981). *The Geometry of Random Fields*. New York: Wiley.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68–72.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the Enigmatic Face: The Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. *Psychological Science*, *16*, 403–410.

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*, 640–654.

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

Argyle, M., Henderson, M., Bond, M., Iizuka, Y., & Contarello, A. (1986). Cross-Cultural Variations in Relationship Rules. *International Journal of Psychology*, *21*, 287–315.

Arizpe, J., Kravitz, D. J., Walsh, V., Yovel, G., & Baker, C. I. (2016). Differences in Looking at Own- and Other-Race Faces Are Subtle and Analysis-Dependent: An Account of Discrepant Reports. *PLoS ONE*, *11*, e0148253.

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start Position Strongly Influences Fixation Patterns during Face Processing: Difficulties with Eye Movements as a Measure of Information Use. *PLoS ONE*, *7*, e31106.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*, 48–53.

Aslin, R. N., & McMurray, B. (2004). Automated Corneal-Reflection Eye Tracking in Infancy: Methodological Developments and Applications to Cognition. *Infancy*, *6*, 155–163.

Aylward, E. H., Park, J. E., Field, K. M., Parsons, A. C., Richards, T. L., Cramer, S. C., & Meltzoff, A. N. (2005). Brain Activation during Face Perception: Evidence of a Developmental Change. *Journal of Cognitive Neuroscience*, *17*, 308–319.

Bahrick, L. E., Netto, D., & Hern, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*, 1263–1275.

Ballieux, H., Wass, S. V., Tomalski, P., Kushnerenko, E., Karmiloff-Smith, A., Johnson, M. H., & Moore, D. G. (2016). Applying gaze-contingent training within community settings to infants from diverse SES backgrounds. *Journal of Applied Developmental Psychology*, *43*, 8–17.

Bambach, S., Franchak, J. M., Crandall, D. J., & Yu, C. (2014). Detecting Hands in Children's Egocentric Views to Understand Embodied Attention during Social Interaction. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*, 134–139.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*, 551–565.

Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience & Biobehavioral Reviews*, *55*, 536–546.

Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & de Haan, E. H. F. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, *12*, 1048–1053.

Bindemann, M., & Lewis, M. B. (2013). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin & Review*, *20*, 1140–1145.

Binetti, N., Harrison, C., Coutrot, A., Johnston, A., & Mareschal, I. (2016). Pupil dilation as an index of preferred mutual gaze duration. *Royal Society Open Science*, *3*, 160086.

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion*, *17*, 1107–1119.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture Shapes How We Look at Faces. *PLoS ONE*, *3*, e3022.

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods*, *46*, 67–80.

Bornstein, M. H., Tamis-Lemonda, C. S., Tal, J., Ludemann, P., Toda, S., Rahn, C. W., … Vardi, D. (1992). Maternal Responsiveness to Infants in Three Societies: The United States, France, and Japan. *Child Development*, *63*, 808–821.

Bradski, G. R. (1998). Computer Vision Face Tracking For Use in a Perceptual User Interface.

Brett, M., Penny, W., & Kiebel, S. (2003). An Introduction to Random Field Theory. *Human Brain Function*, *2*, 1–23.

Brielmann, A. A., Bülthoff, I., & Armann, R. (2014). Looking at faces from different angles: Europeans fixate different features in Asian and Caucasian faces. *Vision Research*, *100*, 105–112.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. New York, NY, US: Oxford University Press.

Caldara, R. (2017). Culture Reveals a Flexible System for Face Processing. *Current Directions in Psychological Science*, *26*, 249–255.

Caldara, R., & Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods, 43*, 864–878.

Caldara, R., Richoz, A.-R., Liu, Y., & Lao, J. (2016). *Cultural Diversity in Eye Movements is Shaped by Nurture not Nature.* Talk presented at the Vision Sciences Society Conference, St Pete Beach, Florida, USA.

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting Culture Under the 'Spotlight' Reveals Universal Information Use for Face Recognition. *PLoS ONE, 5*, e9708.

Chawarska, K., & Shic, F. (2009). Looking But Not Seeing: Atypical Visual Scanning and Recognition of Faces in 2 and 4-Year-Old Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 39*, 1663.

Chen, C., Burton, M., Greenberger, E., & Dmitrieva, J. (1999). Population Migration and the Variation of Dopamine D4 Receptor (DRD4) Allele Frequencies Around the Globe. *Evolution and Human Behavior, 20*, 309–324.

Chen, F. S., & Yoon, J. M. D. (2011). Brief Report: Broader Autism Phenotype Predicts Spontaneous Reciprocity of Direct Gaze. *Journal of Autism and Developmental Disorders*, *41*, 1131–1134.

Chiao, J. Y., & Blizinsky, K. D. (2010). Culture–gene coevolution of individualism–collectivism and the serotonin transporter gene. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 529–537.

Chiao, J. Y., Iidaka, T., Gordon, H. L., Nogawa, J., Bar, M., Aminoff, E., … Ambady, N. (2008). Cultural specificity in amygdala response to fear faces. *Journal of Cognitive Neuroscience*, *20*, 2167–2174.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 12629–12633.

Cohen Kadosh, K. (2011). What can emerging cortical face networks tell us about mature brain organisation? *Developmental Cognitive Neuroscience*, *1*, 246–255.

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology*, *52*, 337–367.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development* (pp. 249–274). Oxford: Oxford University Press.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., … Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*, 519–526.

de Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of Neural Mechanisms Underlying Face Recognition in Human Infants. *Journal of Cognitive Neuroscience*, *14*, 199–209.

Delabarre, E. B. (1898). A Method of Recording Eye-Movements. *The American Journal of Psychology*, *9*, 572–574.

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: An eye-tracking study with 3- and 6-month-old infants and adults. *Journal of Experimental Child Psychology*, *113*, 66–77.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development*, *56*, 868–883.

Diamond, Adele. (2002). Normal Development of Prefrontal Cortex from Birth to Young Adulthood: Cognitive Functions, Anatomy, and Biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 466–503). New York: Oxford University Press.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *11*, 478–484.

Doherty-Sneddon, G., & Phelps, F. G. (2005). Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition*, *33*, 727–733.

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice*. Dordrecht: Springer.

Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *NeuroReport*, *11*, 2319–2324.

Elsabbagh, M., Fernandes, J., Jane Webb, S., Dawson, G., Charman, T., & Johnson, M. H. (2013). Disengagement of Visual Attention in Infancy is Associated with Emerging Autism in Toddlerhood. *Biological Psychiatry*, *74*, 189–194.

Elsabbagh, M., Volein, A., Holmboe, K., Tucker, L., Csibra, G., Baron-Cohen, S., … Johnson, M. H. (2009). Visual orienting in the early broader autism phenotype: disengagement and facilitation. *Journal of Child Psychology and Psychiatry*, *50*, 637–642.

Falck-Ytter, T. (2015). Gaze performance during face-to-face communication: A live eye tracking study of typical children and children with autism. *Research in Autism Spectrum Disorders*, *17*, 78–85.

Falck-Ytter, T., Carlström, C., & Johansson, M. (2015). Eye Contact Modulates Cognitive Processing Differently in Children With Autism. *Child Development*, *86*, 37–47.

Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioural Brain Research*, *76*, 181–189.

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early Commitment of Neural Substrates for Face Recognition. *Cognitive Neuropsychology*, *17*, 117–123.

Farah, M. J., Wilson, K. D., Maxwell Drain, H., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, *35*, 2089–2093.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object Detection with Discriminatively Trained Part Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*, 1627–1645.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*, 1–185.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual*. Baltimore, MD: Brookes.

Fogel, A., Toda, S., & Kawai, M. (1988). Mother-infant face-to-face interaction in Japan and the United States: A laboratory comparison using 3-month-old infants. *Developmental Psychology*, *24*, 398–406.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*, 1920–1931.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye-tracking: A new method to describe infant looking. *Child Development*, *82*, 1738–1750.

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces in early infancy. *Journal of Experimental Child Psychology*, *118*, 13–26.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the Development of Social Attention Using Free-Viewing. *Infancy*, *17*, 355–375.

Freeth, M., & Bugembe, P. (2018). Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults? *Autism*. doi:10.1177/1362361318756786

Freeth, M., Bullock, T., & Milne, E. (2013). The distribution of and relationship between autistic traits and social anxiety in a UK student population. *Autism*, *17*, 571–581.

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLoS ONE*, *8*, e53286.

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, *55*, 119–139.

Fu, G., Hu, C. S., Wang, Q., Quinn, P. C., & Lee, K. (2012). Adults Scan Own- and Other-Race Faces Differently. *PLoS ONE*, *7*, e37688.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*, 568–573.

Geangu, E., Ichikawa, H., Lao, J., Kanazawa, S., Yamaguchi, M. K., Caldara, R., & Turati, C. (2016). Culture shapes 7-month-olds' perceptual strategies in discriminating facial expressions of emotion. *Current Biology*, *26*, R663–R664.

Gintis, H. (2011). Gene–culture coevolution and the nature of human sociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 878–888.

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*, 651–658.

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

Gobel, M. S., Chen, A., & Richardson, D. C. (2017). How different cultures look at faces depends on the interpersonal context. *Canadian Journal of Experimental Psychology*, *71*, 258–264.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364.

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in Cross-Race Face Learning: Insights From Eye Movements and Pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1105–1122.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261–2271.

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye Tracking in Infancy Research. *Developmental Neuropsychology*, *35*, 1–19.

Greenfield, P. M., Keller, H., Fuligni, A., & Maynard, A. (2003). Cultural Pathways Through Universal Development. *Annual Review of Psychology*, *54*, 461–490.

Gutchess, A. H., Welsh, R. C., Boduroĝlu, A., & Park, D. C. (2006). Cultural differences in neural function associated with object processing. *Cognitive, Affective, & Behavioral Neuroscience*, *6*, 102–109.

Hall, J., Harris, J. M., Sprengelmeyer, R., Sprengelmeyer, A., Young, A. W., Santos, I. M., … Lawrie, S. M. (2004). Social cognition and face processing in schizophrenia. *The British Journal of Psychiatry*, *185*, 169–170.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34.

Han, S., & Northoff, G. (2008). Culture-sensitive neural substrates of human cognition: a transcultural neuroimaging approach. *Nature Reviews Neuroscience*, *9*, 646–654.

Hauck, M., Fein, D., Maltby, N., Waterhouse, L., & Feinstein, C. (1998). Memory for Faces in Children with Autism. *Child Neuropsychology*, *4*, 187–198.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*, 1740–1751.

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., & Liebowitz, M. R. (1999). Psychometric properties of the Liebowitz Social Anxiety Scale. *Psychological Medicine*, *29*, 199–212.

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

Hendry, A., Jones, E. J. H., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review*, *42*, 1–33.

Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Canadian Journal of Experimental Psychology*, *71*, 226–242.

Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2018). Eye contact takes two – autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*, *9*, 1–17.

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science*, *5*, 180502.

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2018). The redeployment of attention to the mouth of a talking face during the second year of life. *Journal of Experimental Child Psychology*, *172*, 189–200.

Hirokawa, K., Yagi, A., & Miyata, Y. (2004). Comparison of Blinking Behavior during Listening to and Speaking in Japanese and English. *Perceptual and Motor Skills*, *98*, 463–472.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and Listening with the Eyes: Gaze Signaling during Dyadic Interactions. *PLoS ONE*, *10*, e0136905.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*, 80–84.

Holland, M. K., & Tarlow, G. (1975). Blinking and Thinking. *Perceptual and Motor Skills*, *41*, 403–406.

Holmboe, K., Bonneville-Roussy, A., Csibra, G., & Johnson, M. H. (2018). Longitudinal development of attention and inhibitory control during the first year of life. *Developmental Science*, *21*, e12690.

Holmboe, K., Pasco Fearon, R. M., Csibra, G., Tucker, L. A., & Johnson, M. H. (2008). Freeze-Frame: A new infant inhibition task and its relation to frontal cortex tasks during infancy and early childhood. *Journal of Experimental Child Psychology*, *100*, 89–114.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Hömke, P., Holler, J., & Levinson, S. C. (2017). Eye Blinking as Addressee Feedback in Face-To-Face Conversation. *Research on Language and Social Interaction*, *50*, 54–70.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, *16*, 405–422.

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not see eye to eye: A visual scanpath study of emotional expression processing. *Anxiety Disorders*, *17*, 33–44.

Imada, T., Carlson, S. M., & Itakura, S. (2013). East–West cultural differences in context-sensitivity are evident in early childhood. *Developmental Science*, *16*, 198–208.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*, 1093–1123.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural Confusions Show that Facial Expressions Are Not Universal. *Current Biology*, *19*, 1543–1548.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*, 19–25.

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

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, *17*, 4302–4311.

Kelly, D. J., Jack, R. E., Miellet, S., De Luca, E., Foreman, K., & Caldara, R. (2011). Social Experience Does Not Abolish Cultural Diversity in Eye Movements. *Frontiers in Psychology*, *2*, 1–11.

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*, 105–114.

Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science*, *14*, 1176–1184.

Kelly, D. J., Miellet, S., Caldara, R., Kelly, D. J., Miellet, S., & Caldara, R. (2010). Culture shapes eye movements for visually homogeneous objects. *Perception Science, 1*, 6.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy. *Psychological Science, 18*, 1084–1089.

Kievit, R., Frankenhuis, W. E., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*, 1–14.

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable Neural Pathways Are Involved in the Recognition of Emotion in Static and Dynamic Facial Expressions. *NeuroImage*, *18*, 156–168.

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology*, *19,* 52–56.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99,* 317–340.

Kita, Y., Gunji, A., Sakihara, K., Inagaki, M., Kaga, M., Nakagawa, E., & Hosokawa, T. (2010). Scanning Strategies Do Not Modulate Face Identification: Eye-Tracking and Near-Infrared Spectroscopy Study. *PLoS ONE, 5*, e11050.

Kitayama, S., King, A., Yoon, C., Tompson, S., Huff, S., & Liberzon, I. (2014). The Dopamine D4 Receptor Gene (DRD4) Moderates Cultural Difference in Independent Versus Interdependent Social Orientation. *Psychological Science*, *25*, 1169–1177.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*, 78–100.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism. *Archives of General Psychiatry*, *59*, 809–816.

Kobayashi, C., Glover, G. H., & Temple, E. (2007). Cultural and linguistic effects on neural bases of 'Theory of Mind' in American and Japanese children. *Brain Research*, *1164*, 95–107.

Kovacs, A. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, *106*, 6556–6560.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*, 1503–1518.

Król, M. E. (2018). Auditory noise increases the allocation of attention to the mouth, and the eyes pay the price: An eye-tracking study. *PLoS ONE*, *13*, e0194491.

Kurita, H., Koyama, T., & Osada, H. (2005). Autism-Spectrum Quotient–Japanese version and its short forms for screening normally intelligent persons with pervasive developmental disorders. *Psychiatry and Clinical Neurosciences*, *59*, 490–496.

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, *108*, 5548–5553.

Lan, X., Legare, C. H., Ponitz, C. C., Li, S., & Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: A cross-cultural analysis of Chinese and American preschoolers. *Journal of Experimental Child Psychology*, *108*, 677–692.

Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*, 330–342.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience*, *6*, 1108–1112.

Le, T. T., Farkas, L. G., Ngim, R. C. K., Levin, L. S., & Forrest, C. R. (2002). Proportionality in Asian and North American Caucasian Faces Using Neoclassical Facial Canons as Criteria. *Aesthetic Plastic Surgery*, *26*, 64–69.

Lewis, M., & Edmonds, A. (2005). Searching for faces in scrambled scenes. *Visual Cognition*, *12*, 1309–1336.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*, 1431–1436.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–173.

Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same- and other-race faces as revealed by eye tracking in 4- to 9-month-olds. *Journal of Experimental Child Psychology*, *108*, 180–189.

Lucas, B. D., & Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 674–679.

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

Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology*. London: Chapman & Hall.

Markus, H. R., & Kitayama, S. (1991). Culture and the Self: Implications for Cognition, Emotion, and Motivation. *Psychological Review*, *98*, 224–253.

Marwick, K., & Hall, J. (2008). Social cognition in schizophrenia: a review of face processing. *British Medical Bulletin*, *88*, 43–58.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*, 922–934.

Masuda, T., & Nisbett, R. E. (2006). Culture and Change Blindness. *Cognitive Science*, *30*, 381–399.

McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2006). Cultural Display Rules Drive Eye Gaze During Thinking. *Journal of Cross-Cultural Psychology*, *37*, 717–722.

McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2008). Gaze Display When Thinking Depends on Culture and Context. *Journal of Cross-Cultural Psychology*, *39*, 716–729.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-Specific Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience*, *9*, 605–610.

McCleery, A., Lee, J., Joshi, A., Wynn, J. K., Hellemann, G. S., & Green, M. F. (2015). Meta-Analysis of Face Processing Event-Related Potentials in Schizophrenia. *Biological Psychiatry*, *77*, 116–126.

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*, 3–35.

Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When East meets West: gaze-contingent Blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research*, *5*, 1–12.

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping Face Recognition Information Use across Cultures. *Frontiers in Psychology*, *4*, 1–12.

Miellet, S., Zhou, X., He, L., Rodger, H., & Caldara, R. (2010). Investigating cultural diversity for extrafoveal information use in visual scenes. *Journal of Vision*, *10*, 21–21.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2010). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation*, *3*, 5–24.

Miyamoto, Y., Nisbett, R. E., & Masuda, T. (2006). Culture and the Physical Environment Holistic Versus Analytic Perceptual Affordances. *Psychological Science*, *17*, 113–119.

Moriguchi, Y. (2014). The early development of executive function and its relation to social interaction: a brief review. *Frontiers in Psychology*, *5*, 1–6.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A Two-Process Theory of Infant Face Recognition. *Psychological Review*, *98*, 164–181.

Moukheiber, A., Rautureau, G., Perez-Diaz, F., Soussignan, R., Dubal, S., Jouvent, R., & Pelissolo, A. (2010). Gaze avoidance in social phobia: Objective measure and correlates. *Behaviour Research and Therapy*, *48*, 147–151.

Myllyneva, A., & Hietanen, J. K. (2015). There is more to eye contact than meets the eye. *Cognition*, *134*, 100–109.

Nasiopoulos, E., Risko, E. F., Foulsham, T., & Kingstone, A. (2015). Wearable computing: Will it make people prosocial? *British Journal of Psychology, 106*, 209–216.

Neesgaard, N., Senju, A., & Smith, T. J. (2017). Quantifying the microdynamics of attention during parent-child interaction: practicalities and insights. Presented at the European Conference on Eye Movements.

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, *10*, 3–18.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*, 1–25.

Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences*, *100*, 11163–11170.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*, 467–473.

Nyström, P., Bölte, S., Falck-Ytter, T., Achermann, S., Andersson Konke, L., Brocki, K., … The EASE Team. (2017). Responding to Other People's Direct Gaze: Alterations in Gaze Behavior in Infants at Risk for Autism Occur on Very Short Timescales. *Journal of Autism and Developmental Disorders*, *47*, 3498–3509.

Oakes, L. M., & Ellis, A. E. (2013). An eye-tracking investigation of developmental changes in infants' exploration of upright and inverted human faces. *Infancy*, *18*, 134–148.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with Faces Selectively Enhances Sensitivity to Changes Made to the Eyes. *Perception*, *30*, 755–764.

Oh, S., & Lewis, C. (2008). Korean Preschoolers' Advanced Inhibitory Control and Its Relation to Other Executive Skills and Mental State Understanding. *Child Development*, *79*, 80–99.

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. New York: Psychology Press.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*, 1–27.

Or, C. C.-F., Peterson, M. F., & Eckstein, M. P. (2015). Initial eye movements during face identification are optimal and similar across cultures. *Journal of Vision*, *15*, 1–25.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*, 261–266.

Ousler, G. W., Hagberg, K. W., Schnindelar, M., Welch, D., & Abelson, M. B. (2008). The Ocular Protection Index. *Cornea*, *27*, 509–513.

Peelen, M. V., Glaser, B., Vuilleumier, P., & Eliez, S. (2009). Differential development of selectivity for faces and bodies in the fusiform gyrus. *Developmental Science*, *12*, F16–F25.

Pierce, K., Müller, R.-A., Ambrose, J., Allen, G., & Courchesne, E. (2001). Face processing occurs outside the fusiform `face area' in autism: evidence from functional MRI. *Brain*, *124*, 2059–2073.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*, *26*, 490–498.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, *206*, 468–469.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye Movements When Looking at Unusual/Weird Scenes: Are There Cultural Differences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 254–259.

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, *47*, 2714–2726.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage*, *50*, 1639–1647.

Rhodes, G. (1985). Lateralized processes in face recognition. *British Journal of Psychology*, *76*, 249–271.

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

Riby, D. M., Hancock, P. J., Jones, N., & Hanley, M. (2013). Spontaneous and cued gaze-following in autism and Williams syndrome. *Journal of Neurodevelopmental Disorders*, *5*, 13.

Richardson, D. C., & Gobel, M. S. (2015). Social Attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *Handbook of Attention* (pp. 46–61). Cambridge, MA: MIT Press.

Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, *73*, 291–296.

Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6*, 1–11.

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science: Real-World Social Attention and the Dual Function of Gaze. *Current Directions in Psychological Science*, *25*, 70–74.

Ro, T., Russell, C., & Lavie, N. (2001). Changing Faces: A Detection Advantage in the Flicker Paradigm. *Psychological Science*, *12*, 94–99.

Rodger, H., Kelly, D. J., Blais, C., & Caldara, R. (2010). Inverting Faces Does Not Abolish Cultural Diversity in Eye Movements. *Perception*, *39*, 1491–1503.

Rogers, S. L., Speelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific Reports*, *8*, 4271.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of Infant Cognition for Executive Functions at Age 11. *Psychological Science*, *23*, 1345–1355.

Rossion, B., Dricot, L., Devolder, A., Bodart, J.-M., Crommelinck, M., Gelder, B. de, & Zoontjes, R. (2000). Hemispheric Asymmetries for Whole-Based and Part-Based Face Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience*, *12*, 793–802.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The Development of Executive Functioning and Theory of Mind. *Psychological Science*, *17*, 74–81.

Saez de Urabain, I. R., Johnson, M. H., & Smith, T. J. (2015). GraFIX: A semiautomatic approach for parsing low- and high-quality eye-tracking data. *Behavior Research Methods*, *47*, 53–72.

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*, 393–414.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show Me the Features! Understanding Recognition From the Use of Visual Information. *Psychological Science*, *13*, 402–409.

Sekiyama, K., Tohkura, Y., & Umeda, M. (1996). A few factors which affect the degree of incorporating lip-read information into speech perception. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96* (Vol. 3, pp. 1481–1484). Philadelphia, PA, USA: IEEE.

Sekiyama, Kaoru, & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, *90*, 1797–1805.

Sekiyama, Kaoru, & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*, 427–444.

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*, 668–671.

Senju, A., Vernetti, A., Ganea, N., Hudry, K., Tucker, L., Charman, T., & Johnson, M. H. (2015). Early Social Experience Affects the Development of Eye Gaze Processing. *Current Biology*, *25*, 3086–3091.

Senju, A., Vernetti, A., Kikuchi, Y., Akechi, H., & Hasegawa, T. (2013). Cultural Modulation of Face and Gaze Scanning in Young Children. *PLoS ONE*, *8*, e74017.

Senju, A., Vernetti, A., Kikuchi, Y., Akechi, H., Hasegawa, T., & Johnson, M. H. (2013). Cultural background modulates how we look at other persons' gaze. *International Journal of Behavioral Development*, *37*, 131–136.

Shen, J., & Itti, L. (2012). Top-down influences on visual attention during listening are modulated by observer sex. *Vision Research*, *65*, 62–76.

Shic, F., Macari, S., & Chawarska, K. (2014). Speech Disturbs Face Scanning in 6-Month Olds who Develop Autism Spectrum Disorder. *Biological Psychiatry*, *75*, 231–237.

Shukla, M., Wen, J., White, K. S., & Aslin, R. N. (2011). SMART-T: A system for novel fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods*, *43*, 384–398.

Smith, A. M., Lewis, B. K., Ruttimann, U. E., Ye, F. Q., Sinnwell, T. M., Yang, Y., … Frank, J. A. (1999). Investigation of Low Frequency Drift in fMRI Signal. *NeuroImage*, *9*, 526–533.

Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*, 83–98.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*, 16–16.

Stewart, M., Anderson, J., Beiser, M., Mwakarimba, E., Neufeld, A., Simich, L., & Spitzer, D. (2008). Multicultural Meanings of Social Support among Immigrants and Refugees. *International Migration*, *46*, 123–159.

Sue, D. W., & Sue, D. (1990). *Counseling the culturally different: Theory and Practice*. Oxford, England: John Wiley & Sons.

Sugawara, N., Yasui-Furukori, N., Kaneda, A., Sato, Y., Tsuchimine, S., Fujii, A., … Kaneko, S. (2012). Factor structure of the Liebowitz Social Anxiety Scale in community-dwelling subjects in Japan. *Psychiatry and Clinical Neurosciences*, *66*, 525–528.

Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013). Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. *Infancy*, *18*, 534–553.

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*, 1173–1190.

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*, 657–665.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: a comparison between young and older adults. *Psychology and Aging*, *10*, 215–228.

Thompson, L. A., & Malloy, D. (2004). Attention Resources and Visible Speech Encoding in Older and Younger Adults. *Experimental Aging Research, 30*, 241–252.

Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., Falck-Ytter, T., & The EASE Team. (2016). Altered gaze following during live interaction in infants at risk for autism: an eye tracking study. *Molecular Autism*, *7*, 12.

Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., Falck-Ytter, T., & the EASE Team. (2018). Reduced Alternating Gaze During Social Interaction in Infancy is

Associated with Elevated Symptoms of Autism in Toddlerhood. *Journal of Abnormal Child Psychology, 46*, 1547–1561.

Toda, S., Fogel, A., & Kawai, M. (1990). Maternal speech to three-month-old infants in the United States and Japan. *Journal of Child Language, 17*, 279–294.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495–552.

Tomasi, C., & Kanade, T. (1991). Detection and Tracking of Point Features.

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109.

Vabalas, A., & Freeth, M. (2016). Brief Report: Patterns of Eye Movements in Face to Face Conversation are Associated with Autistic Traits: Evidence from a Student Sample. *Journal of Autism and Developmental Disorders*, *46*, 305–314.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*, 471–491.

Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, *60*, 926–940.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Vol. 1, pp. I-511-I–518). Kauai, HI, USA: IEEE Comput. Soc.

Viola, P., & Jones, M. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision*, *57*, 137–154.

Wagener, D. W., & Herbst, B. (2001). Face Tracking : An implementation of the Kanade-Lucas-Tomasi Tracking algorithm. *PRASA, South Africa*.

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., & Tojo, Y. (2006). The Autism-Spectrum Quotient (AQ) in Japan: A Cross-Cultural Comparison. *Journal of Autism and Developmental Disorders*, *36*, 263–270.

Wang, Q., Xiao, N. G., Quinn, P. C., Hu, C. S., Qian, M., Fu, G., & Lee, K. (2015). Visual scanning and recognition of Chinese, Caucasian, and racially ambiguous faces: Contributions from bottom-up facial physiognomic information and top-down knowledge of racial categories. *Vision Research*, *107*, 67–75.

Wass, S. V., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training Attentional Control in Infancy. *Current Biology*, *21*, 1543–1547.

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*, 229–250.

Wass, S. V., & Smith, T. J. (2014). Individual Differences in Infant Oculomotor Behavior During the Viewing of Complex Naturalistic Scenes. *Infancy*, *19*, 352–384.

Way, B. M., & Lieberman, M. D. (2010). Is there a genetic contribution to cultural differences? Collectivism, individualism and genetic markers of social sensitivity. *Social Cognitive and Affective Neuroscience*, *5*, 203–211.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, *36*, 1060–1084.

Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian Infants Scan Own- and Other-Race Faces Differently. *PLoS ONE*, *6*, e18621.

Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety? An eye-tracking and psychophysiology study. *Journal of Anxiety Disorders*, *23*, 93–103.

Wilcox, T., Stubbs, J. A., Wheeler, L., & Alexander, G. M. (2013). Infants' scanning of dynamic faces during the first year. *Infant Behavior and Development, 36*, 513–516.

Williams, C. C., & Henderson, J. M. (2007). The face inversion effect is not a consequence of aberrant eye movements. *Memory & Cognition, 35*, 1977–1985.

Wörmann, V., Holodynski, M., Kärtner, J., & Keller, H. (2012). A cross-cultural comparison of the development of the social smile: a longitudinal study of maternal and infant imitation in 6- and 12-week-old infants. *Infant Behavior and Development, 35*, 335–347.

Worsley, K. J., Marrett, S., Neelin, P., Friston, K. J., & Evans, A. C. (1996). A Unified Statistical Approach for Determining Significant Signals in Images of Cerebral Activation. *Human Brain Mapping*, *4,* 58–73.

Wu, D. W.-L., Bischof, W. F., & Kingstone, A. (2013). Looking while eating: The importance of social context to social attention. *Scientific Reports, 3,* 2356.

Wu, D. W.-L., Bischof, W. F., & Kingstone, A. (2014). Natural gaze signaling in a social context. *Evolution and Human Behavior, 35,* 211–218.

Wu, E. X. W., Laeng, B., & Magnussen, S. (2012). Through the eyes of the own-race bias: Eye-tracking and pupillometry during face recognition. *Social Neuroscience*, *7*, 202–216.

Xiao, W. S., Xiao, N. G., Quinn, P. C., Anzures, G., & Lee, K. (2013). Development of face scanning for own- and other-race faces in infancy. *International Journal of Behavioral Development, 37*, 100–105.

Yang, G., & Huang, T. S. (1994). Human face detection in a complex background. *Pattern Recognition*, *27*, 53–63.

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In A. L. Yarbus (Ed.), *Eye Movements and Vision* (pp. 171–211). Boston, MA: Springer US.

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE*, *8*, e79659.

Appendix A

Autism Quotient test used in the studies presented in Chapter 3 and Chapter 4.

The Adult Autism Spectrum Quotient (AQ)

Ages 16+

How to fill out the questionnaire

Below is a list of statements. Please read each statement very carefully and rate how strongly you agree or disagree with it by circling your answer.

DO NOT MISS ANY STATEMENT OUT.

Examples

E1. I am willing to take risks.	definitely agree	slightly agree	slightly disagree	definitely disagree
E2. I like playing board games.	definitely agree	slightly agree	slightly disagree	definitely disagree
E3. I find learning to play musical instruments easy.	definitely agree	slightly agree	slightly disagree	definitely disagree
E4. I am fascinated by other cultures.	definitely agree	slightly agree	slightly disagree	definitely disagree

1. I prefer to do things with others rather than on my own.	definitely agree	slightly agree	slightly disagree	definitely disagree
2. I prefer to do things the same way over and over again.	definitely agree	slightly agree	slightly disagree	definitely disagree
3. If I try to imagine something, I find it very easy to create a picture in my mind.	definitely agree	slightly agree	slightly disagree	definitely disagree
4. I frequently get so strongly absorbed in one thing that I lose sight of other things.	definitely agree	slightly agree	slightly disagree	definitely disagree
5. I often notice small sounds when others do not.	definitely agree	slightly agree	slightly disagree	definitely disagree
6. I usually notice car number plates or similar strings of information.	definitely agree	slightly agree	slightly disagree	definitely disagree
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.	definitely agree	slightly agree	slightly disagree	definitely disagree
8. When I'm reading a story, I can easily imagine what the characters might look like.	definitely agree	slightly agree	slightly disagree	definitely disagree
9. I am fascinated by dates.	definitely agree	slightly agree	slightly disagree	definitely disagree
10. In a social group, I can easily keep track of several different people's conversations.	definitely agree	slightly agree	slightly disagree	definitely disagree
11. I find social situations easy.	definitely agree	slightly agree	slightly disagree	definitely disagree
12. I tend to notice details that others do not.	definitely agree	slightly agree	slightly disagree	definitely disagree
13. I would rather go to a library than a party.	definitely agree	slightly agree	slightly disagree	definitely disagree
14. I find making up stories easy.	definitely agree	slightly agree	slightly disagree	definitely disagree
15. I find myself drawn more strongly to people than to things.	definitely agree	slightly agree	slightly disagree	definitely disagree
16. I tend to have very strong interests which I get upset about if I can't pursue.	definitely agree	slightly agree	slightly disagree	definitely disagree
17. I enjoy social chit-chat.	definitely agree	slightly agree	slightly disagree	definitely disagree
18. When I talk, it isn't always easy for others to get a word in edgeways.	definitely agree	slightly agree	slightly disagree	definitely disagree
19. I am fascinated by numbers.	definitely agree	slightly agree	slightly disagree	definitely disagree

20. When I'm reading a story, I find it difficult to work out the characters' intentions.	definitely agree	slightly agree	slightly disagree	definitely disagree
21. I don't particularly enjoy reading fiction.	definitely agree	slightly agree	slightly disagree	definitely disagree
22. I find it hard to make new friends.	definitely agree	slightly agree	slightly disagree	definitely disagree
23. I notice patterns in things all the time.	definitely agree	slightly agree	slightly disagree	definitely disagree
24. I would rather go to the theatre than a museum.	definitely agree	slightly agree	slightly disagree	definitely disagree
25. It does not upset me if my daily routine is disturbed.	definitely agree	slightly agree	slightly disagree	definitely disagree
26. I frequently find that I don't know how to keep a conversation going.	definitely agree	slightly agree	slightly disagree	definitely disagree
27. I find it easy to "read between the lines" when someone is talking to me.	definitely agree	slightly agree	slightly disagree	definitely disagree
28. I usually concentrate more on the whole picture, rather than the small details.	definitely agree	slightly agree	slightly disagree	definitely disagree
29. I am not very good at remembering phone numbers.	definitely agree	slightly agree	slightly disagree	definitely disagree
30. I don't usually notice small changes in a situation, or a person's appearance.	definitely agree	slightly agree	slightly disagree	definitely disagree
31. I know how to tell if someone listening to me is getting bored.	definitely agree	slightly agree	slightly disagree	definitely disagree
32. I find it easy to do more than one thing at once.	definitely agree	slightly agree	slightly disagree	definitely disagree
33. When I talk on the phone, I'm not sure when it's my turn to speak.	definitely agree	slightly agree	slightly disagree	definitely disagree
34. I enjoy doing things spontaneously.	definitely agree	slightly agree	slightly disagree	definitely disagree
35. I am often the last to understand the point of a joke.	definitely agree	slightly agree	slightly disagree	definitely disagree
36. I find it easy to work out what someone is thinking or feeling just by looking at their face.	definitely agree	slightly agree	slightly disagree	definitely disagree
37. If there is an interruption, I can switch back to what I was doing very quickly.	definitely agree	slightly agree	slightly disagree	definitely disagree
38. I am good at social chit-chat.	definitely agree	slightly agree	slightly disagree	definitely disagree

39. People often tell me that I keep going on and on about the same thing.	definitely agree	slightly agree	slightly disagree	definitely disagree
40. When I was young, I used to enjoy playing games involving pretending with other children.	definitely agree	slightly agree	slightly disagree	definitely disagree
41. I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.).	definitely agree	slightly agree	slightly disagree	definitely disagree
42. I find it difficult to imagine what it would be like to be someone else.	definitely agree	slightly agree	slightly disagree	definitely disagree
43. I like to plan any activities I participate in carefully.	definitely agree	slightly agree	slightly disagree	definitely disagree
44. I enjoy social occasions.	definitely agree	slightly agree	slightly disagree	definitely disagree
45. I find it difficult to work out people's intentions.	definitely agree	slightly agree	slightly disagree	definitely disagree
46. New situations make me anxious.	definitely agree	slightly agree	slightly disagree	definitely disagree
47. I enjoy meeting new people.	definitely agree	slightly agree	slightly disagree	definitely disagree
48. I am a good diplomat.	definitely agree	slightly agree	slightly disagree	definitely disagree
49. I am not very good at remembering people's date of birth.	definitely agree	slightly agree	slightly disagree	definitely disagree
50. I find it very easy to play games with children that involve pretending.	definitely agree	slightly agree	slightly disagree	definitely disagree

Appendix B

Liebowitz Social Anxiety Scale used in the study presented in Chapter 4.

This measure assesses the way that social phobia plays a role in your life across a variety of situations. Read each situation carefully and answer two questions about that situation. The first question asks how anxious or fearful you feel in the situation. The second question asks how often you avoid the situation. If you come across a situation that you ordinarily do not experience, we ask that you imagine "what if you were faced with that situation," and then, rate the degree to which you would fear this hypothetical situation and how often you would tend to avoid it. Please base your ratings on the way that the situations have affected you in the last week. Fill out the following scale with the most suitable answer provided below.

Fear or Anxiety:	Avoidance:
0 = None	0 = Never (0%)
1 = Mild	1 = Occasionally (1-33%)
2 = Moderate	2 = Often (33-67%)
3 = Severe	3 = Severe (67-100%)

Item	Fear or Anxiety	Avoidance
1. Telephoning in public.		
2. Participating in small groups.		
3. Eating in public places.		
4. Drinking with others in public places.		
5. Talking to people in authority.		
6. Acting, performing or giving a talk in front of an audience.		
7. Going to a party.		
8. Working while being observed.		
9. Writing while being observed.		
10. Calling someone you don't know very well.		
11. Talking with people you don't know very well.		
12. Meeting strangers.		
13. Urinating in a public bathroom.		
14. Entering a room when others are already seated.		
15. Being the centre of attention.		
16. Speaking up at a meeting.		
17. Taking a test.		
18. Expressing a disagreement or disapproval to people you don't know very well.		
19. Looking at people you don't know very well in the eyes.		
20. Giving a report to a group.		
21. Trying to pick up someone.		
22. Returning goods to a store.		
23. Giving a party.		
24. Resisting a high pressure salesperson.		

Appendix C

Pearson's correlation coefficients for the relationship between autistic or social anxiety traits and fixation time on the (upper) face, for the data presented in Chapter 4.

Table 10.1. *Pearson's correlation coefficients for the relationship between AQ scores and (upper) face fixation time during speaking and listening periods.*

		Japanese	British/Irish
Speaking	AQ – Face looking (Intro)	-0.224	-0.034
	AQ – Face looking (Story)	-0.249	-0.272
	AQ – Upper face looking	-0.064	-0.348
Listening	AQ – Face looking (Intro)	0.101	-0.165
	AQ – Face looking (Story)	-0.322	-0.055
	AQ – Upper face looking	0.110	-0.148

Table 10.2. *Pearson's correlation coefficients for the relationship between LSAS scores and (upper) face fixation time during speaking and listening periods.*

		Japanese	British/Irish
Speaking	LSAS – Face looking (Intro)	-0.442*	-0.360+
	LSAS – Face looking (Story)	-0.090	-0.394*
	LSAS – Upper face looking	-0.206	0.115
Listening	LSAS – Face looking (Intro)	0.045	-0.340
	LSAS – Face looking (Story)	-0.176	-0.115
	LSAS – Upper face looking	-0.061	-0.159

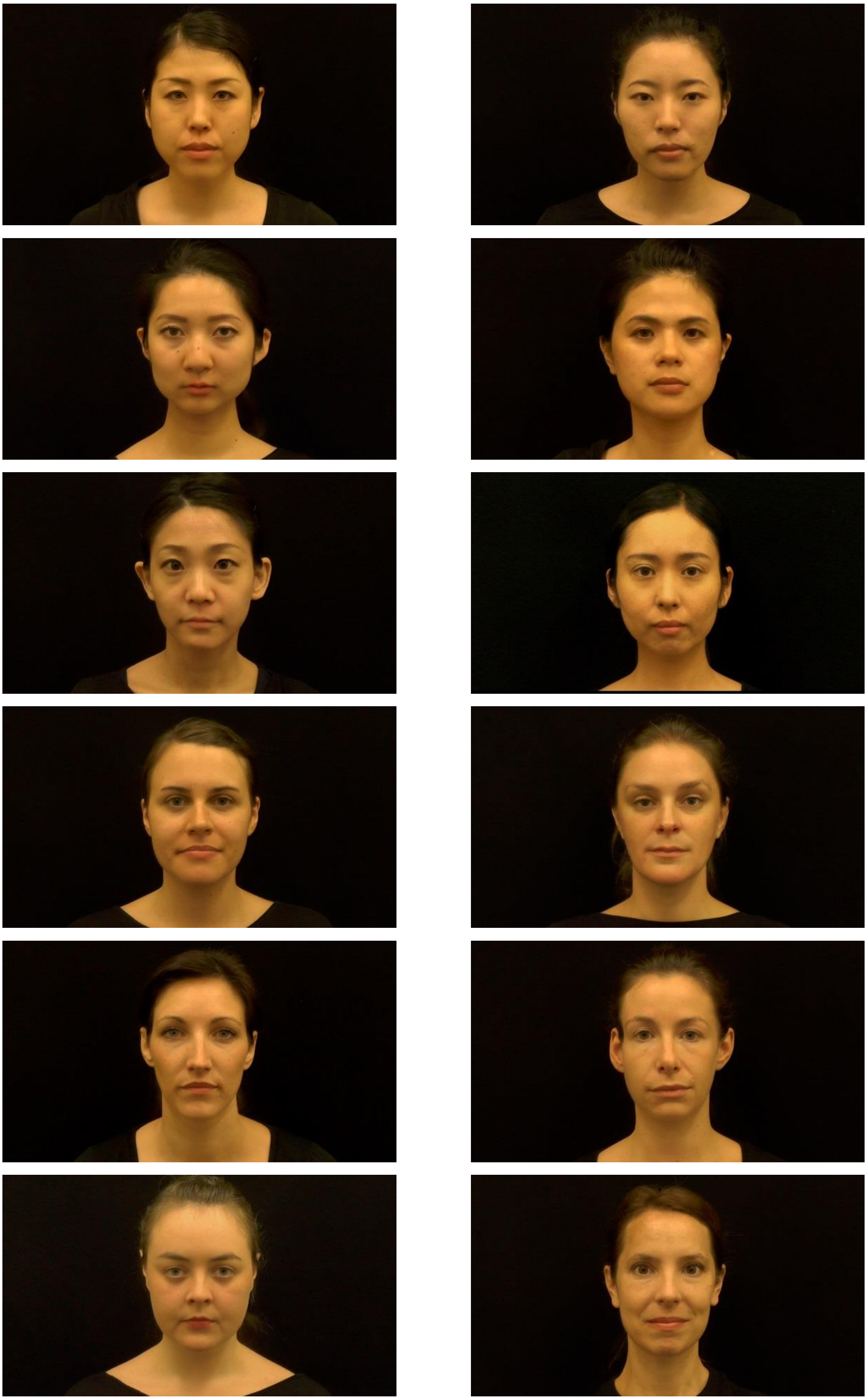
*$p < 0.05$; +$p = 0.055$

Table 10.3. *Pearson's correlation coefficients for the relationship between age and (upper) face fixation time during speaking and listening periods.*

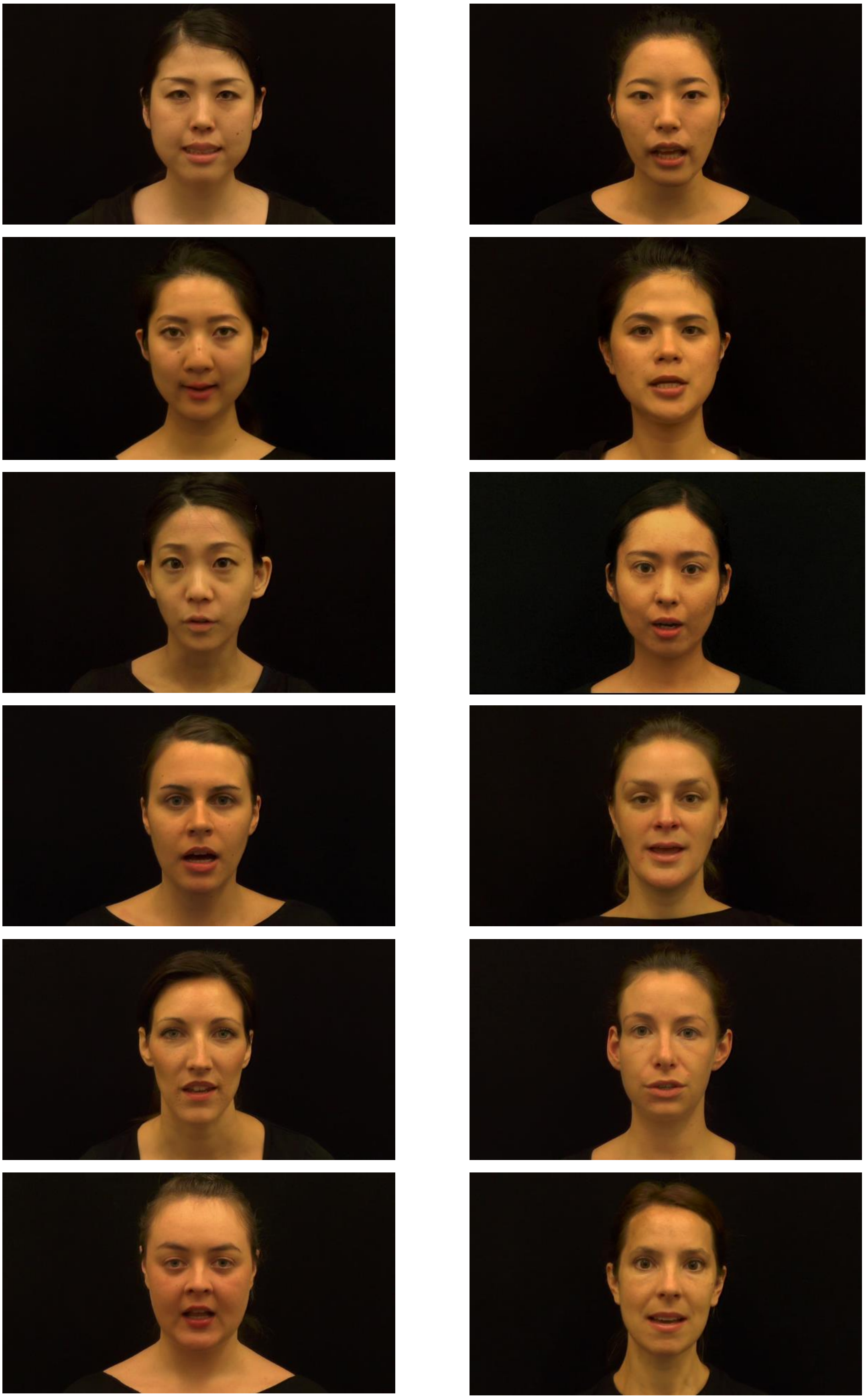
		Japanese	British/Irish
Speaking	Age – Face looking (Intro)	0.295	0.048
	Age – Face looking (Story)	0.211	0.183
	Age – Upper face looking	-0.111	-0.239
Listening	Age – Face looking (Intro)	0.220	0.101
	Age – Face looking (Story)	0.081	-0.071
	Age – Upper face looking	0.176	0.155

Appendix D

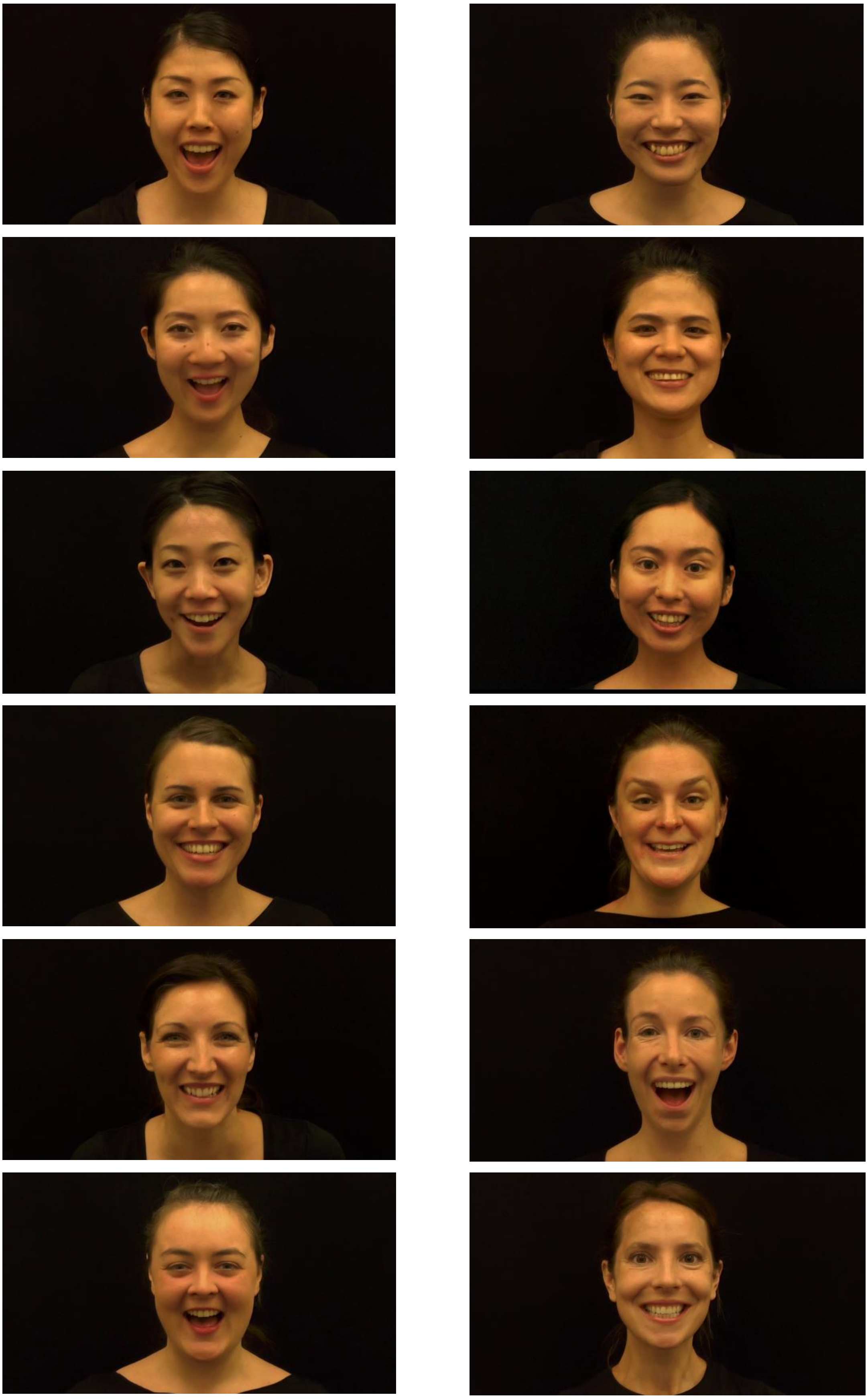
Face stimuli in the static condition used for the study presented in Chapter 5:



Screenshots of the face stimuli in the dynamic-neutral condition used for the study presented in Chapter 5:



Screenshots of the face stimuli in the dynamic-social condition (peekaboo) used for the study presented in Chapter 5:



Screenshots of the face stimuli in the dynamic-social condition (nodding) used for the study presented in Chapter 5:

Appendix E

Four-page demographic questionnaire handed out to caregivers of infants who participated in the studies presented in Chapter 5 and Chapter 6.

Babylab

Baby Culture

Demographic questionnaire

Please read through the following questions carefully and answer each question as best as you can by ticking the boxes. **If there are any questions that you would prefer not to answer,** you can leave these blank. All information that you provide on this questionnaire will be kept confidential. The data will be identified with a code, with personal details kept in locked file or secure computer with access only by the researchers of this study. We will use the answers provided on this questionnaire for demographic background information only.

The first part will ask questions about **you**. The second part will ask questions about **your child.**

Participant ID: ____________
Date completed: ____________
Mother's age: ____________

What is your native language? *If multiple, please indicate which ones.*	☐ English ☐ Other (please specify): ____________
Which language do you use to communicate with your child? *If multiple, please indicate which ones.*	☐ English ☐ Other (please specify): ____________
In which country were you born?	☐ UK ☐ Other (please specify): ____________
In which country were you raised? *If multiple, please state all countries.*	☐ UK ☐ Other (please specify): ____________
Have you ever lived outside the UK? *If 'Yes', please state where and for how long.*	☐ Yes Where? ____________________ For how long? ________________ ☐ No
If you were born or raised outside the UK, or have lived outside the UK, please indicate your UK entry date.	*Date:* ____________________

Please turn over page.

What is <u>your</u> ethnic group?
Choose one option that best describes your ethnic group or background.

White
- ☐ English / Welsh / Scottish / Northern Irish / British
- ☐ Irish
- ☐ Gypsy or Irish Traveller
- ☐ Any other White background (please specify):

Mixed / Multiple ethnic groups
- ☐ White and Black Caribbean
- ☐ White and Black African
- ☐ White and Asian
- ☐ Any other Mixed / Multiple ethnic background (please specify):

Asian / Asian British
- ☐ Indian
- ☐ Pakistani
- ☐ Bangladeshi
- ☐ Chinese
- ☐ Any other Asian background (please specify):

Black / African / Caribbean / Black British
- ☐ African
- ☐ Caribbean
- ☐ Any other Black / African / Caribbean background (please specify):

Other ethnic group
- ☐ Arab
- ☐ Any other ethnic group (please specify):

☐ **Prefer not to say**

Do you have close friends who are Eastern Asian (e.g. Chinese, Japanese, Korean)?

☐ Yes – please answer the following two questions:

<u>How many</u> of your close friends are Eastern Asian? ______________________

How frequently does <u>your child</u> interact with them? ______________________

☐ No

Please turn over page.

The following part will ask questions about **your child.**

In which country was <u>your child</u> born?	☐ UK ☐ Other (please specify): ______________
Has <u>your child</u> ever lived outside the UK? *If 'Yes', please state where and for how long.*	☐ Yes Where? ______________________________ For how long? _________________________ ☐ No
If <u>your child</u> was born outside the UK, or has lived outside the UK, please indicate your child's UK entry date.	*Date:* _________________________

What is <u>your child's</u> ethnic group?
Choose one option that best describes your child's ethnic group or background.

White
- ☐ English / Welsh / Scottish / Northern Irish / British
- ☐ Irish
- ☐ Gypsy or Irish Traveller
- ☐ Any other White background (please specify):

Mixed / Multiple ethnic groups
- ☐ White and Black Caribbean
- ☐ White and Black African
- ☐ White and Asian
- ☐ Any other Mixed / Multiple ethnic background (please specify):

Asian / Asian British
- ☐ Indian
- ☐ Pakistani
- ☐ Bangladeshi
- ☐ Chinese
- ☐ Any other Asian background (please specify):

Black / African / Caribbean / Black British
- ☐ African
- ☐ Caribbean
- ☐ Any other Black / African / Caribbean background (please specify):

Other ethnic group
- ☐ Arab
- ☐ Any other ethnic group (please specify):

☐ **Prefer not to say**

Please turn over page.

Please answer each question as carefully as possible by choosing *one* of the numbers to indicate your degree of agreement or disagreement.

In general, how would you rate <u>YOUR</u> amount of contact with Eastern Asian people?

1 2 3 4 5 6 7

1 = very little 7 = very extensive

In general, how would you rate <u>YOUR CHILD'S</u> amount of contact with Eastern Asian people?

1 2 3 4 5 6 7

1 = very little 7 = very extensive

In general, how would you rate <u>YOUR</u> amount of contact with people of White ethnicity?

1 2 3 4 5 6 7

1 = very little 7 = very extensive

In general, how would you rate <u>YOUR CHILD'S</u> amount of contact with people of White ethnicity?

1 2 3 4 5 6 7

1 = very little 7 = very extensive

End of questionnaire. Thank you!

Appendix F

Detailed descriptive and inferential statistics for the analysis presented in Chapter 5.

Table 13.1. *Means and standard deviations of eye dwell time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Eyes	Japanese	10	British	0.49	0.23
			Japanese	0.65	0.15
		15-17	British	0.39	0.21
			Japanese	0.50	0.18
		Adults	British	0.46	0.16
			Japanese	0.42	0.22
	White-British	10	British	0.55	0.25
			Japanese	0.65	0.20
		15-17	British	0.44	0.24
			Japanese	0.51	0.19
		Adults	British	0.50	0.13
			Japanese	0.48	0.23

Table 13.2. *Means and standard deviations of eye dwell time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Eyes	Japanese	10	British	0.28	0.18
			Japanese	0.31	0.16
		15-17	British	0.16	0.15
			Japanese	0.21	0.12
		Adults	British	0.21	0.15
			Japanese	0.17	0.15
	White-British	10	British	0.24	0.20
			Japanese	0.35	0.19
		15-17	British	0.14	0.12
			Japanese	0.27	0.20
		Adults	British	0.23	0.15
			Japanese	0.21	0.15

Table 13.3. *Means and standard deviations of eye dwell time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Eyes	Japanese	10	British	0.31	0.22
			Japanese	0.40	0.20
		15-17	British	0.18	0.17
			Japanese	0.27	0.16
		Adults	British	0.29	0.18
			Japanese	0.26	0.17
	White-British	10	British	0.25	0.16
			Japanese	0.39	0.23
		15-17	British	0.17	0.16
			Japanese	0.32	0.18
		Adults	British	0.34	0.21
			Japanese	0.24	0.18

Table 13.4. *Means and standard deviations of bridge dwell time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Bridge	Japanese	10	British	0.09	0.06
			Japanese	0.07	0.04
		15-17	British	0.08	0.07
			Japanese	0.09	0.06
		Adults	British	0.08	0.06
			Japanese	0.14	0.11
	White-British	10	British	0.07	0.05
			Japanese	0.06	0.04
		15-17	British	0.07	0.07
			Japanese	0.09	0.06
		Adults	British	0.08	0.07
			Japanese	0.09	0.06

Table 13.5. *Means and standard deviations of bridge dwell time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Bridge	Japanese	10	British	0.08	0.05
			Japanese	0.07	0.05
		15-17	British	0.05	0.04
			Japanese	0.07	0.05
		Adults	British	0.06	0.06
			Japanese	0.07	0.08
	White-British	10	British	0.08	0.05
			Japanese	0.06	0.04
		15-17	British	0.06	0.07
			Japanese	0.05	0.04
		Adults	British	0.06	0.06
			Japanese	0.08	0.09

Table 13.6. *Means and standard deviations of bridge dwell time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Bridge	Japanese	10	British	0.08	0.05
			Japanese	0.09	0.07
		15-17	British	0.06	0.06
			Japanese	0.07	0.03
		Adults	British	0.11	0.11
			Japanese	0.13	0.11
	White-British	10	British	0.08	0.06
			Japanese	0.08	0.05
		15-17	British	0.05	0.05
			Japanese	0.07	0.04
		Adults	British	0.09	0.09
			Japanese	0.10	0.11

Table 13.7. *Means and standard deviations of nose dwell time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Nose	Japanese	10	British	0.14	0.11
			Japanese	0.11	0.08
		15-17	British	0.14	0.08
			Japanese	0.14	0.13
		Adults	British	0.16	0.13
			Japanese	0.21	0.19
	White-British	10	British	0.13	0.12
			Japanese	0.10	0.09
		15-17	British	0.12	0.08
			Japanese	0.15	0.13
		Adults	British	0.14	0.08
			Japanese	0.17	0.14

Table 13.8. *Means and standard deviations of nose dwell time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Nose	Japanese	10	British	0.14	0.13
			Japanese	0.16	0.07
		15-17	British	0.10	0.06
			Japanese	0.16	0.09
		Adults	British	0.15	0.08
			Japanese	0.23	0.17
	White-British	10	British	0.15	0.13
			Japanese	0.14	0.09
		15-17	British	0.09	0.06
			Japanese	0.16	0.10
		Adults	British	0.14	0.10
			Japanese	0.22	0.17

Table 13.9. *Means and standard deviations of nose dwell time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Nose	Japanese	10	British	0.14	0.11
			Japanese	0.16	0.10
		15-17	British	0.11	0.08
			Japanese	0.22	0.08
		Adults	British	0.16	0.11
			Japanese	0.24	0.20
	White-British	10	British	0.13	0.11
			Japanese	0.16	0.12
		15-17	British	0.12	0.07
			Japanese	0.17	0.08
		Adults	British	0.17	0.12
			Japanese	0.28	0.22

Table 13.10. *Means and standard deviations of mouth dwell time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Mouth	Japanese	10	British	0.10	0.15
			Japanese	0.04	0.05
		15-17	British	0.22	0.23
			Japanese	0.13	0.12
		Adults	British	0.14	0.09
			Japanese	0.05	0.04
	White-British	10	British	0.08	0.12
			Japanese	0.04	0.04
		15-17	British	0.19	0.24
			Japanese	0.12	0.14
		Adults	British	0.13	0.10
			Japanese	0.06	0.07

Table 13.11. *Means and standard deviations of mouth dwell time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Mouth	Japanese	10	British	0.31	0.25
			Japanese	0.35	0.15
		15-17	British	0.56	0.22
			Japanese	0.43	0.17
		Adults	British	0.49	0.20
			Japanese	0.38	0.23
	White-British	10	British	0.35	0.24
			Japanese	0.35	0.21
		15-17	British	0.60	0.20
			Japanese	0.42	0.25
		Adults	British	0.49	0.23
			Japanese	0.38	0.24

Table 13.12. *Means and standard deviations of mouth dwell time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Mouth	Japanese	10	British	0.28	0.21
			Japanese	0.22	0.20
		15-17	British	0.51	0.24
			Japanese	0.30	0.19
		Adults	British	0.34	0.22
			Japanese	0.19	0.21
	White-British	10	British	0.34	0.23
			Japanese	0.20	0.11
		15-17	British	0.54	0.22
			Japanese	0.30	0.19
		Adults	British	0.30	0.21
			Japanese	0.23	0.21

Table 13.13. *Summary table of the five-factor ANOVA ROI x Group x Age x Stimulus Condition x Face Ethnicity on dwell time data.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**1.65**	**0.025**	**8.26**	**0.001**	**0.054**
Stimulus Condition x Age	**3.03**	**0.008**	**2.74**	**0.039**	**0.037**
Stimulus Condition x Group	1.65	0.005	1.53	0.220	0.011
Stimulus Condition x Age x Group	3.30	0.004	1.25	0.291	0.017
Error (Stimulus Condition)	237.83	0.003			
Face Ethnicity	1	0.003	3.21	0.075	0.022
Face Ethnicity x Age	2	0.001	1.06	0.351	0.014
Face Ethnicity x Group	1	<0.001	0.14	0.713	0.001
Face Ethnicity x Age x Group	2	0.001	0.83	0.440	0.011
Error (Face Ethnicity)	144	0.001			
ROI	**1.84**	**19.483**	**105.85**	**<0.001**	**0.424**
ROI x Age	**3.69**	**1.332**	**7.24**	**<0.001**	**0.091**
ROI x Group	**1.84**	**1.725**	**9.37**	**<0.001**	**0.061**
ROI x Age x Group	**3.69**	**0.456**	**2.48**	**0.049**	**0.033**
Error (ROI)	265.35	0.184			
Stimulus Condition x Face Ethnicity	2	0.002	1.79	0.169	0.012
Stimulus Condition x Face Ethnicity x Age	4	<0.001	0.33	0.858	0.005
Stimulus Condition x Face Ethnicity x Group	2	<0.001	0.28	0.759	0.002
Stimulus Condition x Face Ethnicity x Age x Group	4	<0.001	0.34	0.848	0.005
Error (Stimulus Condition x Face Ethnicity)	288	0.001			
Stimulus Condition x ROI	**3.86**	**4.406**	**238.60**	**<0.001**	**0.624**
Stimulus Condition x ROI x Age	**7.73**	**0.073**	**2.55**	**0.011**	**0.034**
Stimulus Condition x ROI x Group	**3.86**	**0.089**	**3.09**	**0.017**	**0.021**

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition x ROI x Age x Group	7.73	0.033	1.16	0.320	0.016
Error (Stimulus Condition x ROI)	556.23	0.029			
Face Ethnicity x ROI	**2.03**	**0.032**	**3.81**	**0.023**	**0.026**
Face Ethnicity x ROI x Age	4.06	0.011	0.87	0.486	0.012
Face Ethnicity x ROI x Group	2.03	0.008	0.64	0.532	0.004
Face Ethnicity x ROI x Age x Group	4.06	0.016	1.30	0.269	0.018
Error (Face Ethnicity x ROI)	292.45	0.012			
Stimulus Condition x Face Ethnicity x ROI	4.01	0.020	1.74	0.140	0.012
Stimulus Condition x Face Ethnicity x ROI x Age	8.02	0.008	0.75	0.649	0.010
Stimulus Condition x Face Ethnicity x ROI x Group	4.01	0.022	1.98	0.095	0.014
Stimulus Condition x Face Ethnicity x ROI x Age x Group	**8.02**	**0.026**	**2.29**	**0.020**	**0.031**
Error (Stimulus Condition x Face Ethnicity x ROI)	577.71	0.011			
Age	2	0.017	1.71	0.184	0.023
Group	1	<0.001	0.03	0.857	<0.001
Age x Group	**2**	**0.043**	**4.38**	**0.014**	**0.057**
Error	144	0.010			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.14. *Summary table of the four-factor ANOVA ROI x Group x Age x Stimulus Condition for face stimuli of Japanese ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	1.78	0.006	2.99	0.058	0.020
Stimulus Condition x Age	3.56	0.004	1.94	0.113	0.026
Stimulus Condition x Group	1.78	0.002	0.81	0.434	0.006
Stimulus Condition x Age x Group	3.56	0.001	0.56	0.671	0.008
Error (Stimulus Condition)	256.51	0.002			
ROI	**1.89**	**5.416**	**90.76**	**<0.001**	**0.387**
ROI x Age	**3.77**	**0.719**	**7.57**	**<0.001**	**0.095**
ROI x Group	**1.89**	**0.790**	**8.33**	**<0.001**	**0.055**
ROI x Age x Group	3.77	0.196	2.06	0.090	0.028
Error (ROI)	271.67	0.095			
Stimulus Condition x ROI	**4.13**	**2.944**	**161.43**	**<0.001**	**0.529**
Stimulus Condition x ROI x Age	8.25	0.035	1.93	0.051	0.026
Stimulus Condition x ROI x Group	**4.13**	**0.057**	**3.11**	**0.014**	**0.021**
Stimulus Condition x ROI x Age x Group	8.25	0.021	1.14	0.335	0.016
Error (Stimulus Condition x ROI)	594.28	0.018			
Age	2	0.005	0.91	0.407	0.012
Group	1	<0.001	<0.05	0.950	<0.001
Age x Group	**2**	**0.027**	**5.11**	**0.001**	**0.066**
Error	144	0.005			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.15. *Summary table of the four-factor ANOVA ROI x Group x Age x Stimulus Condition for face stimuli of White-British ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**1.82**	**0.019**	**9.37**	**<0.001**	**0.061**
Stimulus Condition x Age	3.64	0.004	1.98	0.104	0.027
Stimulus Condition x Group	1.82	0.003	1.45	0.237	0.010
Stimulus Condition x Age x Group	3.64	0.003	1.36	0.251	0.019
Error (Stimulus Condition)	261.85	0.002			
ROI	**1.83**	**10.79**	**106.54**	**<0.001**	**0.425**
ROI x Age	**3.66**	**0.612**	**6.04**	**<0.001**	**0.077**
ROI x Group	**1.83**	**0.931**	**9.19**	**<0.001**	**0.060**
ROI x Age x Group	**3.66**	**0.275**	**2.72**	**0.035**	**0.036**
Error (ROI)	263.55	0.101			
Stimulus Condition x ROI	**3.93**	**3.654**	**178.19**	**<0.001**	**0.553**
Stimulus Condition x ROI x Age	**7.86**	**0.043**	**2.11**	**0.034**	**0.028**
Stimulus Condition x ROI x Group	**3.93**	**0.050**	**2.45**	**0.047**	**0.017**
Stimulus Condition x ROI x Age x Group	7.86	0.037	1.82	0.072	0.025
Error (Stimulus Condition x ROI)	565.95	0.021			
Age	2	0.013	2.39	0.095	0.032
Group	1	<0.001	0.08	0.779	0.001
Age x Group	2	0.016	3.05	0.051	0.041
Error	144	0.005			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.16. *Means and standard deviations of eye fixation time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Eyes	Japanese	10	British	0.53	0.29
			Japanese	0.69	0.18
		15-17	British	0.39	0.26
			Japanese	0.51	0.19
		Adults	British	0.47	0.18
			Japanese	0.45	0.22
	White-British	10	British	0.58	0.28
			Japanese	0.66	0.20
		15-17	British	0.47	0.28
			Japanese	0.60	0.22
		Adults	British	0.51	0.14
			Japanese	0.50	0.24

Table 13.17. *Means and standard deviations of eye fixation time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Eyes	Japanese	10	British	0.33	0.22
			Japanese	0.32	0.17
		15-17	British	0.18	0.17
			Japanese	0.25	0.14
		Adults	British	0.23	0.16
			Japanese	0.20	0.19
	White-British	10	British	0.25	0.21
			Japanese	0.38	0.22
		15-17	British	0.14	0.14
			Japanese	0.28	0.23
		Adults	British	0.25	0.15
			Japanese	0.22	0.16

Table 13.18. *Means and standard deviations of eye fixation time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Eyes	Japanese	10	British	0.31	0.26
			Japanese	0.40	0.21
		15-17	British	0.19	0.19
			Japanese	0.26	0.17
		Adults	British	0.31	0.18
			Japanese	0.28	0.19
	White-British	10	British	0.23	0.17
			Japanese	0.41	0.24
		15-17	British	0.17	0.18
			Japanese	0.32	0.19
		Adults	British	0.35	0.21
			Japanese	0.25	0.18

Table 13.19. *Means and standard deviations of bridge fixation time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Bridge	Japanese	10	British	0.10	0.09
			Japanese	0.06	0.05
		15-17	British	0.08	0.11
			Japanese	0.09	0.06
		Adults	British	0.08	0.06
			Japanese	0.13	0.11
	White-British	10	British	0.09	0.08
			Japanese	0.05	0.03
		15-17	British	0.06	0.08
			Japanese	0.06	0.04
		Adults	British	0.07	0.07
			Japanese	0.08	0.04

Table 13.20. *Means and standard deviations of bridge fixation time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Bridge	Japanese	10	British	0.09	0.09
			Japanese	0.05	0.05
		15-17	British	0.06	0.11
			Japanese	0.06	0.06
		Adults	British	0.05	0.06
			Japanese	0.06	0.08
	White-British	10	British	0.09	0.08
			Japanese	0.04	0.03
		15-17	British	0.05	0.08
			Japanese	0.04	0.04
		Adults	British	0.05	0.06
			Japanese	0.08	0.09

Table 13.21. *Means and standard deviations of bridge fixation time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Bridge	Japanese	10	British	0.07	0.06
			Japanese	0.09	0.07
		15-17	British	0.06	0.07
			Japanese	0.07	0.03
		Adults	British	0.10	0.10
			Japanese	0.12	0.10
	White-British	10	British	0.10	0.10
			Japanese	0.07	0.06
		15-17	British	0.04	0.06
			Japanese	0.07	0.04
		Adults	British	0.08	0.08
			Japanese	0.11	0.15

Table 13.22. *Means and standard deviations of nose fixation time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Nose	Japanese	10	British	0.11	0.15
			Japanese	0.11	0.08
		15-17	British	0.14	0.09
			Japanese	0.15	0.13
		Adults	British	0.14	0.13
			Japanese	0.20	0.20
	White-British	10	British	0.12	0.14
			Japanese	0.12	0.10
		15-17	British	0.09	0.09
			Japanese	0.13	0.12
		Adults	British	0.14	0.08
			Japanese	0.19	0.17

Table 13.23. *Means and standard deviations of nose fixation time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Nose	Japanese	10	British	0.10	0.12
			Japanese	0.15	0.08
		15-17	British	0.06	0.05
			Japanese	0.17	0.10
		Adults	British	0.14	0.08
			Japanese	0.22	0.18
	White-British	10	British	0.12	0.16
			Japanese	0.14	0.09
		15-17	British	0.07	0.09
			Japanese	0.16	0.11
		Adults	British	0.14	0.10
			Japanese	0.23	0.19

Table 13.24. *Means and standard deviations of nose fixation time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Nose	Japanese	10	British	0.13	0.15
			Japanese	0.15	0.10
		15-17	British	0.10	0.09
			Japanese	0.22	0.10
		Adults	British	0.16	0.12
			Japanese	0.24	0.22
	White-British	10	British	0.12	0.14
			Japanese	0.13	0.11
		15-17	British	0.11	0.10
			Japanese	0.17	0.08
		Adults	British	0.17	0.13
			Japanese	0.25	0.20

Table 13.25. *Means and standard deviations of mouth fixation time proportional to inner face looking time of static faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Mouth	Japanese	10	British	0.12	0.20
			Japanese	0.04	0.06
		15-17	British	0.22	0.25
			Japanese	0.12	0.14
		Adults	British	0.14	0.09
			Japanese	0.05	0.04
	White-British	10	British	0.10	0.17
			Japanese	0.03	0.03
		15-17	British	0.22	0.26
			Japanese	0.10	0.09
		Adults	British	0.14	0.10
			Japanese	0.06	0.06

Table 13.26. *Means and standard deviations of mouth fixation time proportional to inner face looking time of dynamic-neutral faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Mouth	Japanese	10	British	0.30	0.29
			Japanese	0.37	0.14
		15-17	British	0.59	0.24
			Japanese	0.43	0.17
		Adults	British	0.50	0.20
			Japanese	0.39	0.26
	White-British	10	British	0.36	0.28
			Japanese	0.35	0.21
		15-17	British	0.61	0.25
			Japanese	0.42	0.27
		Adults	British	0.49	0.22
			Japanese	0.39	0.26

Table 13.27. *Means and standard deviations of mouth fixation time proportional to inner face looking time of dynamic-social faces.*

ROI	Face Ethnicity	Age	Group	*M*	*SD*
Mouth	Japanese	10	British	0.32	0.22
			Japanese	0.22	0.20
		15-17	British	0.52	0.28
			Japanese	0.30	0.21
		Adults	British	0.35	0.22
			Japanese	0.19	0.22
	White-British	10	British	0.35	0.25
			Japanese	0.23	0.15
		15-17	British	0.57	0.23
			Japanese	0.30	0.21
		Adults	British	0.31	0.23
			Japanese	0.24	0.23

Table 13.28. *Summary table of the five-factor ANOVA ROI x Group x Age x Stimulus Condition x Face Ethnicity on fixation data.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**1.83**	**0.025**	**5.84**	**0.004**	**0.039**
Stimulus Condition x Age	**3.65**	**0.017**	**4.11**	**0.004**	**0.054**
Stimulus Condition x Group	1.83	0.007	1.55	0.215	0.011
Stimulus Condition x Age x Group	3.65	0.008	1.87	0.123	0.025
Error (Stimulus Condition)	263.11	0.004			
Face Ethnicity	1	0.002	0.72	0.396	0.005
Face Ethnicity x Age	2	0.002	0.67	0.515	0.009
Face Ethnicity x Group	1	<0.001	0.03	0.862	<0.001
Face Ethnicity x Age x Group	2	0.001	0.22	0.802	0.003
Error (Face Ethnicity)	144	0.003			
ROI	**1.86**	**22.941**	**110.49**	**<0.001**	**0.434**
ROI x Age	**3.71**	**1.359**	**6.54**	**<0.001**	**0.083**
ROI x Group	**1.86**	**2.100**	**10.11**	**<0.001**	**0.066**
ROI x Age x Group	3.71	0.493	2.37	0.057	0.032
Error (ROI)	267.43	0.208			
Stimulus Condition x Face Ethnicity	2	0.001	0.23	0.797	0.002
Stimulus Condition x Face Ethnicity x Age	4	0.002	0.83	0.505	0.011
Stimulus Condition x Face Ethnicity x Group	2	0.003	1.26	0.286	0.009
Stimulus Condition x Face Ethnicity x Age x Group	4	0.001	0.23	0.924	0.003
Error (Stimulus Condition x Face Ethnicity)	288	0.002			
Stimulus Condition x ROI	**3.99**	**7.040**	**198.40**	**<0.001**	**0.579**
Stimulus Condition x ROI x Age	**7.99**	**0.089**	**2.52**	**0.011**	**0.034**
Stimulus Condition x ROI x Group	**3.99**	**0.093**	**2.62**	**0.034**	**0.018**
Stimulus Condition x ROI x Age x Group	7.99	0.042	1.17	0.313	0.016
Error (Stimulus Condition x ROI)	575.12	0.035			

Source	df	MS	F	p	η_p^2
Face Ethnicity x ROI	2.14	0.046	2.73	0.063	0.019
Face Ethnicity x ROI x Age	4.27	0.020	1.20	0.312	0.16
Face Ethnicity x ROI x Group	2.14	0.028	1.69	0.183	0.12
Face Ethnicity x ROI x Age x Group	4.27	0.030	1.77	0.131	0.024
Error (Face Ethnicity x ROI)	307.41	0.017			
Stimulus Condition x Face Ethnicity x ROI	**4.64**	**0.030**	**2.37**	**0.042**	**0.016**
Stimulus Condition x Face Ethnicity x ROI x Age	9.29	0.008	0.65	0.758	0.009
Stimulus Condition x Face Ethnicity x ROI x Group	4.64	0.020	1.56	0.173	0.011
Stimulus Condition x Face Ethnicity x ROI x Age x Group	9.29	0.023	1.84	0.057	0.025
Error (Stimulus Condition x Face Ethnicity x ROI)	668.65	0.013			
Age	2	0.007	0.54	0.584	0.007
Group	1	0.003	0.21	0.646	0.001
Age x Group	2	0.036	2.77	0.066	0.037
Error	144	0.010			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.29. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of White-British ethnicity in the static condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.64**	**11.831**	**251.18**	**<0.001**	**0.629**
ROI x Group	**1.64**	**0.154**	**3.27**	**0.049**	**0.022**
Error (ROI)	242.96	0.047			
Group	1	<0.001	0.11	0.736	0.001
Error	148	0.004			

Table 13.30. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of White-British ethnicity in the dynamic-neutral condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.71**	**6.228**	**104.10**	**<0.001**	**0.413**
ROI x Group	**1.71**	**0.375**	**6.27**	**0.004**	**0.041**
Error (ROI)	253.44	0.060			
Group	1	0.001	0.421	0.518	0.003
Error	148	0.003			

Table 13.31. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of White-British ethnicity in the dynamic-social condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.92**	**2.523**	**45.01**	**<0.001**	**0.233**
ROI x Group	**1.93**	**0.591**	**10.54**	**<0.001**	**0.066**
Error (ROI)	284.77	0.056			
Group	1	0.004	1.16	0.284	0.008
Error	148	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05

Table 13.32. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of Japanese ethnicity in the static condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.90**	**7.842**	**185.99**	**<0.001**	**0.557**
ROI x Group	**1.90**	**0.269**	**6.37**	**0.002**	**0.041**
Error (ROI)	280.45	0.042			
Group	1	0.001	0.18	0.672	0.001
Error	148	0.004			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.33. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of Japanese ethnicity in the dynamic-neutral condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.69**	**5.751**	**110.42**	**<0.001**	**0.427**
ROI x Group	**1.69**	**0.237**	**4.55**	**0.016**	**0.030**
Error (ROI)	250.42	0.052			
Group	1	<0.001	0.042	0.837	<0.001
Error	148	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.34. *Summary table of the two-factor ANOVA ROI x Group for face stimuli of Japanese ethnicity in the dynamic-social condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.86**	**2.303**	**39.71**	**<0.001**	**0.212**
ROI x Group	**1.86**	**0.575**	**9.91**	**<0.001**	**0.063**
Error (ROI)	275.19	0.058			
Group	1	0.003	1.15	0.286	0.008
Error	148	0.002			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.35. *Summary table of independent t-tests contrasting British > Japanese participants in the dwell time measure.*

Face Ethnicity	Stimulus Condition	ROI	*t*	*p*	Cohen's *d*
Japanese	Static	Eyes	-2.00	0.048	0.328
		Bridge	-1.59	0.114	0.303
		Nose	-0.59	0.556	0.099
		Mouth	**4.48**	**<0.001**	**0.710**
	Dynamic-neutral	Eyes	-0.32	0.751	0.051
		Bridge	-0.98	0.330	0.146
		Nose	**-3.25**	**0.002**	**0.542**
		Mouth	2.32	0.022	0.377
	Dynamic-social	Eyes	-1.35	0.178	0.218
		Bridge	-1.36	0.175	0.218
		Nose	**-3.16**	**0.002**	**0.530**
		Mouth	**4.10**	**<0.001**	**0.667**
White-British	Static	Eyes	-1.35	0.180	0.225
		Bridge	-0.98	0.328	0.167
		Nose	-0.79	0.431	0.151
		Mouth	**3.11**	**0.002**	**0.494**
	Dynamic-neutral	Eyes	-2.13	0.035	0.370
		Bridge	-0.12	0.906	0.018
		Nose	**-2.60**	**0.011**	**0.432**
		Mouth	**2.68**	**0.008**	**0.449**
	Dynamic-social	Eyes	-1.51	0.134	0.255
		Bridge	-1.11	0.269	0.171
		Nose	**-2.91**	**0.004**	**0.489**
		Mouth	**4.43**	**<0.001**	**0.717**

Note: Rows in bold indicate significant effect at Bonferroni-adjusted alpha-level of 0.0125.

Table 13.36. *Summary table of the two-factor ANOVA ROI x Group for static faces.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.71**	**12.044**	**244.00**	**<0.001**	**0.622**
ROI x Group	**1.71**	**0.272**	**5.51**	**0.007**	**0.036**
Error (ROI)	252.32	0.049			
Group	1	<0.001	0.003	0.958	<0.001
Error	148	0.004			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.37. *Summary table of the two-factor ANOVA ROI x Group for dynamic-neutral faces.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.75**	**6.313**	**115.97**	**<0.001**	**0.439**
ROI x Group	**1.75**	**0.333**	**6.13**	**0.004**	**0.040**
Error (ROI)	258.86	0.054			
Group	1	0.001	0.389	0.534	0.003
Error	148	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.38. *Summary table of the two-factor ANOVA ROI x Group for dynamic-social faces.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.89**	**2.852**	**48.91**	**<0.001**	**0.248**
ROI x Group	**1.89**	**0.625**	**10.72**	**<0.001**	**0.068**
Error (ROI)	279.17	0.058			
Group	1	0.004	1.68	0.197	0.011
Error	148	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.39. *Summary table of independent t-tests contrasting British > Japanese participants in the fixation time measure.*

Stimulus Condition	ROI	t	p	Cohen's d
Static	Eyes	-1.70	0.091	0.295
	Bridge	0.09	0.929	0.013
	Nose	-1.38	0.170	0.194
	Mouth	**4.15**	**<0.001**	**0.653**
Dynamic-neutral	Eyes	-1.23	0.220	0.198
	Bridge	0.932	0.353	0.141
	Nose	**-4.08**	**<0.001**	**0.684**
	Mouth	2.45	0.015	0.376
Dynamic-social	Eyes	-1.64	0.103	0.265
	Bridge	-1.41	0.161	0.216
	Nose	**-2.86**	**0.005**	**0.531**
	Mouth	**4.38**	**<0.001**	**0.712**

Note: Rows in bold indicate significant effect at Bonferroni-adjusted alpha-level of 0.0125.

Table 13.40. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of White-British ethnicity in the static condition.*

Source	df	MS	F	p	η_p^2
ROI	**1.67**	**11.271**	**252.99**	**<0.001**	**0.632**
ROI x Age	**3.34**	**0.229**	**5.14**	**0.001**	**0.065**
Error (ROI)	245.37	0.045			
Age	2	<0.001	0.12	0.889	0.002
Error	147	0.004			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.41. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of White-British ethnicity in the dynamic-neutral condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.73**	**6.568**	**113.74**	**<0.001**	**0.436**
ROI x Age	**3.46**	**0.327**	**5.67**	**<0.001**	**0.072**
Error (ROI)	254.09	0.058			
Age	**2**	**0.010**	**3.97**	**0.021**	**0.051**
Error	147	0.002			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.42. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of White-British ethnicity in the dynamic-social condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.94**	**2.832**	**51.66**	**<0.001**	**0.260**
ROI x Age	**3.88**	**0.373**	**6.80**	**<0.001**	**0.085**
Error (ROI)	285.49	0.055			
Age	**2**	**0.011**	**3.77**	**0.025**	**0.049**
Error	147	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.43. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of Japanese ethnicity in the static condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.92**	**7.433**	**184.01**	**<0.001**	**0.556**
ROI x Age	**3.84**	**0.240**	**5.95**	**<0.001**	**0.075**
Error (ROI)	282.51	0.040			
Age	2	0.001	0.136	0.873	0.002
Error	147	0.004			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.44. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of Japanese ethnicity in the dynamic-neutral condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.75**	**5.833**	**123.47**	**<0.001**	**0.456**
ROI x Age	**3.50**	**0.366**	**7.75**	**<0.001**	**0.095**
Error (ROI)	257.45	0.047			
Age	**2**	**0.010**	**3.27**	**0.041**	**0.043**
Error	147	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.45. *Summary table of the two-factor ANOVA ROI x Age for face stimuli of Japanese ethnicity in the dynamic-social condition.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.90**	**2.570**	**45.94**	**<0.001**	**0.238**
ROI x Age	**3.79**	**0.377**	**6.73**	**<0.001**	**0.084**
Error (ROI)	278.83	0.056			
Age	2	0.004	1.411	0.247	0.019
Error	147	0.002			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.46. *Summary table of the one-way ANOVA with factor Age using dwell time data.*

Face Ethnicity	Stimulus Condition	ROI	MS	*F*	*p*	η^2
Japanese	Static	**Eyes**	**0.248**	**5.97**	**0.003**	**0.075**
		Bridge	0.016	2.85	0.061	0.037
		Nose	0.041	2.57	0.080	0.034
		Mouth	**0.157**	**8.58**	**<0.001**	**0.104**
	Dynamic-neutral	**Eyes**	**0.172**	**7.28**	**0.001**	**0.090**
		Bridge	0.002	0.72	0.489	0.010
		Nose	**0.062**	**4.93**	**0.008**	**0.063**
		Mouth	**0.414**	**8.99**	**<0.001**	**0.109**
	Dynamic-social	**Eyes**	**0.202**	**5.80**	**0.004**	**0.073**
		Bridge	**0.042**	**6.42**	**0.002**	**0.080**
		Nose	0.044	2.58	0.079	0.034
		Mouth	**0.430**	**8.55**	**<0.001**	**0.104**
White-British	Static	**Eyes**	**0.218**	**4.89**	**0.009**	**0.062**
		Bridge	0.009	2.47	0.089	0.032
		Nose	0.024	2.03	0.135	0.027
		Mouth	**0.132**	**7.36**	**0.001**	**0.091**
	Dynamic-neutral	**Eyes**	**0.122**	**4.08**	**0.019**	**0.053**
		Bridge	0.004	1.00	0.370	0.013
		Nose	**0.065**	**4.56**	**0.012**	**0.058**
		Mouth	**0.384**	**7.12**	**0.001**	**0.088**
	Dynamic-social	Eyes	0.104	2.67	0.072	0.035
		Bridge	**0.017**	**3.14**	**0.046**	**0.041**
		Nose	**0.124**	**6.49**	**0.002**	**0.081**
		Mouth	**0.490**	**10.64**	**<0.001**	**0.126**

Note: MS = Mean squares, effect size = η^2. Rows in bold indicate significant effect at 0.05.

Table 13.47. *Summary table of independent t-tests contrasting each age group for Japanese face stimuli using dwell time data.*

Face Ethnicity	Stimulus Condition	ROI	Contrast	*t*	*p*	Cohen's *d*
Japanese	Static	Eyes	**10 vs 15**	**2.94**	**0.004**	**0.637**
			10 vs adults	**2.97**	**0.004**	**0.541**
			15 vs adults	-0.39	0.696	0.082
		Mouth	**10 vs 15**	**-3.15**	**0.003**	**0.681**
			10 vs adults	-1.12	0.267	0.178
			15 vs adults	**2.77**	**0.008**	**0.597**
	Dynamic-neutral	Eyes	**10 vs 15**	**3.33**	**0.001**	**0.795**
			10 vs adults	**3.12**	**0.002**	**0.641**
			15 vs adults	-0.53	0.599	0.104
		Nose	10 vs 15	1.36	0.178	0.352
			10 vs adults	-1.85	0.067	0351
			15 vs adults	**-3.28**	**0.001**	**0.627**
		Mouth	**10 vs 15**	**-4.31**	**<0.001**	**0.898**
			10 vs adults	**-2.69**	**0.008**	**0.546**
			15 vs adults	1.79	0.077	0.355
	Dynamic-social	Eyes	**10 vs 15**	**3.36**	**0.001**	**0.708**
			10 vs adults	1.87	0.064	0.365
			15 vs adults	-1.80	0.074	0.355
		Bridge	10 vs 15	1.47	0.147	0.345
			10 vs adults	-2.36	0.020	0.441
			15 vs adults	**-3.50**	**0.001**	**0.661**
		Mouth	**10 vs 15**	**-3.68**	**<0.001**	**0.714**
			10 vs adults	-0.27	0.791	0.054
			15 vs adults	**3.50**	**0.001**	**0.738**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167.

Table 13.48. *Summary table of independent t-tests contrasting each age group for White-British face stimuli using dwell time data.*

Face Ethnicity	Stimulus Condition	ROI	Contrast	*t*	*p*	Cohen's *d*
White-British	Static	Eyes	**10 vs 15**	**2.61**	**0.011**	**0.564**
			10 vs adults	**2.67**	**0.009**	**0.458**
			15 vs adults	-0.51	0.614	0.112
		Mouth	**10 vs 15**	**-3.23**	**0.002**	**0.514**
			10 vs adults	-2.15	0.034	0.413
			15 vs adults	2.04	0.047	0.438
	Dynamic-neutral	Eyes	**10 vs 15**	**2.56**	**0.012**	**0.622**
			10 vs adults	1.94	0.056	0.380
			15 vs adults	-1.05	0.297	0.218
		Nose	10 vs 15	1.40	0.165	0.359
			10 vs adults	-1.69	0.095	0.321
			15 vs adults	**-3.19**	**0.002**	**0.612**
		Mouth	**10 vs 15**	**-3.85**	**<0.001**	**0.796**
			10 vs adults	-1.85	0.067	0.373
			15 vs adults	2.16	0.033	0.433
	Dynamic-social	Bridge	10 vs 15	2.24	0.028	0.472
			10 vs adults	-0.76	0.450	0.202
			15 vs adults	**-2.60**	**0.011**	**0.494**
		Nose	10 vs 15	0.71	0.943	0.015
			10 vs adults	**-2.89**	**0.005**	**0.543**
			15 vs adults	**-3.20**	**0.002**	**0.600**
		Mouth	**10 vs 15**	**-3.65**	**<0.001**	**0.784**
			10 vs adults	0.40	0.692	0.080
			15 vs adults	**4.15**	**<0.001**	**0.893**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167.

Table 13.49. *Summary table of the two-factor ANOVA ROI x Age for static faces.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.76**	**11.35**	**247.98**	**<0.001**	**0.628**
ROI x Age	**3.51**	**0.315**	**6.89**	**<0.001**	**0.086**
Error (ROI)	258.03	0.046			
Age	2	0.008	1.90	0.153	0.025
Error	147	0.004			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.50. *Summary table of the two-factor ANOVA ROI x Age for dynamic-neutral faces.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.81**	**6.424**	**129.83**	**<0.001**	**0.469**
ROI x Age	**3.63**	**0.411**	**8.31**	**<0.001**	**0.102**
Error (ROI)	266.48	0.049			
Age	2	0.008	3.03	0.051	0.040
Error	147	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.51. *Summary table of the two-factor ANOVA ROI x Age for dynamic-social faces.*

Source	*df*	MS	*F*	*p*	η_p^2
ROI	**1.91**	**3.183**	**55.48**	**<0.001**	**0.274**
ROI x Age	**3.81**	**0.361**	**6.29**	**<0.001**	**0.079**
Error (ROI)	280.24	0.057			
Age	**2**	**0.010**	**3.86**	**0.023**	**0.050**
Error	147	0.003			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.52. *Summary table of the one-way ANOVA with factor Age using fixation time data.*

Stimulus Condition	ROI	MS	*F*	*p*	η^2
Static	**Eyes**	**0.371**	**8.10**	**<0.001**	**0.099**
	Bridge	0.004	1.00	0.369	0.013
	Nose	**0.045**	**3.28**	**0.040**	**0.043**
	Mouth	**0.140**	**6.91**	**0.001**	**0.086**
Dynamic-neutral	**Eyes**	**0.209**	**7.80**	**0.001**	**0.096**
	Bridge	0.006	1.20	0.303	0.016
	Nose	**0.83**	**6.45**	**0.002**	**0.081**
	Mouth	**0.455**	**9.50**	**<0.001**	**0.115**
Dynamic-social	**Eyes**	**0.131**	**3.56**	**0.031**	**0.046**
	Bridge	**0.020**	**3.26**	**0.041**	**0.042**
	Nose	**0.084**	**4.42**	**0.014**	**0.057**
	Mouth	**0.463**	**9.24**	**<0.001**	**0.112**

Note: MS = Mean squares, effect size = η^2. Rows in bold indicate significant effect at 0.05.

Table 13.53. *Summary table of independent t-tests contrasting each age group.*

Stimulus Condition	ROI	Contrast	*t*	*p*	Cohen's *d*
Static	Eyes	**10 vs 15**	**3.10**	**0.003**	**0.622**
		10 vs adults	**3.85**	**<0.001**	**0.813**
		15 vs adults	-0.10	0.920	0.021
	Nose	10 vs 15	-0.74	0.463	0.169
		10 vs adults	-2.28	0.025	0.549
		15 vs adults	-1.61	0.110	0.283
	Mouth	**10 vs 15**	**-2.86**	**0.006**	**0.618**
		10 vs adults	-1.41	0.162	0.218
		15 vs adults	2.32	0.024	0.503
Dynamic-neutral	Eyes	**10 vs 15**	**3.34**	**0.001**	**0.798**
		10 vs adults	**3.11**	**0.003**	**0.610**
		15 vs adults	-0.70	0.488	0.145
	Nose	10 vs 15	0.88	0.384	0.198
		10 vs adults	**-2.61**	**0.010**	**0.491**
		15 vs adults	**-3.37**	**0.001**	**0.647**
	Mouth	**10 vs 15**	**-4.40**	**<0.001**	**0.914**
		10 vs adults	**-2.49**	**0.014**	**0.494**
		15 vs adults	2.18	0.032	0.439
Dynamic-social	Eyes	**10 vs 15**	**2.55**	**0.013**	**0.539**
		10 vs adults	0.77	0.445	0.149
		15 vs adults	-2.10	0.039	0.420
	Bridge	10 vs 15	2.03	0.046	0.487
		10 vs adults	-1.08	0.285	0.201
		15 vs adults	**-2.65**	**0.010**	**0.499**
	Nose	10 vs 15	-0.38	0.708	0.088
		10 vs adults	**-2.50**	**0.014**	**0.588**
		15 vs adults	-2.35	0.021	0.454
	Mouth	**10 vs 15**	**-3.47**	**0.001**	**0.744**
		10 vs adults	0.26	0.795	0.053
		15 vs adults	**3.72**	**<0.001**	**0.762**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167.

Table 13.54. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI eyes.*

Source	*df*	MS	*F*	*p*	η_p^2
Age	**2**	**0.101**	**4.58**	**0.12**	**0.060**
Group	**1**	**0.100**	**4.54**	**0.035**	**0.031**
Age x Group	**2**	**0.093**	**4.22**	**0.017**	**0.055**
Error	144	0.022			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.55. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI bridge.*

Source	*df*	MS	*F*	*p*	η_p^2
Age	2	0.006	2.82	0.063	0.038
Group	1	0.001	0.65	0.422	0.004
Age x Group	2	0.002	0.71	0.491	0.010
Error	144	0.002			

Note: MS = Mean squares, effect size = η_p^2.

Table 13.56. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI nose.*

Source	*df*	MS	*F*	*p*	η_p^2
Age	**2**	**0.062**	**7.54**	**0.001**	**0.095**
Group	**1**	**0.053**	**6.44**	**0.012**	**0.043**
Age x Group	2	0.015	1.85	0.162	0.025
Error	144	0.008			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.57. *Summary table of two-way ANOVA (Age x Group) for White-British faces: ROI mouth.*

Source	*df*	MS	*F*	*p*	η_p^2
Age	**2**	**0.164**	**7.63**	**0.001**	**0.096**
Group	**1**	**0.303**	**14.12**	**<0.001**	**0.089**
Age x Group	2	0.023	1.07	0.345	0.015
Error	144	0.021			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.58. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for Japanese participants viewing face stimuli of Japanese ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	2	0.002	1.91	0.152	0.028
Error (Stimulus Condition)	132	0.001			
ROI	**2.18**	**3.104**	**37.10**	**<0.001**	**0.360**
Error (ROI)	144.08	0.084			
Stimulus Condition x ROI	**3.98**	**1.611**	**80.65**	**<0.001**	**0.550**
Error (Stimulus Condition x ROI)	262.88	0.020			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.59. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for British participants viewing face stimuli of Japanese ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	1.63	0.009	2.92	0.068	0.034
Error (Stimulus Condition)	133.28	0.003			
ROI	**1.46**	**7.934**	**54.58**	**<0.001**	**0.400**
Error (ROI)	120.05	0.145			
Stimulus Condition x ROI	**4.07**	**1.616**	**88.84**	**<0.001**	**0.520**
Error (Stimulus Condition x ROI)	333.73	0.018			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.60. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for Japanese participants viewing face stimuli of White-British ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**2**	**0.013**	**6.45**	**0.002**	**0.089**
Error (Stimulus Condition)	132	0.002			
ROI	**1.93**	**4.670**	**44.65**	**<0.001**	**0.404**
Error (ROI)	127.33	0.105			
Stimulus Condition x ROI	**3.99**	**1.588**	**73.25**	**<0.001**	**0.526**
Error (Stimulus Condition x ROI)	262.26	0.022			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.61. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for British participants viewing face stimuli of White-British ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**1.63**	**0.013**	**6.33**	**0.004**	**0.072**
Error (Stimulus Condition)	133.27	0.002			
ROI	**1.53**	**8.563**	**63.18**	**<0.001**	**0.435**
Error (ROI)	125.14	0.136			
Stimulus Condition x ROI	**3.66**	**2.542**	**115.99**	**<0.001**	**0.586**
Error (Stimulus Condition x ROI)	300.49	0.022			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.62. *Summary table of the one-way ANOVA with factor Stimulus Condition using dwell time data.*

Face Ethnicity	Group	ROI	MS	*F*	*p*	η^2
Japanese	British	**Eyes**	**1.226**	**101.81**	**<0.001**	**0.554**
		Bridge	**0.020**	**5.97**	**0.007**	**0.068**
		Nose	0.006	0.74	0.461	0.009
		Mouth	**2.049**	**114.31**	**<0.001**	**0.582**
	Japanese	**Eyes**	**1.501**	**102.89**	**<0.001**	**0.609**
		Bridge	**0.025**	**9.09**	**<0.001**	**0.121**
		Nose	**0.040**	**5.60**	**0.005**	**0.078**
		Mouth	**1.645**	**100.08**	**<0.001**	**0.603**
White-British	British	**Eyes**	**2.266**	**127.17**	**<0.001**	**0.608**
		Bridge	0.002	0.73	0.472	0.009
		Nose	0.007	1.27	0.283	0.015
		Mouth	**2.667**	**146.16**	**<0.001**	**0.641**
	Japanese	**Eyes**	**1.639**	**98.41**	**<0.001**	**0.599**
		Bridge	0.009	2.85	0.069	0.041
		Nose	**0.079**	**8.22**	**<0.001**	**0.111**
		Mouth	**1.819**	**90.83**	**<0.001**	**0.579**

Note: MS = Mean squares, effect size = η^2. Rows in bold indicate significant effect at 0.05.

Table 13.63. *Summary table of paired t-tests contrasting each stimulus condition: British participants.*

Face Ethnicity	ROI	Contrast	*t*	*p*	Cohen's *d*
Japanese	Eyes	**Stat vs Dyn-neut**	**13.07**	**<0.001**	**1.242**
		Stat vs Dyn-soc	**10.05**	**<0.001**	**0.912**
		Dyn-neut vs Dyn-soc	**-3.12**	**0.002**	**0.263**
	Mouth	**Stat vs Dyn-neut**	**-14.36**	**<0.001**	**1.455**
		Stat vs Dyn-soc	**-11.01**	**<0.001**	**1.066**
		Dyn-neut vs Dyn-soc	**3.91**	**<0.001**	**0.341**
White-British	Eyes	**Stat vs Dyn-neut**	**14.39**	**<0.001**	**1.525**
		Stat vs Dyn-soc	**10.88**	**<0.001**	**1.178**
		Dyn-neut vs Dyn-soc	**-3.31**	**0.001**	**0.288**
	Mouth	**Stat vs Dyn-neut**	**-16.20**	**<0.001**	**1.682**
		Stat vs Dyn-soc	**-11.79**	**<0.001**	**1.227**
		Dyn-neut vs Dyn-soc	**4.70**	**<0.001**	**0.393**

Table 13.64. *Summary table of paired t-tests contrasting each stimulus condition: Japanese participants.*

Face Ethnicity	ROI	Contrast	*t*	*p*	Cohen's *d*
Japanese	Eyes	**Stat vs Dyn-neut**	**12.93**	**<0.001**	**1.553**
		Stat vs Dyn-soc	**9.78**	**<0.001**	**1.042**
		Dyn-neut vs Dyn-soc	**-4.37**	**<0.001**	**0.472**
	Mouth	**Stat vs Dyn-neut**	**-14.45**	**<0.001**	**2.120**
		Stat vs Dyn-soc	**-8.11**	**<0.001**	**1.088**
		Dyn-neut vs Dyn-soc	**6.10**	**<0.001**	**0.744**
White-British	Eyes	**Stat vs Dyn-neut**	**13.23**	**<0.001**	**1.358**
		Stat vs Dyn-soc	**9.84**	**<0.001**	**1.113**
		Dyn-neut vs Dyn-soc	-2.13	0.037	0.202
	Mouth	**Stat vs Dyn-neut**	**-11.78**	**<0.001**	**1.756**
		Stat vs Dyn-soc	**-8.95**	**<0.001**	**1.158**
		Dyn-neut vs Dyn-soc	**6.12**	**<0.001**	**0.685**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Table 13.65. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 10-month-olds viewing face stimuli of Japanese ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	2	<0.001	0.193	0.824	0.004
Error (Stimulus Condition)	94	0.002			
ROI	**1.41**	**5.943**	**45.60**	**<0.001**	**0.492**
Error (ROI)	66.03	0.130			
Stimulus Condition x ROI	**3.93**	**0.909**	**53.62**	**<0.001**	**0.533**
Error (Stimulus Condition x ROI)	184.78	0.017			

Table 13.66. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 15- to 17-month-olds viewing face stimuli of Japanese ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	2	0.003	3.03	0.054	0.070
Error (Stimulus Condition)	80	0.001			
ROI	**1.31**	**5.389**	**35.40**	**<0.001**	**0.470**
Error (ROI)	52.29	0.152			
Stimulus Condition x ROI	**3.40**	**1.155**	**60.122**	**<0.001**	**0.600**
Error (Stimulus Condition x ROI)	135.86	0.019			

Table 13.67. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for adults viewing face stimuli of Japanese ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	1.56	0.015	5.70	0.008	0.087
Error (Stimulus Condition)	93.72	0.003			
ROI	**2.43**	**1.87**	**23.94**	**<0.001**	**0.285**
Error (ROI)	145.97	0.078			
Stimulus Condition x ROI	**4.04**	**1.388**	**61.42**	**<0.001**	**0.506**
Error (Stimulus Condition x ROI)	242.15	0.023			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.68. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 10-month-olds viewing face stimuli of White-British ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	2	0.003	2.09	0.130	0.043
Error (Stimulus Condition)	94	0.002			
ROI	**1.44**	**6.108**	**45.91**	**<0.001**	**0.494**
Error (ROI)	67.88	0.133			
Stimulus Condition x ROI	**4.06**	**1.226**	**64.64**	**<0.001**	**0.579**
Error (Stimulus Condition x ROI)	190.62	0.019			

Table 13.69. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 15- to 17-month-olds viewing face stimuli of White-British ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	1.56	0.007	3.35	0.053	0.077
Error (Stimulus Condition)	62.25	0.002			
ROI	**1.28**	**6.276**	**36.49**	**<0.001**	**0.477**
Error (ROI)	51.30	0.172			
Stimulus Condition x ROI	**3.22**	**1.551**	**65.81**	**<0.001**	**0.622**
Error (Stimulus Condition x ROI)	128.69	0.024			

Table 13.70. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for adults viewing face stimuli of White-British ethnicity.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**1.69**	**0.024**	**9.42**	**<0.001**	**0.136**
Error (Stimulus Condition)	101.61	0.003			
ROI	**2.32**	**2.712**	**33.30**	**<0.001**	**0.357**
Error (ROI)	139.25	0.081			
Stimulus Condition x ROI	**3.83**	**1.542**	**66.45**	**<0.001**	**0.522**
Error (Stimulus Condition x ROI)	229.99	0.024			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.71. *Summary table of the one-way ANOVA with factor Stimulus Condition using dwell time data.*

Face Ethnicity	Group	ROI	MS	*F*	*p*	η^2
Japanese	10 months	**Eyes**	**0.976**	**76.30**	**<0.001**	**0.619**
		Bridge	0.001	0.73	0.483	0.015
		Nose	0.006	0.91	0.406	0.019
		Mouth	**0.804**	**56.80**	**<0.001**	**0.547**
	15 to 17 months	**Eyes**	**0.751**	**69.76**	**<0.001**	**0.636**
		Bridge	0.009	3.30	0.056	0.076
		Nose	0.011	1.65	0.205	0.039
		Mouth	**1.198**	**78.51**	**<0.001**	**0.662**
	Adults	**Eyes**	**1.089**	**64.15**	**<0.001**	**0.517**
		Bridge	**0.053**	**16.24**	**<0.001**	**0.213**
		Nose	0.005	0.57	0.569	0.009
		Mouth	**1.764**	**87.75**	**<0.001**	**0.594**
White-British	10 months	**Eyes**	**1.386**	**101.14**	**<0.001**	**0.683**
		Bridge	0.005	2.87	0.062	0.057
		Nose	0.009	1.16	0.318	0.024
		Mouth	**1.089**	**65.31**	**<0.001**	**0.582**
	15 to 17 months	**Eyes**	**1.336**	**62.39**	**<0.001**	**0.609**
		Bridge	0.009	3.09	0.065	0.072
		Nose	0.009	1.82	0.169	0.043
		Mouth	**1.516**	**91.35**	**<0.001**	**0.695**
	Adults	**Eyes**	**1.187**	**72.43**	**<0.001**	**0.547**
		Bridge	0.011	3.10	0.054	0.049
		Nose	**0.068**	**7.88**	**0.001**	**0.116**
		Mouth	**1.710**	**89.75**	**<0.001**	**0.599**

Note: MS = Mean squares, effect size = η^2. Rows in bold indicate significant effect at 0.05.

Table 13.72. *Summary table of paired t-tests contrasting each stimulus condition: 10-month-olds.*

Face Ethnicity	ROI	Contrast	*t*	*p*	Cohen's *d*
Japanese	Eyes	**Stat vs Dyn-neut**	**12.18**	**<0.001**	**1.399**
		Stat vs Dyn-soc	**8.31**	**<0.001**	**0.997**
		Dyn-neut vs Dyn-soc	**-2.79**	**0.008**	**0.309**
	Mouth	**Stat vs Dyn-neut**	**-10.15**	**<0.001**	**1.482**
		Stat vs Dyn-soc	**-8.85**	**<0.001**	**1.089**
		Dyn-neut vs Dyn-soc	2.45	0.018	0.324
White-British	Eyes	**Stat vs Dyn-neut**	**12.65**	**<0.001**	**1.404**
		Stat vs Dyn-soc	**10.97**	**<0.001**	**1.279**
		Dyn-neut vs Dyn-soc	-1.21	0.232	0.128
	Mouth	**Stat vs Dyn-neut**	**-9.85**	**<0.001**	**1.711**
		Stat vs Dyn-soc	**-9.49**	**<0.001**	**1.439**
		Dyn-neut vs Dyn-soc	**2.73**	**0.009**	**0.346**

Table 13.73. *Summary table of paired t-tests contrasting each stimulus condition: 15- to 17-month-olds.*

Face Ethnicity	ROI	Contrast	*t*	*p*	Cohen's *d*
Japanese	Eyes	**Stat vs Dyn-neut**	**9.63**	**<0.001**	**1.416**
		Stat vs Dyn-soc	**10.47**	**<0.001**	**1.140**
		Dyn-neut vs Dyn-soc	-1.62	0.113	0.231
	Mouth	**Stat vs Dyn-neut**	**-13.58**	**<0.001**	**1.591**
		Stat vs Dyn-soc	**-8.79**	**<0.001**	**1.100**
		Dyn-neut vs Dyn-soc	**2.78**	**0.008**	**0.355**
White-British	Eyes	**Stat vs Dyn-neut**	**9.00**	**<0.001**	**1.434**
		Stat vs Dyn-soc	**7.83**	**<0.001**	**1.210**
		Dyn-neut vs Dyn-soc	-1.99	0.054	0.196
	Mouth	**Stat vs Dyn-neut**	**-12.94**	**<0.001**	**1.654**
		Stat vs Dyn-soc	**-9.09**	**<0.001**	**1.240**
		Dyn-neut vs Dyn-soc	**3.36**	**0.002**	**0.362**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Table 13.74. *Summary table of paired t-tests contrasting each stimulus condition: Adults.*

Face Ethnicity	ROI	Contrast	*t*	*p*	Cohen's *d*
Japanese	Eyes	**Stat vs Dyn-neut**	**10.14**	**<0.001**	**1.455**
		Stat vs Dyn-soc	**7.05**	**<0.001**	**0.899**
		Dyn-neut vs Dyn-soc	**-4.46**	**<0.001**	**0.508**
	Mouth	**Stat vs Dyn-neut**	**-12.86**	**<0.001**	**2.026**
		Stat vs Dyn-soc	**-6.89**	**<0.001**	**1.029**
		Dyn-neut vs Dyn-soc	**6.57**	**<0.001**	**0.749**
White-British	Eyes	**Stat vs Dyn-neut**	**12.31**	**<0.001**	**1.595**
		Stat vs Dyn-soc	**7.52**	**<0.001**	**1.036**
		Dyn-neut vs Dyn-soc	**-3.36**	**0.001**	**0.396**
	Mouth	**Stat vs Dyn-neut**	**-12.16**	**<0.001**	**1.857**
		Stat vs Dyn-soc	**-7.40**	**<0.001**	**1.018**
		Dyn-neut vs Dyn-soc	**6.84**	**<0.001**	**0.757**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Table 13.75. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for Japanese participants.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**1.77**	**0.012**	**8.66**	**0.001**	**0.116**
Error (Stimulus Condition)	116.75	0.001			
ROI	**2.05**	**4.813**	**50.57**	**<0.001**	**0.434**
Error (ROI)	135.41	0.095			
Stimulus Condition x ROI	**3.85**	**1.820**	**105.38**	**<0.001**	**0.615**
Error (Stimulus Condition x ROI)	253.94	0.017			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.76. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for British participants.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	1.57	0.006	2.24	0.122	0.027
Error (Stimulus Condition)	128.87	0.003			
ROI	**1.49**	**9.755**	**62.69**	**<0.001**	**0.433**
Error (ROI)	121.82	0.156			
Stimulus Condition x ROI	**4.01**	**2.110**	**105.23**	**<0.001**	**0.562**
Error (Stimulus Condition x ROI)	328.90	0.020			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.77. *Summary table of the one-way ANOVA with factor Stimulus Condition using fixation data.*

Group	ROI	MS	*F*	*p*	η^2
British	**Eyes**	**2.039**	**116.45**	**<0.001**	**0.587**
	Bridge	0.002	0.67	0.497	0.008
	Nose	0.017	2.74	0.067	0.032
	Mouth	**2.433**	**139.98**	**<0.001**	**0.631**
Japanese	**Eyes**	**1.681**	**146.73**	**<0.001**	**0.690**
	Bridge	**0.020**	**11.94**	**<0.001**	**0.153**
	Nose	**0.036**	**5.17**	**0.007**	**0.073**
	Mouth	**1.774**	**123.93**	**<0.001**	**0.653**

Note: MS = Mean squares, effect size = η^2. Rows in bold indicate significant effect at 0.05.

Table 13.78. *Summary table of paired t-tests contrasting each stimulus condition: British participants.*

ROI	Contrast	*t*	*p*	Cohen's *d*
Eyes	**Stat vs Dyn-neut**	**13.89**	**<0.001**	**1.320**
	Stat vs Dyn-soc	**10.78**	**<0.001**	**1.108**
	Dyn-neut vs Dyn-soc	-1.81	0.074	0.155
Mouth	**Stat vs Dyn-neut**	**-16.46**	**<0.001**	**1.524**
	Stat vs Dyn-soc	**-11.64**	**<0.001**	**1.174**
	Dyn-neut vs Dyn-soc	**3.79**	**<0.001**	**0.309**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Table 13.79. *Summary table of paired t-tests contrasting each stimulus condition: Japanese participants.*

ROI	Contrast	t	p	Cohen's d
Eyes	**Stat vs Dyn-neut**	**15.29**	**<0.001**	**1.523**
	Stat vs Dyn-soc	**12.39**	**<0.001**	**1.239**
	Dyn-neut vs Dyn-soc	**-2.90**	**0.005**	**0.254**
Mouth	**Stat vs Dyn-neut**	**-14.59**	**<0.001**	**2.128**
	Stat vs Dyn-soc	**-9.34**	**<0.001**	**1.256**
	Dyn-neut vs Dyn-soc	**7.04**	**<0.001**	**0.704**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Table 13.80. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 10-month-olds.*

Source	df	MS	F	p	η_p^2
Stimulus Condition	**1.12**	**0.125**	**6.00**	**0.015**	**0.113**
Error (Stimulus Condition)	52.65	0.021			
ROI	**1.59**	**3.63**	**50.02**	**<0.001**	**0.516**
Error (ROI)	74.60	0.072			
Stimulus Condition x ROI	**3.60**	**2.241**	**75.64**	**<0.001**	**0.617**
Error (Stimulus Condition x ROI)	169.21	0.033			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.81. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for 15- to 17-month-olds.*

Source	df	MS	F	p	η_p^2
Stimulus Condition	**2**	**0.005**	**4.89**	**0.020**	**0.093**
Error (Stimulus Condition)	80	0.001			
ROI	**1.34**	**6.380**	**35.34**	**<0.001**	**0.469**
Error (ROI)	53.39	0.181			
Stimulus Condition x ROI	**3.84**	**1.289**	**71.56**	**<0.001**	**0.641**
Error (Stimulus Condition x ROI)	153.57	0.018			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.82. *Summary table of the two-factor ANOVA ROI x Stimulus Condition for adults.*

Source	*df*	MS	*F*	*p*	η_p^2
Stimulus Condition	**1.48**	**0.027**	**11.99**	**<0.001**	**0.167**
Error (Stimulus Condition)	88.71	0.002			
ROI	**2.34**	**2.821**	**33.84**	**<0.001**	**0.361**
Error (ROI)	140.18	0.083			
Stimulus Condition x ROI	**3.72**	**1.581**	**79.48**	**<0.001**	**0.570**
Error (Stimulus Condition x ROI)	223.06	0.020			

Note: MS = Mean squares, effect size = η_p^2. Rows in bold indicate significant effect at 0.05.

Table 13.83. *Summary table of the one-way ANOVA with factor Stimulus Condition using fixation data.*

Group	ROI	MS	*F*	*p*	η^2
10 months	**Eyes**	**1.489**	**104.18**	**<0.001**	**0.689**
	Bridge	0.002	0.803	0.451	0.017
	Nose	0.004	0.644	0.527	0.014
	Mouth	**0.989**	**66.65**	**<0.001**	**0.586**
15 to 17 months	**Eyes**	**1.164**	**81.34**	**<0.001**	**0.670**
	Bridge	0.007	1.64	0.207	0.039
	Nose	0.013	2.29	0.108	0.054
	Mouth	**1.710**	**99.69**	**<0.001**	**0.714**
Adults	**Eyes**	**1.252**	**85.71**	**<0.001**	**0.588**
	Bridge	**0.028**	**10.41**	**<0.001**	**0.148**
	Nose	0.020	2.77	0.066	0.044
	Mouth	**1.822**	**110.69**	**<0.001**	**0.648**

Note: MS = Mean squares, effect size = η^2. Rows in bold indicate significant effect at 0.05.

Table 13.84. *Summary table of paired t-tests contrasting stimulus conditions: 10-month-olds.*

ROI	Contrast	*t*	*p*	Cohen's *d*
Eyes	**Stat vs Dyn-neut**	**13.21**	**<0.001**	**1.485**
	Stat vs Dyn-soc	**11.22**	**<0.001**	**1.400**
	Dyn-neut vs Dyn-soc	-0.07	0.948	0.007
Mouth	**Stat vs Dyn-neut**	**-10.22**	**<0.001**	**1.531**
	Stat vs Dyn-soc	**-9.54**	**<0.001**	**1.300**
	Dyn-neut vs Dyn-soc	1.95	0.057	0.237

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Table 13.85. *Summary table of paired t-tests contrasting stimulus conditions: 15- to 17-month-olds.*

ROI	Contrast	*t*	*p*	Cohen's *d*
Eyes	**Stat vs Dyn-neut**	**10.28**	**<0.001**	**1.341**
	Stat vs Dyn-soc	**9.69**	**<0.001**	**1.194**
	Dyn-neut vs Dyn-soc	-1.09	0.282	0.118
Mouth	**Stat vs Dyn-neut**	**-15.72**	**<0.001**	**1.669**
	Stat vs Dyn-soc	**-8.89**	**<0.001**	**1.203**
	Dyn-neut vs Dyn-soc	**3.19**	**0.003**	**0.341**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Table 13.86. *Summary table of paired t-tests contrasting stimulus conditions: Adults.*

ROI	Contrast	*t*	*p*	Cohen's *d*
Eyes	**Stat vs Dyn-neut**	**11.97**	**<0.001**	**1.573**
	Stat vs Dyn-soc	**8.11**	**<0.001**	**1.014**
	Dyn-neut vs Dyn-soc	**-4.64**	**<0.001**	**0.456**
Mouth	**Stat vs Dyn-neut**	**-14.08**	**<0.001**	**2.071**
	Stat vs Dyn-soc	**-7.82**	**<0.001**	**1.099**
	Dyn-neut vs Dyn-soc	**7.50**	**<0.001**	**0.742**

Note: Rows in bold indicate significant effect at Bonferroni-corrected alpha of 0.0167; stat = static; dyn-neut = dynamic-neutral; dyn-soc = dynamic-social).

Appendix G

Pearson's correlation coefficients for the relationship between disengagement latencies and mouth looking for the data presented in Chapter 5.

Table 14.1. *Pearson's correlation coefficients and corresponding p-values for the relationship between disengagement latencies and proportional mouth looking using dwell time data.*

Age group	Variables	Japanese	British
10 months (Japanese N = 22; British N = 26)	Disengagement – Mouth (Japanese, static)	r = -0.301, p = 0.174	r = -0.097, p = 0.637
	Disengagement – Mouth (White-British, static)	r = -0.206, p = 0.357	r = -0.086, p = 0.677
	Disengagement – Mouth (Japanese, dynamic-neutral)	r = -0.060, p = 0.792	r = -0.255, p = 0.208
	Disengagement – Mouth (White-British, dynamic-neutral)	r = 0.261, p = 0.240	r = -0.086, p = 0.677
	Disengagement – Mouth (Japanese, dynamic-social)	r = -0.067, p = 0.767	r = -0.130, p = 0.526
	Disengagement – Mouth (White-British, dynamic-social)	r = -0.344, p = 0.117	r = -0.271, p = 0.181
15 to 17 months (Japanese N = 15; British N = 26)	Disengagement – Mouth (Japanese, static)	r = 0.097, p = 0.731	r = 0.047, p = 0.821
	Disengagement – Mouth (White-British, static)	r = 0.032, p = 0.910	r = 0.001, p = 0.995
	Disengagement – Mouth (Japanese, dynamic-neutral)	r = -0.090, p = 0.749	r = -0.045, p = 0.828
	Disengagement – Mouth (White-British, dynamic-neutral)	r = -0.264, p = 0.341	r = 0.105, p = 0.609
	Disengagement – Mouth (Japanese, dynamic-social)	r = 0.122, p = 0.664	r = 0.084, p = 0.685
	Disengagement – Mouth (White-British, dynamic-social)	r = -0.157, p = 0.576	r = 0.059, p = 0.776

Age group	Variables	Japanese	British
Adults (Japanese N = 30; British N = 31)	Disengagement – Mouth (Japanese, static)	r = -0.048, p = 0.800	r = -0.239, p = 0.196
	Disengagement – Mouth (White-British, static)	r = 0.077, p = 0.687	r = -0.107, p = 0.567
	Disengagement – Mouth (Japanese, dynamic-neutral)	r = 0.086, p = 0.652	r = 0.282, p = 0.125
	Disengagement – Mouth (White-British, dynamic-neutral)	r = -0.182, p = 0.335	r = 0.074, p = 0.691
	Disengagement – Mouth (Japanese, dynamic-social)	r = -0.118, p = 0.535	r = 0.021, p = 0.912
	Disengagement – Mouth (White-British, dynamic-social)	r = -0.161, p = 0.394	r = 0.009, p = 0.963

Table 14.2. *Pearson's correlation coefficients and corresponding p-values for the relationship between disengagement latencies and proportional mouth looking using fixation time data.*

Age group	Variables	Japanese	British
10 months	Disengagement – Mouth (static)	r = -0.406, p = 0.061	r = 0.020, p = 0.924
(Japanese N = 22;	Disengagement – Mouth (dynamic-neutral)	r = -0.014, p = 0.950	r = -0.012, p = 0.954
British N = 26)	Disengagement – Mouth (dynamic-social)	r = -0.319, p = 0.148	r = -0.160, p = 0.436
15 to 17 months	Disengagement – Mouth (static)	r = -0.120, p = 0.669	r = -0.030, p = 0.884
(Japanese N = 15;	Disengagement – Mouth (dynamic-neutral)	r = -0.167, p = 0.552	r = 0.005, p = 0.979
British N = 26)	Disengagement – Mouth (dynamic-social)	r = -0.022, p = 0.937	r = 0.106, p = 0.607
Adults	Disengagement – Mouth (static)	r = 0.116, p = 0.541	r = -0.133, p = 0.476
(Japanese N = 30;	Disengagement – Mouth (dynamic-neutral)	r = -0.084, p = 0.658	r = 0.168, p = 0.367
British N = 31)	Disengagement – Mouth (dynamic-social)	r = -0.139, p = 0.463	r = 0.025, p = 0.893

Pearson's correlation coefficients for the relationship between the performance on the cognitive control task and mouth looking for the data presented in Chapter 5.

Table 14.3. *Pearson's correlation coefficients and corresponding p-values for the relationship between cognitive control task performance and proportional mouth looking using dwell time data.*

Age group	Variables	Japanese	British
10 months (Japanese N = 19; British N = 21)	CCT – Mouth (Japanese, static)	r = 0.190, p = 0.436	r = -0.445, p = 0.043
	CCT – Mouth (White-British, static)	r = 0.318, p = 0.184	r = -0.258, p = 0.259
	CCT – Mouth (Japanese, dynamic-neutral)	r = 0.374, p = 0.115	r = -0.414, p = 0.062
	CCT – Mouth (White-British, dynamic-neutral)	r = 0.399, p = 0.090	r = -0.391, p = 0.080
	CCT – Mouth (Japanese, dynamic-social)	r = 0.318, p = 0.184	r = -0.290, p = 0.201
	CCT – Mouth (White-British, dynamic-social)	r = 0.475, p = 0.040	r = -0.172, p = 0.457
15 to 17 months (Japanese N = 15; British N = 24)	CCT – Mouth (Japanese, static)	r = 0.297, p = 0.282	r = -0.230, p = 0.280
	CCT – Mouth (White-British, static)	r = 0.407, p = 0.132	r = -0.241, p = 0.256
	CCT – Mouth (Japanese, dynamic-neutral)	r = 0.438, p = 0.103	r = -0.183, p = 0.392
	CCT – Mouth (White-British, dynamic-neutral)	r = 0.125, p = 0.656	r = -0.201, p = 0.347
	CCT – Mouth (Japanese, dynamic-social)	r = 0.405, p = 0.135	r = -0.119, p = 0.580
	CCT – Mouth (White-British, dynamic-social)	r = 0.293, p = 0.289	r = -0.189, p = 0.375

Age group	Variables	Japanese	British
Adults (Japanese N = 30; British N = 31)	CCT – Mouth (Japanese, static)	r = 0.202, p = 0.285	r = 0.006, p = 0.976
	CCT – Mouth (White-British, static)	r = 0.357, p = 0.053	r = 0.078, p = 0.676
	CCT – Mouth (Japanese, dynamic-neutral)	r = 0.410, p = 0.024	r = 0.058, p = 0.756
	CCT – Mouth (White-British, dynamic-neutral)	r = 0.178, p = 0.345	r = 0.333, p = 0.067
	CCT – Mouth (Japanese, dynamic-social)	r = 0.222, p = 0.239	r = -0.083, p = 0.655
	CCT – Mouth (White-British, dynamic-social)	r = 0.224, p = 0.235	r = 0.010, p = 0.958

Table 14.4. *Pearson's correlation coefficients and corresponding p-values for the relationship between cognitive control task performance and proportional mouth looking using fixation time data.*

Age group	Variables	Japanese	British
10 months	CCT – Mouth (static)	$r = 0.207$, $p = 0.395$	$r = -0.329$, $p = 0.146$
(Japanese $N = 22$;	CCT – Mouth (dynamic-neutral)	$r = 0.474$, $p = 0.040$	$r = -0.321$, $p = 0.156$
British $N = 26$)	CCT – Mouth (dynamic-social)	$r = 0.323$, $p = 0.178$	$r = -0.145$, $p = 0.531$
15 to 17 months	CCT – Mouth (static)	$r = 0.376$, $p = 0.168$	$r = -0.252$, $p = 0.235$
(Japanese $N = 15$;	CCT – Mouth (dynamic-neutral)	$r = 0.234$, $p = 0.401$	$r = -0.255$, $p = 0.228$
British $N = 24$)	CCT – Mouth (dynamic-social)	$r = 0.385$, $p = 0.156$	$r = -0.142$, $p = 0.507$
Adults	CCT – Mouth (static)	$r = 0.316$, $p = 0.088$	$r = -0.019$, $p = 0.921$
(Japanese $N = 30$;	CCT – Mouth (dynamic-neutral)	$r = 0.316$, $p = 0.088$	$r = 0.104$, $p = 0.577$
British $N = 31$)	CCT – Mouth (dynamic-social)	$r = 0.211$, $p = 0.263$	$r = -0.122$, $p = 0.513$