



Gosetti, Valentina (2025) *In the face of diversity: revealing the influence of ethnicity and culture on social trait face perception*. PhD thesis.

<https://theses.gla.ac.uk/85597/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

**In the face of diversity:
Revealing the influence of ethnicity and culture
on social trait face perception**

Valentina Gosetti

B.Sc. Psychology

M.Sc. Research Methods in Psychology

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Psychology and Neuroscience
College of Medical, Veterinary & Life Sciences
University of Glasgow



August 2025

Abstract

On a daily basis, people make spontaneous judgements about who to trust or avoid based on facial appearance. Because of their considerable downstream consequences, a longstanding goal has been to understand which facial features drive social trait judgements. Despite emerging evidence showing that ethnic and cultural diversity influence social trait perception, most knowledge remains centred on White Western observers perceiving White faces. This bias questions the generalizability of prominent theories and feature-based models. In this thesis, I combine a data-driven reverse correlation approach with a generative model of 3D human faces to model the specific facial features of 3D shape and 2D complexion that drive perceptions of trustworthiness and dominance from three face ethnicities—Black African, East Asian, and White European—in two observer cultures—White Western and East Asian. Using information-theoretic analyses, I show that both White Western and East Asian observers perceive trustworthiness and dominance from a core set of facial features which are shared across face ethnicities, and map onto previous findings, plus novel face ethnicity-specific variations. These variations challenge the generalizability of prominent feature-based models and characterize the causal influence of face ethnicity on social trait perception. Further, while conceptually similar, the results for White Western and East Asian observers comprise different facial features. To formally test these differences, I next examine the cultural specificity of the modelled facial features across face ethnicities. Results show that, while White Western and East Asian observers provide similar social trait ratings for each face ethnicity, the features they base their ratings on differ. This questions previous claims of universality based only on rating comparisons. Further analyses reveal that the facial features specific to Western culture resemble specific emotion cues (e.g., smiling, frowning), whereas those specific to East Asian culture do not. This contrasts prominent theories such as emotion overgeneralization and highlights the Western-centric bias of current knowledge. Finally, I use a machine-learning approach and information-theoretic analyses to examine how face ethnicity, observer culture, and their synergistic interaction causally influence social trait perception. Results show that, across face ethnicities and observer cultures, social trait perception is driven by four feature sets: those that are shared, those that are face ethnicity-specific, those that are culture-specific, and those that are synergistic. Subsequent examinations of these feature sets confirm that they represent key sources of variance in social trait perception. These findings extend current efforts to quantify the relative contributions of the face, the observer, and their interaction and offer direct empirical support for modern theories of social trait perception. Together, this thesis responds to mounting calls to

diversify psychological science by showing that ethnic and cultural diversity systematically alter the causal facial features for perception of key social traits, with direct implications for current knowledge and theory development.

Contents

Abstract	i
List of Tables	vii
List of Figures	x
Publications	xi
Acknowledgements	xii
Declaration	xiv
List of abbreviations	xv
1 General introduction	1
1.1 Social trait perception	1
1.1.1 Prevalence	2
1.1.2 Consequences	2
1.1.3 Accuracy	3
1.2 Theories and models of social trait perception	4
1.2.1 Dominance-trustworthiness model	5
1.2.2 Ecological theory and the overgeneralization hypothesis	7
1.2.3 Stereotype Content Model	8
1.2.4 Social vision and the shared signal hypothesis	9
1.2.5 Dynamic Interactive model	11
1.2.6 Trait Inference Mapping framework	12
1.3 Drivers of social trait perception	14
1.3.1 3D face shape	14
1.3.2 2D face complexion	17
1.3.3 Changeable facial features	19
1.3.4 Other facial features	20
1.3.5 Non-facial features	21
1.4 Social trait perception knowledge is White- and Western-centric	23
1.4.1 Ethnic diversity in the face	23
1.4.2 Ethnic and cultural diversity in the observer	25

1.5	Data-driven methods to diversify knowledge of social trait perception . . .	26
1.5.1	Reverse correlation	27
1.5.2	Generative model of 3D human faces	28
1.5.3	Per-observer modelling	28
1.6	The present work	29
2	Ethnic diversity in facial features challenges generalizability of social trait perception models	31
2.1	Introduction	32
2.2	Methods	33
2.2.1	Observers	35
2.2.2	Stimulus generation	35
2.2.3	Experimental procedure	36
2.2.4	Modelling procedure	36
2.2.5	Model validation	37
2.2.6	Model visualization	37
2.2.7	Population prevalence	38
2.3	Results	38
2.3.1	Social trait face perception is driven by shared plus specific facial features across face ethnicities	41
2.3.2	Validating ethnic variance in the facial features that drive social trait judgements in White Western observers	46
2.4	Discussion	49
3	Ethnic variance in social trait facial features replicates in East Asian culture	53
3.1	Introduction	54
3.2	Methods	55
3.2.1	Observers	55
3.2.2	Stimulus generation	56
3.2.3	Experimental procedure	56
3.2.4	Modelling procedure	56
3.2.5	Model validation	57
3.3	Results	57
3.3.1	East Asian observers use shared plus face ethnicity-specific facial features to perceive social traits	60
3.3.2	Comparing the shared facial features used by East Asian and White Western observers	65
3.3.3	Validating ethnic variance in the facial features that drive social trait judgements in East Asian observers	66
3.4	Discussion	68

4	Cultural variance questions the universality of emotion overgeneralization	71
4.1	Introduction	72
4.2	Methods	73
4.3	Results	73
4.3.1	Social trait perception is driven by culturally shared plus culture-specific facial features	78
4.3.2	The use of emotion cues in social trait perception is specific to Western culture	83
4.4	Discussion	90
5	Face ethnicity and observer culture interact to causally influence social trait perception	93
5.1	Introduction	94
5.2	Methods	95
5.3	Results	95
5.3.1	Facial feature variance predicts differences in face ethnicity and observer culture	99
5.3.2	Measuring the individual and joint influence of face ethnicity and observer culture on the facial features used for social trait perception	103
5.3.3	Face ethnicity and observer culture individually influence social trait perception	105
5.3.4	Face ethnicity and observer culture synergistically influence social trait perception	108
5.3.5	Social trait perception is driven by features shared across face ethnicities and observer cultures	112
5.3.6	Examining the relative contributions of face ethnicity, observer culture, and their combination to social trait perception	113
5.4	Discussion	117
6	General discussion	121
6.1	Main findings and contributions	121
6.2	Limitations	124
6.2.1	Western-centric social trait dimensions	124
6.2.2	Limited number of ethnic and cultural groups	124
6.2.3	Computer-generated vs real faces	125
6.2.4	Lack of context and further considerations for ecological validity	126
6.3	Future directions	127
6.3.1	Examining other facets of diversity and intersectionality	127
6.3.2	Diversity in multimodal social perception	128
6.3.3	Designing ethnically and culturally sensitive virtual agents	129
6.4	Conclusion	130

7	Supplementary materials	131
7.1	Chapter 1	131
7.1.1	Expressivity of the Generative Model of 3D human Faces	131
7.2	Chapter 2	134
7.2.1	Screening questionnaire for White Western observers	134
7.2.2	Determining the appropriate number of experimental trials	134
7.2.3	Checking the linearity assumption to model 3D shape and 2D complexion for White Western observers	135
7.2.4	Cross-validating the 3D face models of White Western observers	138
7.2.5	Number and distribution of shared vs face ethnicity-specific identity components	139
7.2.6	Ethnic phenotypes in the 3D Generative Model of human Faces	142
7.2.7	Identifying above-chance face vertex correlations	143
7.3	Chapter 3	144
7.3.1	Screening questionnaire for East Asian observers	144
7.3.2	Checking the linearity assumption to model 3D shape and 2D complexion for East Asian observers	144
7.3.3	Cross-validating the 3D face models of East Asian observers	147
7.3.4	Number and distribution of shared vs face ethnicity-specific identity components	147
7.4	Chapter 4	150
7.4.1	Number and distribution of culturally shared vs culture-specific identity components used across face ethnicities	150
7.4.2	Relevant 3D face vertices and cartesian dimension of bilateral Action Units	153
7.4.3	Action Unit patterns representing the culture-specific facial features	154
7.5	Chapter 5	156
7.5.1	Number and distribution of face ethnicity-specific identity components across observer cultures	156
7.5.2	Number and distribution of synergistic identity components	157
7.5.3	Number and distribution of shared identity components	160
	References	199

List of Tables

7.1	Number of shared vs specific identity components used by White Western observers	139
7.2	Number of shared vs specific identity components used by East Asian observers	148
7.3	Number of culturally shared vs culture-specific identity components used across face ethnicities	150
7.4	Number of face ethnicity-specific identity components across observer cultures	156
7.5	Number of synergistic identity components for male faces	158
7.6	Number of synergistic identity components for female faces	159
7.7	Number of identity components shared across face ethnicities and observer cultures	161

List of Figures

2.1	Modelling the 3D facial features that drive social trait perception across different face ethnicities	34
2.2	Facial features that drive social trait perception from male faces of different ethnicities	39
2.3	Facial features that drive social trait perception from female faces of different ethnicities	41
2.4	Social trait facial features that are shared vs specific to face ethnicity for male faces	43
2.5	Social trait facial features that are shared vs specific to face ethnicity for female faces	45
2.6	Face ethnicity influences which features of male faces are diminished or exaggerated for social trait perception	47
2.7	Face ethnicity influences which features of female faces are diminished or exaggerated for social trait perception	49
3.1	Facial features that drive social trait perception from male faces of different ethnicities	58
3.2	Facial features that drive social trait perception from female faces of different ethnicities	60
3.3	Social trait facial features that are shared vs specific to face ethnicity for male faces	61
3.4	Social trait facial features that are shared vs specific to face ethnicity for female faces	64
3.5	Correlation between the shared features used by East Asian and White Western observers	65
3.6	Correlation between the face ethnicity-specific features used by East Asian and White Western observers	67
4.1	Relationship between social trait ratings given by East Asian and White Western observers	74
4.2	Facial features that drive social trait perception across cultures from male faces of different ethnicities	76

4.3	Facial features that drive social trait perception across cultures from female faces of different ethnicities	78
4.4	Social trait facial features that are shared vs specific to observer culture across ethnically diverse male faces	80
4.5	Social trait facial features that are shared vs specific to observer culture across ethnically diverse female faces	82
4.6	Representing social trait features as Action Units	84
4.7	Action Unit representation of the male culture-specific social trait facial features and their correlation with basic emotions	86
4.8	Action Unit representation of the female culture-specific social trait facial features and their correlation with basic emotions	89
5.1	Facial features that drive social trait perception across cultures from male faces of different ethnicities	96
5.2	Facial features that drive social trait perception across cultures from female faces of different ethnicities	98
5.3	Ethnicity and culture SVM classification performance for male face models	100
5.4	Ethnicity and culture SVM classification performance for female face models	102
5.5	Identifying the influence of culture, ethnicity, and their combination of the facial features used for social trait perception	103
5.6	Culture-specific and ethnicity-specific facial features of male faces	106
5.7	Culture-specific and ethnicity-specific facial features of female faces	108
5.8	Synergistic facial features used for social trait perception from male faces .	109
5.9	Synergistic facial features used for social trait perception from female faces	111
5.10	Facial features that are shared across observer cultures and face ethnicities .	112
5.11	Proportion of social trait features that are shared, observer culture-specific, face ethnicity-specific, and synergistic	113
5.12	Reconstructing the original social trait models using feature sets	115
5.13	Correlation between the original and reconstructed social trait models	116
7.1	Expressivity of the Generative Model of 3D human Faces (GMF)	132
7.2	Beta coefficients bootstrapped CIs of the pilot 3D shape and 2D complexion models	135
7.3	Linearity of identity components for male faces	136
7.4	Linearity of identity components for female faces	137
7.5	LOOCV results for White Western observers' models	138
7.6	Distribution of shared vs specific identity components used by White Western observers	141
7.7	Ethnic phenotypic facial features	142
7.8	Random correlation (Pearson r) distribution	143
7.9	Linearity of identity components for male faces	145
7.10	Linearity of identity components for female faces	146

7.11 LOOCV results for East Asian observers' models	147
7.12 Distribution of shared vs specific identity components used by East Asian observers	149
7.13 Distribution of culturally shared vs culture-specific identity components used across face ethnicities	152
7.14 Relevant 3D face vertices of bilateral Action Units	154
7.15 Action Unit pattern of the shared social trait facial features	155
7.16 Distribution of face ethnicity-specific identity components across observer cultures	157
7.17 Distribution of synergistic identity components	160
7.18 Distribution of shared identity components across face ethnicities and ob- server cultures	162

Publications

Parts of this thesis have been published as peer-reviewed conference abstracts and proceedings, pre-prints, and/or academic book chapters:

Chapter 1

Gosetti, V., & Jack, R. E. (2023). Using data-driven methods to advance knowledge of social face perception. In U. Hess, R. B., Adams Jr, & R. E. Kleck (Eds.), *Emotion Communication by the Aging Face and Body: A Multidisciplinary View* (pp. 286-299). Cambridge University Press.

Chapter 2

Gosetti, V., Hensel, L. B., Ince, R. A. A., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2024). *Face ethnicity influences which facial features drive social judgements*. Research-Square. <https://doi.org/10.21203/rs.3.rs-4680996/v1>

Gosetti, V., Hensel, L. B., Ince, R. A. A., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2023). In the face of diversity: face ethnicity influences which facial features are used for social trait perception. *Journal of Vision*, 23(9), 5247-5247.

Chapter 6

Gosetti, V. & Jack, R. E. (2024). Using psychological science to diversify IVAs and address bias. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents*.

Acknowledgements

Completing this PhD would not have been possible without the help, guidance, and support of many. I would first like to thank my supervisors, Prof. Rachael Jack and Prof. Philippe Schyns, for sharing their immense knowledge, wisdom, and unwavering dedication to research. I would also like to thank my co-authors, Dr Robin Ince and Dr Oliver Garrod, for sharing their expertise and taking the time to do so repeatedly, and often painstakingly, across my PhD. Finally, I would like to thank Dr Pablo Arias-Sarah for showing me that enthusiasm, kindness, and success are not mutually exclusive.

I would also like to thank past and present members of FaceSyntax: Dr Jonas Nölle, Yichen Wu, Max Christou, and Monica Duta. Your academic feedback, personal insights, and friendship have made these four years much more enjoyable and I am honoured to have shared my PhD journey with you. I would also like to extend a special thank you to my academic big sisters, Dr Laura Hensel and Dr Chaona Chen, who always found the time to help when I was stuck, listen when I was worried, and laugh when I needed it.

I would not be writing this thesis had I not received continuous support from my wonderful group of friends. I particularly owe four people: Beata Ciesluk, Aidan Gold, Sapir Zichner, and Tobias Thejll-Madsen. Beata, who would have guessed that those two ridiculous undergraduates would end up as PhDs? Thank you for always sticking by me no matter how far away I pulled, and for being a constant source of inspiration. You are an incredible person and an even better friend, and I will always be grateful to have you in my life. Aidan, I can't believe we have managed to deal with each other's presence for almost a decade. Thank you for being my brother in arms against all the brain demons and blood ghosts, and for keeping a very accurate countdown. But most importantly, thank you for being a friend from the start, and for showing me that moving to a different country does not have to be scary if you meet the right people. Sapir, my catastrophe companion, I think we have shared enough laughter and tears to last a lifetime. Thank you for always being there for me, and for reminding me that life does not start and end with academia. I would not be here without you and I can only hope to be able to repay you someday. Finally, Tobias, my franken-friend, can you believe we made it? There are not enough words in the English language (nor in Italian or my fluent understanding of Danish) to express how grateful I am to you, but I will nevertheless try. Thank you for all the coffee dates, movie nights, and cigarette breaks, and thank you for letting these be the pretext for all the emotional labour you do for me. Thank you for always knowing when to listen, offer advice, or just crack stupid jokes. Thank you for being the harshest Reviewer 2 you could be for this thesis. But above all, thank you for

simply being my best friend.

Finally, an enormous thank you to my family, Daniele, Stefania, and Grazia Gosetti, who believed in me and supported me since the start, even if I did not always give them much reason to. Grazie di tutto, ma forse soprattutto per la genetica testardaggine che mi ha permesso di completare questo percorso non sempre facile.

Declaration

All empirical work presented in this thesis has been carried out by the author. Co-authors and colleagues have provided inputs and suggestions for all empirical chapters, however the author alone has led the project and made the primary contribution to it. This work has not been previously submitted as application for any other higher education degree.

Valentina Gosetti

List of abbreviations

ASD	Autism Spectrum Disorder
AU	Action Unit
BA	Black African
CI	Confidence Interval
Co-I	Co-Information
DI	Dynamic Interactive model
EA	East Asian
fWHR	Facial Width-to-Height Ratio
GLM	General Linear Model
GMF	Generative Model of 3D human Faces
GUI	Graphical User Interface
IELTS	International English Language Testing System
LOOCV	Leave-One-Out Cross-Validation
MI	Mutual Information
ORE	Other Race Effect
PC	Principal Component
PCA	Principal Component Analysis
RSA	Representational Similarity Analysis
SCM	Stereotype Content Model
SD	Standard Deviation
SF	Spatial Frequency
SVM	Support Vector Machine
TIM	Trait Inference Mapping framework
WE	White European

Chapter 1

General introduction

Human faces are as socially salient as they are visually complex. This is because no one face looks exactly alike, owing to their large natural variations in both 3D shape—including overall size as well as the size and internal configuration of each individual feature (e.g., eyes, mouth, nose)—and 2D complexion—including overall skin tone, localized colouration (e.g., cheek flushing) and texture (e.g., wrinkles, blemishes). From this considerable variance, observers infer a commensurately large body of information about others, including their sex (e.g., Komori et al., 2011; Little et al., 2008), ethnicity (e.g., MacLin & Malpass, 2001; Rhodes et al., 2009), age (e.g., van Rijsbergen et al., 2014; Watson et al., 2016), and individual identity (e.g., Zhan, Garrod, et al., 2019). Beyond this relatively stable and objective information, observers also infer much more variable and subjective characteristics, including attractiveness (e.g., Zhan et al., 2021), social class (e.g., Bjornsdottir et al., 2024), and—importantly to the present—social traits.

1.1 Social trait perception

On a daily basis, humans make spontaneous judgments about the character of others, such as how trustworthy, dominant, competent, or likeable they are. Given the high social salience of human faces (e.g., see Jack & Schyns, 2017; Zebrowitz & Montepare, 2008), observers spontaneously (e.g., Klapper et al., 2016), implicitly (e.g., Swe et al., 2020), and readily (e.g., Willis & Todorov, 2006) infer these social traits from facial appearance. Despite the large natural variance of human faces, social traits are judged with high levels of inter-observer agreement by both adults and children (e.g., Cogsdill et al., 2014; Vernon et al., 2014). Further, although social trait perception typically refers to judgements made at zero-acquaintance and thus often leads to inaccurate inferences (e.g., Olivola & Todorov, 2010), social trait judgements are nonetheless update-resistant (e.g., Jaeger et al., 2020) and highly consequential (e.g., Menegatti et al., 2021). Taken together, social trait perception is a fundamental social process that, despite its unlikely reliability, develops early in life, influences social interactions throughout the life span, and appears guided by at least partly similar mechanisms across observers. In this section, I substantiate these claims by reviewing existing evidence for the prevalence, consequences, and accuracy of social trait perception.

1.1.1 Prevalence

Human faces are highly salient social stimuli that command attention (e.g., Jack & Schyns, 2017; Little et al., 2011; see also Farah et al., 1998). It is therefore perhaps unsurprising that observers infer social traits from faces after minimal exposure (e.g., Albert et al., 2021; Dzhelyova et al., 2012; Rule et al., 2012; Todorov et al., 2009; Willis & Todorov, 2006) and without being directly instructed to do so (e.g., Freeman et al., 2014; Klapper et al., 2016; Stewart et al., 2012; Swe et al., 2020, 2022; Verosky et al., 2020). For example, perceptions of dominance are reliably inferred after observing a face for 100ms (Albert et al., 2021; Willis & Todorov, 2006) or less (Rule et al., 2012). Similarly, task-irrelevant trustworthy faces impact neural responses in oddball tasks (Swe et al., 2022; Verosky et al., 2020). Further, observers infer social traits from minimal cues (e.g., Albert et al., 2021; Becker et al., 2007), and often rely on such inferences to guide their behaviour in spite of having access to more diagnostic information (e.g., Jaeger et al., 2019, 2020). Finally, the propensity to perceive social traits is strong enough to elicit similar social judgments of non-human characters, including Greebles (R. Lee et al., 2021) and robots (Reeves et al., 2020). Together, this evidence shows that social trait perception is not effortful, often automatic, and hard to override.

Reiterating the primacy of social trait judgements, developmental data shows that children as young as three years old infer social traits from faces (Cogsdill et al., 2014) and that their judgements converge with those made by adults (Cogsdill & Banaji, 2015), suggesting that social trait perception develops early in life in a robust manner. Among adults, social trait judgements also often produce high levels of inter-observer agreement (e.g., M. Mileva, Kramer, & Burton, 2019; Oosterhof & Todorov, 2008; Vernon et al., 2014), and this can hold true across different cultural/ethnic observer groups (e.g., Albright et al., 1997; Zebrowitz et al., 1993)—though more recent data suggests that using coarse agreement measures may hide cultural (e.g., Maeng et al., 2022; Rostovtseva et al., 2024) or observer-level idiosyncrasies (e.g., Hehman et al., 2017; Xie et al., 2019; see also Kramer et al., 2018, for further discussion on agreement measures). Indeed, high levels of inter-observer agreement could still occur when observers associate different conceptual knowledge with the same social trait label, or when the same social trait is inferred using different facial features (see Jack & Schyns, 2017; Schyns et al., 2022, for further discussion, see also section 1.5). Nevertheless, despite these more poorly understood causal mechanisms, current evidence shows that social trait perception is a fundamental process that colours our first impressions of others throughout the life span and across cultural boundaries.

1.1.2 Consequences

Though rapid and often subconscious, social trait judgements can have considerable consequences on individual lives and broader society. First, lab-based data shows that perceptions of facial trustworthiness influence partner selection (e.g., FeldmanHall et al., 2018) and cooperation in trust games (e.g., De Neys et al., 2017; Okubo et al., 2017)—though implicit

bias against other-ethnicity players can override this (Stanley et al., 2011). Perceptions of dominance similarly influence ally selection (Watkins & Jones, 2016) and strategy use in ultimatum games (Tang & Schmeichel, 2015). Second, social trait judgements can impact real-world outcomes, including dating preferences (e.g., South Palomares & Young, 2018; K. A. Valentine et al., 2014, 2020), professional success (e.g., Graham et al., 2017; Menegatti et al., 2021; Rule & Ambady, 2009), voting choices (e.g., Joo et al., 2015; Na et al., 2015; Olivola et al., 2012; Rule et al., 2010), and even sentencing decisions (e.g., Blair et al., 2004; Eberhardt et al., 2006; Wilson & Rule, 2015). Critically, many of these outcomes are further influenced by other types of face information, namely face ethnicity and sex. For example, sentencing decisions for Black defendants hinge not only on facial trustworthiness (Wilson & Rule, 2015) but also on the extent to which their faces appear more Afrocentric (Blair et al., 2004; Eberhardt et al., 2006). In a similar vein, perceptions of attractiveness as well as competence influence the hiring of female, but not male, applicants (Menegatti et al., 2021). Finally, variance in the observer can also nuance the consequences of social trait judgements. For example, past election outcomes are best predicted by judgements of competence—and particularly task-oriented vs social competence—in the USA relative to East Asian countries such as Taiwan (F. F. Chen et al., 2016) and South Korea (Na et al., 2015). Together, this body of work not only highlights the considerable downstream consequences of social trait judgements, but also underscores the central role of diversity in social trait perception (see section 1.4 for further discussion).

1.1.3 Accuracy

Given their prevalence and considerable downstream consequences, earlier assumptions held that face-based social trait judgements contain "kernels of truth"—that is, they are reliable indicators of a person's true intentions and dispositions. Supporting this, observers' judgements of extraversion (Penton-Voak et al., 2006) and conscientiousness (Little & Perrett, 2007) correlate with self-reported scores. Social trait judgements can also reflect more direct behavioural tendencies—for example, infidelity is associated with both facial untrustworthiness (Foo et al., 2022) and dominance (Arnocky et al., 2018). Nevertheless, these trait-behaviour associations typically have small effects and are context-specific (Foo et al., 2022). Further, while some context-specific associations may be veridical—for example, both facial dominance and sexual behaviour are testosterone-mediated (e.g., Lefevre et al., 2013)—others may be the result of self-fulfilling prophecies. For example, untrustworthy judgements correlate with lying behaviour to the extent that individuals already expect to be judged as untrustworthy (Slepian & Ames, 2016). Finally, evidence for trait-behaviour mappings is often confounded—for instance, while CEOs' facial competence has been associated with company profitability (Rule & Ambady, 2008), the effect disappears when past company profits are controlled for (Graham et al., 2017).

Despite these fallacies, observers still largely base their first impressions and subsequent treatment of others on facial appearance (e.g., Jaeger & Jones, 2022; Todorov et al., 2015;

Zebrowitz & Montepare, 2008). This remains true when more diagnostic information is provided (Jaeger et al., 2019, 2020) and in spite of the fact that face-based judgements reduce overall accuracy (Olivola & Todorov, 2010). Further, once established, social trait judgements are difficult to update (e.g., Ferguson et al., 2019), particularly when these initial judgements are negative (Zabag et al., 2023). This may be due to the fact that, regardless of actual accuracy, observers nevertheless *believe* to be accurate (Ames et al., 2010; Olivola et al., 2014). Known as "meta-accuracy", this belief captures observers' estimation of their own ability to accurately infer social traits yet bears no relation to actual performance—for example, the accuracy of leadership judgements of real-world leaders is unrelated to each observer's meta-accuracy (Olivola et al., 2014). Thus, evidence indicates that social trait judgements are seldom accurate and reliably map onto behaviour only in specific contexts. Nevertheless, observers' meta-accuracy, coupled with typically high levels of inter-observer consensus (see subsection 1.1.1), may inflate the perceived reliability of social trait perception, which in turn perpetuates its impact on real-world outcomes (see subsection 1.1.2).

1.2 Theories and models of social trait perception

Given the importance of social trait judgments for human social life and their significant downstream consequences for both individuals and broader society, a longstanding goal in the human behavioural sciences has been to understand how these judgments are made. Consequently, several theories and models have been put forth over the years to make sense of one or more of the following: *why* observers make social judgments, *how* observers make social judgments, and *what* types of faces drive social judgments in observers.

Historically, theories of social trait perception have focused on understanding *how* observers judge the social character of others. Grounded in the belief that social trait judgments are mostly accurate, the Lens Model introduced by Brunswik (1956) in environmental perception and later tailored to social perception (e.g., Gifford, 1994; Nestler & Back, 2013; Scherer, 1978) argues that observers use information that is readily available in their environment—here, the physical appearance and behaviour of others—to infer unperceivable information—here, the personality and intentions of others. To form their judgments, observers must incorporate available cues in a weighted manner that should, through experience, reflect the validity of these cues for a given judgment (Brunswik, 1956; Gifford, 1994; Nestler & Back, 2013). Notably, the focus of the Lens Model on accuracy opposes extensive evidence to the contrary (see subsection 1.1.3). In response to this, subsequent models such as the Weighted Average Model (Kenny, 1991) and the Stage Model of Dispositional Inferences (Trope & Higgins, 1993) argue that while social judgments arise from weighted combinations of available cues, these weightings are not dictated by the validity of each cue, but rather by each observer's current disposition, past experiences, and subjective interpretations. In doing so, both the Weighted Average Model and the Stage Model of Dispositional Inferences account for the recurrent inaccuracy of social trait judgments as well as interpersonal variability. Nevertheless, the correlational nature of all three models together with

their ambiguity in both the type and extent to which different cues are used to make social judgments has made them largely fall out of favour.

More recent theories and models have tried to correct this ambiguity by more precisely addressing the *why*, *how*, and/or *what* of social trait perception. Among these, the dominance-trustworthiness model (Oosterhof & Todorov, 2008, see subsection 1.2.1) is the only model that causally links social judgments with specific facial features, thereby directly addressing the questions of *what* types of faces drive social trait perception. For this reason, the empirical work presented in this thesis focuses on testing and extending this model. However, as discussed later, Oosterhof and Todorov (2008)'s model is not without limitation. Rather, other modern theories offer relevant insights and context into social trait perception that I will also discuss in relation to my work. Specifically, both ecological theory (Gibson, 1986; L. McArthur & Baron, 1983, see subsection 1.2.2) and the Stereotype Content Model (Cuddy et al., 2007; Fiske et al., 2002, see subsection 1.2.3) address *why* observers make social judgments, respectively taking an evolutionary and social hierarchical perspective. Additionally, Social vision (Adams & Kveraga, 2015; Adams et al., 2011, see subsection 1.2.4), the Dynamic Interactive model (Freeman & Ambady, 2011; Freeman et al., 2020; Stoller et al., 2020, see subsection 1.2.5), and the Trait Inference Mapping framework (Cook et al., 2022; Over & Cook, 2018, see subsection 1.2.6) extend and refine historical theorizing of *how* social judgments are made, with a relevant focus on how social categories such as ethnicity and culture nuance social trait perception.

1.2.1 Dominance-trustworthiness model

The dominance-trustworthiness model of social trait perception (Oosterhof & Todorov, 2008) is arguably the most prominent model in the field. Put simply, the model posits that all social trait judgments are structured along two orthogonal dimensions of trustworthiness and dominance, each characterized by distinct facial features and theorized to respectively reflect one's intentions toward others and their ability to enact them (Oosterhof & Todorov, 2008). Notably, unlike most other theories and models of social trait perception, the dominance-trustworthiness model was derived in a bottom-up manner. Specifically, Oosterhof and Todorov (2008) first asked participants to provide unconstrained judgments of 66 different real faces, which were then refined into the 14 most frequent judgements (e.g., attractive, aggressive, intelligent). A second group of participants was then asked to rate the same 66 faces according to the refined list of 14 judgments plus "dominant"—which was not found to be a frequent spontaneous judgment, but rather deemed theoretically relevant by the authors. Applying Principal Component Analysis (PCA) to this data, Oosterhof and Todorov (2008) revealed two Principal Components (PCs) which could account for over 80% of the variance in social trait judgments and respectively correlated with trustworthiness (63.3% explained variance) and dominance (18.3% explained variance). Finally, to reveal the specific facial features characterizing these two distinct judgments (i.e., PCs), a new sample of participants was asked to rate 300 computer-generated faces on trustworthiness and dominance alone. By

modelling the statistical relationship between the participants' ratings and the face stimuli's shape variations (broad complexion features such as luminance were included in later work, e.g., Todorov & Oosterhof, 2011), the authors (2008) identified which face shape features drive judgments of trustworthiness and dominance, and revealed their association with other relevant facial cues including emotion and physical strength respectively (see subsection 1.3.1 for more details).

Though undoubtedly influential, the dominance-trustworthiness model is not without limitations. Firstly, the two-dimensional solution proposed by Oosterhof and Todorov (2008) has only been replicated in Western observer samples (e.g., Morrison et al., 2017; H. Wang et al., 2016). Rather, when observers from different cultural and/or ethnic backgrounds are included, the fundamental dimensions which underpin social judgments range in both number (two to four; e.g., see B. C. Jones et al., 2021; Lin et al., 2021; Sutherland et al., 2013, 2018) and content (e.g., replacing dominance with capability, adding dimensions for attractiveness/youthfulness, masculinity/femininity; Lin et al., 2021; Sutherland et al., 2018; H. Wang et al., 2019). Secondly, the orthogonality of trustworthiness and dominance disappears once these dimensions are modelled from judgments of faces ranging in sex (e.g., M. Mil-eva, Kramer, & Burton, 2019; Sutherland et al., 2015) and age (e.g., S. Y. Ng et al., 2016). Finally, the causal facial features identified by Oosterhof and Todorov (2008) are by design only representative of White European young male faces. Though few studies have used diverse face stimulus sets to identify causal features, evidence suggests that social categories such as ethnicity influence how faces are socially evaluated (e.g., Hutchings et al., 2024; Sutherland et al., 2018; Xie et al., 2021). Thus, though widely embraced across the human behavioural sciences, the dominance-trustworthiness model fails to smoothly generalize to non-Western observers and non-White faces.

Nevertheless, extensive behavioural and neurological evidence supports the primacy of both trustworthiness and dominance in guiding social judgments of faces, at least within a White and Western context. First, perceptions of trustworthiness and dominance occur incredibly rapidly (e.g., Albert et al., 2021; Dzhelyova et al., 2012; Rule et al., 2012; Todorov et al., 2009; Willis & Todorov, 2006), suggesting that such perceptions are not effortful. Second, observers infer these traits readily from faces (e.g., Oosterhof & Todorov, 2008), and in the case of trustworthiness this applies cross-culturally (e.g., B. C. Jones et al., 2021; Sutherland et al., 2018). Third, these perceptions occur unconsciously (e.g., Freeman et al., 2014; Stewart et al., 2012) and implicitly (e.g., Klapper et al., 2016; Swe et al., 2020, 2022; Verosky et al., 2020)—that is, without directing one's attention to these perceptions neither intentionally nor through task demands. Finally, updating initial judgments of trustworthiness and dominance is possible but difficult, requiring intensive training (Chua & Freeman, 2021; Hong et al., 2024), or extensive diagnostic counter-evidence (Ferguson et al., 2019) that is still not always successful (Jaeger et al., 2020). Together, such evidence points to the fundamental role that trustworthiness and dominance play in structuring (at least White Western) observers' judgments of others.

1.2.2 Ecological theory and the overgeneralization hypothesis

Ecological theory, as applied to both object (Gibson, 1986) and social perception (L. McArthur & Baron, 1983), states that perception and action are fundamentally linked through co-evolution, and this link thus necessitates perception to be adaptive. In other words, by directly informing action, perception must support the survival of the species and/or goal attainment for the individual. To do this, perception must focus on detecting *useful* information in the environment—that is, information which is relevant to the observer’s goals and intentions and thereby affords opportunities for adaptive action (L. McArthur & Baron, 1983). Notably, though the ecological stance on perception inherently accounts for individual differences according to one’s goals, intentions, and experiential learning, it also predicts that perception should be at least in part guided by attunements to universally adaptive information (Gibson, 1986; L. McArthur & Baron, 1983). In the context of social face perception, the ecological approach therefore posits that certain facial characteristics that signal highly adaptive information such as age, attractiveness, and emotions should drive perception (and action) in similar ways across observers. To illustrate, angry faces are typically judged as threatening and thus elicit avoidance (e.g., Kaltwasser et al., 2017; Marsh, Ambady, & Kleck, 2005; Stins et al., 2011), while infant faces are typically evaluated positively and thus elicit nurturing (e.g., Glocker et al., 2009; Senese et al., 2013).

Critically, though adaptive in nature, such universal attunements can lead to biased perceptions. The overgeneralization hypothesis (see e.g., Zebrowitz, 2004, 2017; Zebrowitz & Montepare, 2008, for reviews) states that judgments which are accurately elicited by faces that reliably signal adaptive cues—for example, aggression from anger-expressing faces—are also elicited by faces which merely resemble these cues—for example, neutral faces with a naturally lower brow bone resembling a scowl. In other words, by being highly attuned to adaptive facial cues, observers erroneously attribute cue-congruent social traits to faces that are structurally similar to—but not necessarily indicative of—those cues. Supporting this view, neutral faces that structurally resemble angry (vs happy) facial expressions are judged to be less (vs more) likeable and trustworthy (e.g., Adams et al., 2012; Oosterhof & Todorov, 2008; Said et al., 2009; Zebrowitz et al., 2010); baby-like (vs mature-looking) faces are judged to be more (vs less) submissive, trustworthy, and incompetent (e.g., Jaeger et al., 2020; Zebrowitz & McDonald, 1991; Zebrowitz et al., 2003); and attractive (vs unattractive) faces are judged to be more (vs less) intelligent, competent, and powerful (e.g., Batres & Shiramizu, 2023; Little et al., 2012; Rhodes, 2006; Zebrowitz, 2004).

Structural resemblance to sexually dimorphic cues presents an interesting case of overgeneralization. As with other adaptive cues, faces which appear more feminine (vs masculine) are judged to be more (vs less) trustworthy, submissive, and warm (e.g., Sutherland et al., 2015; Zebrowitz et al., 2010). Critically, the facial features that make a face appear feminine (vs masculine) overlap with those that make a face appear baby-like (vs mature-looking) and resemble happy (vs angry) facial expressions—for example, thinner eyebrows, bigger eyes, rounder cheeks, and fuller lips are associated femininity as well as younger age and/or positive emotion cues (e.g., Becker et al., 2007; Hess et al., 2009a, 2009b; Zebrowitz

et al., 2003). It therefore is unclear whether judgments of feminine vs masculine faces are guided only by their resemblance to sexually dimorphic cues or by their overlap with age and emotion cues. Further, the social judgments elicited by feminine vs masculine faces reflect gender stereotypes: for example, feminine faces which elicit counter-stereotypical judgments (e.g., dominant) are evaluated more negatively than dominant-looking masculine faces (e.g., Oh, Buck, & Todorov, 2019; Oh, Dotsch, et al., 2020; Oliveira et al., 2020; Sutherland et al., 2015). Thus, while the overgeneralization hypothesis states that bottom-up information—here, facial features that resemble adaptive cues—guides social trait perception, social judgments of feminine vs masculine faces suggest that social trait perception could also be in part driven by top-down information—here, gender stereotype knowledge.

1.2.3 Stereotype Content Model

Opposing the ecological view on social trait perception, the Stereotype Content Model (SCM; Cuddy et al., 2008, 2009; Fiske et al., 2002) proposes that social judgments are guided in a top-down manner by group stereotype knowledge, which is in turn synthesized according to two universal dimensions of warmth and competence. As dimensions, warmth and competence respectively reflect one's intentions and ability to harm others, and their combination captures one's standing in social hierarchies and their consequent treatment in society (Cuddy et al., 2008, 2009; Fiske et al., 2002). For example, in the USA elderly people are typically judged as warm but incompetent and are therefore treated with pity, while Asian immigrants are judged as cold but competent and therefore treated with envy (Cuddy et al., 2007; Fiske et al., 2007). These examples highlight a key tenet of SCM: stereotypes are ambivalent constructs, meaning that stereotyped groups will often be rated high on one dimension and low on the other in line with the group's perceived status (associated with competence) and competitiveness (associated with warmth; Cuddy et al., 2009). Further, although SCM stems from the field of person—rather than face—perception, subsequent work has shown that people associate specific facial features with varying levels of warmth and competence (e.g., Hensel et al., 2020; Imhoff et al., 2013; Walker & Wänke, 2017; F. Wen et al., 2020). For example, both male and female faces with more feminine features are perceived as warm (F. Wen et al., 2020), and male faces with textured (e.g., acne) complexion are perceived as incompetent (Fetscherin et al., 2020).

Cultural differences in values (e.g., Hanel et al., 2018; Schwartz, 2014), ideology (e.g., Guimond et al., 2013), and collectivism vs individualism (e.g., Cuddy et al., 2014) can all impact which stereotypes are formed and about whom. Despite this, a large body of evidence supports the notion that warmth and competence form a universal structure for stereotype knowledge (though its content can vary considerably both across and within cultures; see e.g., Durante et al., 2017; Fiske, 2017; Stanciu et al., 2017). In an initial series of studies, Cuddy et al. (2009) showed that judgments of different social groups provided by USA nationals as well as Europeans and East Asians could be clustered according to warmth and competence. Subsequent work conducted in Romania (Stanciu, 2015), Russia (Grigoryev

et al., 2019), and mainland China (Guan et al., 2010) further supports the cross-cultural validity of the two dimensions. Judgements of warmth and competence from faces also show some cultural similarities. For example, feminine facial features drive perceptions of warmth in both Western (Walker & Wänke, 2017) and East Asian observers (Luo et al., 2024; F. Wen et al., 2020); both American and Korean observers rely on facial competence when deciding who to vote for in hypothetical election scenarios (Na et al., 2015); and face-based judgements of real political candidates in the USA and Japan show high inter-rater consensus (Rule et al., 2010). Together, this suggests that the conceptual universality of warmth and competence could translate into at least some cultural similarities in the way these traits are inferred from faces.

Notably, the dimensions of warmth and competence are conceptually similar to the dimensions of trustworthiness and dominance proposed by Oosterhof and Todorov (2008, see subsection 1.2.1). Specifically, both warmth and trustworthiness relate to one's good vs bad intentions towards others—though some argue a distinction between moral intentions (i.e., trustworthiness) and social intentions (i.e., warmth; e.g., see Brambilla et al., 2011)—while both competence and dominance cover one's ability to enact their intentions (Fiske et al., 2002; Oosterhof & Todorov, 2008). However, comparisons of face judgments along these similar dimensions show that while judgments of warmth and trustworthiness correlate highly, those of competence and dominance do not (Oliveira et al., 2019; Sutherland et al., 2016). This in turn suggests that while conceptually similar, judgments of warmth/competence and trustworthiness/dominance are at least in part underpinned by different facial features. Nevertheless, the feature similarities between warmth and trustworthiness (Oliveira et al., 2019), together with their spontaneous inference cross-culturally (e.g., B. C. Jones et al., 2021; Lin et al., 2021; Sutherland et al., 2018) and the considerably higher explained variance of these judgments relative to both competence/dominance as well as other dimensions (e.g., B. C. Jones et al., 2021; Lin et al., 2021; Sutherland et al., 2013, 2018; H. Wang et al., 2019) suggests that inferring the valence of others' intentions toward us—whether this is labelled as warmth or trustworthiness—is a fundamental aspect of social perception.

1.2.4 Social vision and the shared signal hypothesis

In its simplest form, Social vision (Adams & Kveraga, 2015; Adams et al., 2011) is a research framework that proposes an alternative to understanding social face perception, where classically separate areas of study—such as face recognition, impression formation, and emotion perception—should instead be examined as a combinatorial process. Inspired by the neo-Darwinian view on emotion perception (e.g., Ekman & Friesen, 1986) and the ecological theory of social perception (e.g., L. McArthur & Baron, 1983; Zebrowitz & Montepare, 2008, see also subsection 1.2.2), Social vision argues that that our visual system has co-evolved with the social complexity of our species (i.e., the social brain hypothesis; Dunbar, 1998; see also Albohn & Adams, 2022), leading to the propensity to co-process

visual social cues to integrate their social meaning (e.g., dominance vs affiliation, approach vs avoid; Adams et al., 2011). Extending from this, the shared signal hypothesis states that multiple sources of social information, including static (e.g., shape features) and dynamic facial features (e.g., facial expressions, eye gaze) as well as other bodily cues, combine and interact in perception to bias and/or nuance social judgments depending on their social meaning (Adams & Kveraga, 2015; Adams et al., 2011, 2017; Albohn et al., 2022). Supporting this, facial cues to age, gender, and ethnicity affect emotion perception in line with the congruency of the social meaning of the emotion-categorical cue pair (e.g., Becker et al., 2007; Hess et al., 2009b; Hugenberg, 2005; Marsh, Adams, & Kleck, 2005).

In the context of social trait perception, the shared signal hypothesis places particular importance on integrating meaning according to dominance and affiliation (Adams & Kveraga, 2015). These two motivational stances easily map onto previously proposed 2D social trait structures—including the dominance-trustworthiness model (Oosterhof & Todorov, 2008, see also subsection 1.2.1) and SCM's competence and warmth dimensions (Fiske et al., 2002, see also subsection 1.2.3)—and are indeed perceived from multiple cues, including static facial appearance (e.g., Zebrowitz & Montepare, 2008), dynamic facial expressions (e.g., Knutson, 1996), bodily movements (e.g., Vacharkulksemsuk et al., 2016), and the voice (e.g., Wood et al., 2017). In this view, social trait perception is therefore the result of a combinatorial process that integrates otherwise distinct social cues to forecast others' fundamental tendencies to dominate or affiliate (Adams & Kveraga, 2015). For example, masculine facial features as well as features that make a face appear mature-looking or anger-resembling guide perceptions of dominance in line with their underlying social meaning (e.g., Adams et al., 2015). Their shared signal is further evidenced by their ability to disambiguate one another—for example, observers categorize androgynous faces as "male" both when they actually display facial expressions of anger (Hess et al., 2009b) as well as when they simply resemble anger expressions (e.g., lower brow bone; Becker et al., 2007).

Critically, the integration of multiple cues in social perception has two possible explanations. First, as explained above, multiple cues may be integrated according to the (in)congruency of their fundamental social meaning. Second, in line with evolutionary perspectives (Darwin, 1872), cues to distinct social categories such as gender, age, and emotion can visually overlap, particularly within the face (Hess et al., 2009b). To illustrate, masculine, mature-looking, and anger-resembling facial features share not only the same social meaning (i.e., dominance) but also similar structural characteristics (e.g., thin lips, lower brow-bone; Adams et al., 2015). Notably, many proponents of Social vision support both explanations through the notion of functional equivalence—that is, the cues that observers use for social perception have co-evolved to resemble one another to the extent that they share similar meaning (Adams et al., 2011, 2017; Hess et al., 2009b). For example, cues to facial maturity drive perceptions of dominance (e.g., Jaeger et al., 2020) and yield systematic differences in emotion perception (Marsh, Adams, & Kleck, 2005) in line with structural similarities between specific emotions and age cues (Zebrowitz et al., 2007). Nevertheless, structural resemblance is not a prerequisite for integration in social perception. For example,

Black African faces are perceived as more threatening (Kleider-Offutt et al., 2018) and angry (Hugenberg, 2005) by White observers in spite of the fact that anger (vs happy) facial expressions structurally overlap most with White (vs Black African) faces (Adams et al., 2022; Zebrowitz et al., 2010). Therefore, neither explanation of why multiple cues are integrated in perception precludes the other, and both are supported by a considerable body of evidence (see subsection 1.3.1 for more details). However, a lack of precise feature-based accounts of social trait perception limits further understanding of these mechanisms (see section 1.5 for further discussion).

1.2.5 Dynamic Interactive model

Bridging debates on the bottom-up vs top-down nature of social perception, the Dynamic Interactive model (DI; Freeman & Ambady, 2011; Freeman et al., 2020) argues that social judgments arise from the interaction between bottom-up visual inputs and top-down cognitive processes, including the observer's prior experiences, stereotype knowledge, and current affective and motivational states. In other words, social perception is not a unilateral process (whether bottom-up or top-down), but rather a rapid negotiation between the information afforded by the visible features of a face and each observer's prior knowledge and current state (Freeman & Ambady, 2011; Freeman et al., 2020). Importantly, though rapid, such negotiations are both gradual—meaning that multiple partially activated concepts (e.g., gender, ethnicity, emotion, social traits) compete in perception before arriving at a discrete social judgment—and dynamic—meaning that social judgments fluctuate over time as different concepts are further activated or dampened (Freeman & Ambady, 2011; Freeman et al., 2020). Because of their gradual and dynamic nature, social judgments—including simple categorizations (e.g., male vs female) and higher-level perceptions (e.g., trustworthy)—are inherently nuanced and/or biased by potential overlaps in cue-diagnostic facial features and their associated stereotype knowledge (Freeman & Ambady, 2011; Freeman & Johnson, 2016; Freeman et al., 2020). For example, disambiguation of ethnically diverse androgynous faces is influenced by both feature overlaps between ethnicity and sex cues (e.g., the typically larger features of both Black African and male faces) and shared stereotypes between ethnicity and sex categories (e.g., aggression is stereotypically associated with both "Black African" and "male"; K. L. Johnson et al., 2012).

Subsequent elaborations of DI (Freeman et al., 2018; Stoler et al., 2020) formalize the bidirectional influence of bottom-up and top-down information within the context of social trait perception by proposing the existence of a shared representational structure between social trait concepts and social trait features. That is, the more a given observer believes two social traits to be conceptually related, the more they will use physically similar facial features to infer these traits (Stoler et al., 2018, 2020). Supporting this, Stoler et al. (2018) used a reverse correlation paradigm to show that the visual similarity in the facial features driving judgments of the Big Five personality traits was positively correlated with the conceptual similarity of these traits for observers. Further, because certain social traits

are stereotypically associated with specific social categories (e.g., female-submissive, Black African-aggressive; see e.g., Devine & Elliot, 1995; Eagly & Steffen, 1984), facial features indicative of these categories can also drive social trait perception to the extent that the observer holds such stereotypical associations (Stolier et al., 2020). Indeed, social trait judgments of ethnically diverse male and female faces covary with the extent to which observers associate sex and ethnicity with specific social traits (Xie et al., 2021). Finally, the influence of categorical facial features on social trait perception varies according to how prototypical such features are for the category. For example, Black African faces with more Afrocentric features are perceived as having more traits stereotypically associated with the category "Black African" (e.g., aggression; Blair et al., 2002; Kleider-Offutt et al., 2018).

Empirical support for DI has been primarily obtained using two approaches: mouse-tracking (see e.g., Freeman et al., 2008) and Representational Similarity Analysis (RSA; see e.g., Freeman et al., 2018). In a mouse-tracking task, observers are asked to categorize a face stimulus according to a binary option (e.g., male vs female) by clicking on one of the two category labels presented on opposite sides of a computer screen; the trajectory of the mouse movements—specifically, how straight vs skewed the movement is—then serves as a proxy for how much the non-selected category influences perception of the selected category (Freeman et al., 2008). To illustrate, mouse-tracking studies have shown that sex categorizations (Freeman et al., 2008) as well as categorization of gender-stereotypical traits (Freeman & Ambady, 2009) and sexual orientation (Freeman et al., 2010) are influenced by how sex-typical a face is. On the other hand, RSA is an analytical—rather than experimental—approach that, put simply, first computes all pairwise (dis)similarities across judgments in a given domain (e.g., perceptual decisions, conceptual knowledge) and then compares the vectorized results *across* domains to provide a quantitative representation of how similar two representational structures are (see e.g., Freeman et al., 2018, for more details). For example, the work described above by Kleider-Offutt et al. (2018), Stolier et al. (2018), and Xie et al. (2021) all employ RSA. Critically, though both approaches provide compelling evidence for the fact that multiple competing concepts can be activated by the same face prior to arriving at a final social judgment, they are not suited to explain why this competition occurs. That is, because neither approach is designed to unveil which facial features elicit activation of competing concepts, it remains unclear whether this is caused (at least in part) by an overlap in concept-diagnostic facial features or solely by semantic associations between competing concepts (see section 1.5 for further discussion).

1.2.6 Trait Inference Mapping framework

Akin to DI, the Trait Inference Mapping framework (TIM; Cook et al., 2022; Over & Cook, 2018; Over et al., 2020) also proposes that social judgments can be explained by the interaction between two correlated representational structures: face space and trait space. Aligned with earlier propositions in face recognition research (e.g., Johnston et al., 1997; T. Valentine, 1991), TIM defines face space as a multidimensional structure that codes vari-

ations in facial appearance (Over & Cook, 2018). Notably, though some basic aspects of face representation can be observed from birth (see e.g., M. H. Johnson et al., 2015, for a review), the structure of face space is shaped by each observer's prior experiences (Over & Cook, 2018). That is, the facial features variations that are coded across the dimensions of face space vary according to the types of faces each observer encounters (Over & Cook, 2018), with more common (e.g., own-ethnicity) faces being better represented than uncommon (e.g., other-ethnicity) faces (e.g., Caldara & Abdi, 2006; T. Valentine, 1991). Parallel to—yet dissociable from—face space, trait space defines each observer's conceptual knowledge of others' traits, including their likely co-occurrence and associated behaviours (Over & Cook, 2018; Over et al., 2020). Like face space, trait space develops through experience and can thus vary across observers (Over & Cook, 2018; Over et al., 2020). However, beyond *personal* experiences, TIM emphasizes the role of *cultural* experience in the development of trait concepts—that is, repeated exposure to cultural representations of different traits in popular media, traditional folk-tales, and even political propaganda powerfully shapes trait space (Cook et al., 2022).

Within the TIM framework, social judgments reflect learned mappings between face space and trait space. These mappings can form in a controlled manner when, upon encountering a new person, the observer represents their facial appearance at a certain point in face space and simultaneously represents (at least parts of) their character at a point in trait space (Over et al., 2020). However, over time and experience, social trait judgments can also become automatic through established face-trait mappings—that is, repeatedly encountering individuals with specific facial features and specific traits establishes predictive links between the two spaces, such that representing a novel face at a point in face space is sufficient to trigger the perception of the trait(s) represented in the linked point in trait space (Over & Cook, 2018; Over et al., 2020). Importantly, these established face-trait mappings do not only arise through direct personal experience—indeed, if this was the case it would be difficult to explain the high levels of inter-observer consensus (see subsection 1.1.1)—but also learned cultural associations (Cook et al., 2022). For example, cartoon depictions of heroes and villains tend to be designed according to the "beautiful is good" archetype (e.g., England et al., 2011; Sharmin & Sattar, 2018), and therefore establish a cultural link between specific types of character appearance (e.g., beautiful vs ugly) and specific traits (e.g., brave vs cunning). Thus, TIM argues that social judgments are learned—rather than innate—processes which are fundamentally shaped—rather than secondarily nuanced—by culture.

A growing body of evidence supports the learned nature of social judgments. For example, brief lab-based training sessions are sufficient to establish automatic social trait associations with novel non-human characters (i.e., Greebles; R. Lee et al., 2021) as well as human faces (Chua & Freeman, 2022). Further, previous experiences with trustworthy vs untrustworthy partners in a trust game influences new partner selection based on facial similarity—that is, the appearance of those we have learned to trust is generalized as a cue to trustworthiness in novel faces (FeldmanHall et al., 2018). In a similar vein, the typicality of a face for a given observer (Sofer et al., 2015) or broader cultural group (Sofer et al.,

2017) can also affect trustworthiness judgments, suggesting that both personal and cultural learning can affect social trait judgments. Supporting the causal influence of culture, faces known to elicit specific social trait judgments (e.g., trustworthy) in Western culture do not do so to the same extent in East Asian culture (Rostovtseva et al., 2024). On the other hand, evidence from a twin study shows that variations in trustworthy perceptions are better explained by differences in personal experiences rather than by shared environment, though both factors still explain more variance than genetics (Sutherland et al., 2020). Together, this body of work highlights the likely role of learning in how observers judge others' character, though results conflict on the relative importance of individual vs cultural learning. Critically, further understanding of this is impeded by the field's overreliance on Western culture (see section 1.4 for further discussion).

1.3 Drivers of social trait perception

Despite theoretical debates on their origins and underlying mechanisms, all theories and models of social trait perception align on the notion that social judgements are driven by specific cues. Within the face, these cues can be generally categorized as invariant—that is, stable structural characteristics that may denote one's individual identity as well as broader social categories such as ethnicity, sex, and age—or changeable—that is, momentary and often dynamic characteristics such as facial expressions, eye gaze, and head tilt (e.g., Adams et al., 2011; Sutherland & Young, 2022; Sutherland et al., 2017; see also Haxby et al., 2000). Beyond these, aesthetical modifications such as makeup, tattoos, and hairstyles can also impact which social judgements are inferred from faces (e.g., Etcoff et al., 2011; Funk & Todorov, 2013). Finally, though human faces are highly salient and rich sources of information for social trait perception, non-facial cues including the voice (e.g., McAleer et al., 2014) and body (e.g., Hu et al., 2018) have also been shown to influence social judgements. Given the focus of this thesis, in this section I first review in detail the invariant facial features of 3D shape and 2D complexion that have been proposed as drivers for social trait judgements. I then briefly review changeable social trait facial features, before discussing other facial and non-facial features.

1.3.1 3D face shape

Variations in 3D face shape are arguably the most extensively studied type of social trait cue. Owing to the complexity of this feature space—which includes not only variations in overall face shape, but also the shape, size, and configuration of its internal features (e.g., eyes, nose, mouth; e.g., see Farkas et al., 2005; Jack & Schyns, 2017; Jaeger & Jones, 2022)—a large portion of existing work has focused on specific subsets of 3D shape features. For example, early work on the overgeneralization hypothesis (see subsection 1.2.2) has tested the role of baby-like facial appearance—which comprises a rounder face shape with a smaller chin, a bigger head, larger eyes, and thin, higher eyebrows—on social trait

perception (e.g., Berry & McArthur, 1985; Zebrowitz & McDonald, 1991; Zebrowitz & Montepare, 1992). These features consistently elicit higher ratings of warmth and kindness (Berry & McArthur, 1985) as well as lower ratings of dominance and intelligence (L. Z. McArthur & Apatow, 1984), suggesting that at least some aspects of baby-like appearance influence social trait perception. Similarly, anomalous facial appearance—which broadly refers to asymmetric features that diverge from the average as well as visible wounds or scarring—decreases ratings of competence, trustworthiness and intelligence (e.g., Zebrowitz et al., 2003) as well as pro-social treatment (e.g., Hartung et al., 2019; Workman et al., 2021). Parallel and opposite effects have also been reported for mature-looking (Berry & McArthur, 1986; Friedman & Zebrowitz, 1992) and attractive faces (e.g., Batres & Shiramizu, 2023; Little et al., 2012). Critically, though both baby-like and anomalous facial appearance influence social judgements, the breadth of their characterizing features makes it difficult to understand whether specific features within these subsets drive social trait perception more than others.

A more constrained feature subset that has received much attention is the Facial Width-to-Height Ratio (fWHR), which codes the bizygomatic width of the face divided by its length from the lips to the eyebrows (Weston et al., 2007). fWHR is often referred to as a sexually dimorphic cue due to its typically higher value in male vs female faces (e.g., Carré & McCormick, 2008; Geniole et al., 2015) and its association with testosterone levels (e.g., Lefevre et al., 2013) and testosterone-mediated behaviour (i.e., aggressive and/or antisocial behaviour; e.g., Carré et al., 2009; Haselhuhn & Wong, 2011; Haselhuhn et al., 2015). Although more recent evidence questions its reliability as a sexually dimorphic cue (e.g., see Hodges-Simeon et al., 2016; Kramer, 2017), fWHR still appears to influence social trait perception differently for male and female faces. Specifically, higher fWHR in male faces elicits higher ratings of aggression, dominance, and untrustworthiness (e.g., Alrajih & Ward, 2014; McCormick et al., 2010; Stirrat & Perrett, 2010), while in female faces it only elicits higher ratings of aggression and untrustworthiness but not dominance (Durkee & Ayers, 2021; V. R. Mileva et al., 2014). This divergence might be because perceptions of dominance are driven by masculine features beyond higher fWHR, including a sharper jaw and thinner lips (Burris et al., 2007). Indeed, dominance perceptions increase when female faces have more masculine features beyond higher fWHR (Quist et al., 2011), and masculinizing male faces increases dominance ratings even at 100ms exposure (Albert et al., 2021). Nevertheless, recent evidence suggests that social trait perception may be most influenced by another subset of features—namely, features that resemble facial expressions of emotion (Jaeger & Jones, 2022).

Structural resemblance to facial expressions of emotion refers to those static facial features that make an otherwise neutral face appear to be expressing a given emotion. For example, a naturally lower brow bone may resemble an angry scowl, slightly upturned mouth corners may resemble a happy smile, and thin, high-set eyebrows may resemble the shock of fear (e.g., Adams et al., 2012; Albohn & Adams, 2021b; Zebrowitz, 2017). An extensive body of evidence shows that these features impact social trait perception in valence-congruent

ways. Specifically, anger-resembling features—which comprise a lower brow bone, angled eyebrows, thinner lips, and a sharper, juttied jaw—drive perceptions of both dominance and untrustworthiness (e.g., Albohn & Adams, 2021b; Montepare & Dobish, 2003; Windmann et al., 2023); happy-resembling features—which comprise arched eyebrows, fuller cheeks, and slightly upturned mouth corners—drive perception of both trustworthiness and warmth (e.g., Albohn & Adams, 2021b; Jaeger & Jones, 2022; Said et al., 2009); and fear-resembling features—which comprise bigger, rounder eyes and high-set eyebrows—drive perceptions of submissiveness (e.g., Albohn & Adams, 2021a; Keating et al., 1977; Montepare & Dobish, 2003). The link between emotion cues and social trait perception is further evidenced by the fact that observers readily infer the same social traits from actual facial expressions of emotion (e.g., Hareli et al., 2009; Hess et al., 2009a; Knutson, 1996; Sutherland et al., 2017). However, though evidently powerful, emotion-resembling features carry the same limitation as the other feature subsets: the breadth of the features they comprise. Further complicating the issue, these subsets tend to overlap at the feature and/or concept level. For example, mature-looking, masculine, and anger-resembling features overlap both structurally (e.g., lower brow-bone, thinner lips) and conceptually (e.g., associated with power and/or aggression; e.g., Adams et al., 2015; Becker et al., 2007; Hess et al., 2009b). Thus, though examination of these feature subsets has provided numerous insights into which facial features influence social trait perception, their inherent coarseness and natural overlaps limit finer-tuned understanding.

Data-driven methods have gained popularity in the field because they can overcome the limitations imposed by a priori assumptions, such as only testing the impact of coarse feature subsets on social trait perception (e.g., see Jack & Schyns, 2017; Todorov et al., 2011, for reviews, see also section 1.5 for further discussion). As a key example, Oosterhof and Todorov (2008, see also subsection 1.2.1) used reverse correlation to reveal which types of faces drive perceptions of dominance and trustworthiness. Visual inspection of these seminal results shows that perceptions of dominance are driven by a bigger face with a more protruding brow bone, smaller eyes, thinner lips, and a stronger jaw and chin; while perceptions of trustworthiness are driven by a smaller face with arched eyebrows, bigger eyes, upturned mouth corners, and fuller lips. Importantly, though these results map onto previously investigated feature subsets and were indeed rated as indicative of specific emotion, sex, and age cues by a secondary set of observers (Oosterhof & Todorov, 2008), they were derived in a bottom-up manner that was therefore unconstrained by prior assumptions or specific hypotheses. Subsequent data-driven investigations have extended these results to show how the sex of the face (e.g., M. Mileva, Kramer, & Burton, 2019; Sutherland et al., 2015) and the culture of the observer (e.g., Sutherland et al., 2018; H. Wang et al., 2019) can influence these features. Similarly, reverse correlation has been employed to identify which facial features drive perceptions of competence (Oliveira et al., 2019), criminality (Dotsch et al., 2011), and different personality traits (Stolier et al., 2018). However, despite the valuable insights that reverse correlation can provide, the ways in which these results are interpreted can significantly impact their specificity, and consequently the extent to which they can provide causal

explanations of social trait perception.

Classic reverse correlation approaches model variations in the entire face. To identify more specific facial features, it is common to rely on visual inspection of the resulting classification image, which is based on the sample average (e.g., Oosterhof & Todorov, 2008; Sutherland et al., 2013, 2018). Critically, doing so cannot provide an objective demonstration of which specific facial features drive social perception. Further, because classification images are based on the group mean, they erase any potentially relevant individual differences (e.g., see Jack & Schyns, 2017; Medin et al., 1993, for further discussion). Additionally, averaging across participants in this way can inflate Type I errors (Cone et al., 2021). Thus, interpreting reverse correlation results by visually inspecting classification images limits objective understanding, risks erasure of important observer-level idiosyncrasies, and increases the rate of false positives (see subsection 1.5 for further discussion). An alternative to visual inspection is formal comparison of pixel luminance values (e.g., Dotsch & Todorov, 2012; Dotsch et al., 2011; Oliveira et al., 2019), where clustering approaches can reveal which face regions predict perception. Though this approach is less subjective than visual inspection, it still fails to explain *how* a given face region drives social judgements. To illustrate, if clustering highlights the eye region, does this entail that eye size, eye shape, distance between the eyes, their combination, or yet another characteristic of the eyes is relevant? Further, clustering pixel luminance values does not always yield easily interpretable results. For example, in Dotsch and Todorov (2012), most pixel clusters highlighted small, disjointed regions such as the mid-portion of the right eyebrow for perceptions of trustworthiness, and a small section of the left under-eye area for perceptions of dominance. Thus, though more formal than visual inspection, clustering approaches also fail to robustly and objectively identify which specific facial features drive social trait perception. Together, though a large body of work has investigated which 3D shape features drive social judgements, current knowledge lacks precise accounts.

1.3.2 2D face complexion

Relative to 3D face shape, variations in 2D complexion—which comprise both skin colour and texture (e.g., blemishes, wrinkles, scarring)—are less frequently investigated as cues for social trait perception. This may be due to the fact that, unlike 3D shape, experimentally manipulating 2D complexion poses considerable technical challenges. First, stimulus sets that capture real-life variations in 2D complexion can vary significantly depending on the type of camera used, the external lighting conditions, the image processing techniques used (if any), and the final resolution of the image file (Thorstenson, 2018). Second, though skin colour varies across face regions (e.g., Fink & Matts, 2008; A. L. Jones et al., 2016), experimental manipulations are typically applied to the entire face due to technological constraints (e.g., Thorstenson & Pazda, 2021; Thorstenson et al., 2019). Third, identifying changes in skin colour yields relatively poor observer performance (e.g., Cooper & Gerlach, 2013) which can further be affected by even small difference in colour display settings across

monitors (Thorstenson, 2018). Finally, when compared to 3D face shape, 2D complexion influences social trait perception much less (e.g., M. Mileva, Young, et al., 2019; Oh, Dotsch, & Todorov, 2019). Together, these limitations have led a large proportion of extant social trait perception research to disregard the role of 2D complexion. However, this is not the case for related social judgements, including age (e.g., Puccetti et al., 2011), health (e.g., A. L. Jones et al., 2016), gender (e.g., M. Wang et al., 2022), attractiveness (e.g., Zhan et al., 2021), and—more recently—emotions (e.g., Thorstenson et al., 2019). Given the influence that these judgements have on social trait perception from 3D shape features, it is likely that they play a similar role from 2D complexion.

Given the impact that the natural ageing process has on human skin, a large body of work has examined the role of 2D complexion features on perceptions of age. Age judgements rely on both more obvious characteristics such as wrinkles and folds (e.g., Fink & Matts, 2008) as well as skin colour variations, where brighter and more homogenous colouration increases perceptions of youthfulness (e.g., Fink et al., 2006; Matts et al., 2007; Puccetti et al., 2011). Higher facial contrast—that is, differences in luminance and colour between specific facial features such as the eyes and lips and the surrounding skin—similarly increases perceptions of youthfulness (Porcheron et al., 2013). Notably, though many of these features result from the natural process of ageing, they can influence age perception differently depending on other characteristics of the face—for example, patchier skin tones drive perceptions of old age most in Chinese faces, while wrinkles and folds do so most for White faces (Porcheron et al., 2014). Similarly, the true age of the face can influence effects—for example, smiling-related wrinkling (e.g., crows feet) increases perception of age in young adult faces but not older adult faces (Ganel & Goodale, 2021). Despite these differences, recent evidence shows that age-related complexion features impact social trait perception in similar ways for young and older adult faces regardless of other age-related shape features (e.g., lower jowls), with more wrinkled texture decreasing ratings of trustworthiness and pleasantness, as well as attractiveness (Hess et al., 2023). Thus, though seldom investigated in the context of social trait perception, age-related 2D complexion features appear to influence social trait judgements in similar ways to mature-looking shape features.

The notion that age-related complexion features decrease attractiveness ratings can be explained by the fact that many of the 2D complexion features that drive age perception also influence perceptions of health, gender, and attractiveness. Specifically, homogenous skin tone in female faces increases perceptions of both overall health (Fink et al., 2006) and fertility (Samson et al., 2011). Still in female faces, localized colouration, including brighter eye regions and redder cheeks, increase perceptions of health (A. L. Jones et al., 2016). These area-specific effects, together with higher facial contrast and overall brighter skin tones, also guide "female" categorizations (Dupuis-Roy et al., 2009; Nestor & Tarr, 2008; M. Wang et al., 2022). Finally, together with increased youthfulness, health, and femininity, these 2D complexion features drive perceptions of attractiveness (Fink et al., 2006; Matts et al., 2007), and do so cross-culturally with the exception of skin tone lightness (Zhan et al., 2021). Critically, all these complexion-driven judgements can impact social

trait perception—for example, when including stimuli that also vary in 2D complexion, additional social trait dimensions that capture youthfulness/attractiveness (Sutherland et al., 2013, 2018) and femininity (Lin et al., 2021) become relevant. Additionally, though Oosterhof and Todorov (2008) did not initially investigate 2D complexion variations, subsequent elaborations revealed that at least some broad complexion characteristics such as skin tone luminance impact perceptions of both trustworthiness (lighter skin tones) and dominance (darker skin tones; Todorov & Oosterhof, 2011). Thus, much like 3D shape, variations of 2D complexion that are indicative of adaptive cues such as age, health, and attractiveness can influence fundamental social trait perceptions.

A final adaptive cue that has received much attention within 3D shape and has been more recently investigated within 2D complexion is resemblance to emotions. Unlike 3D shape, the features of 2D complexion that resemble emotions typically serve to augment or disambiguate facial expressions of emotion (Thorstenon, 2018). For example, in line with psychophysiological data (e.g., Drummond & Quah, 2001; Kreibig et al., 2007), pallor might distinguish fear from surprise, while skin reddening might increase the perceived intensity of anger. Consistent with this, observers associate increased facial reddening with anger and decreased facial reddening with sadness (Thorstenon et al., 2018). Interestingly, other approach- (e.g., happiness) and avoidance-oriented emotion (e.g., fear) showed similar patterns of red associations (Thorstenon et al., 2018), suggesting that the link between skin colouration and emotion may be better explained by broader motivational stances. Supporting this, manipulating the redness of neutral and emotionally expressive faces increases perceptions of health and attractiveness regardless of expression, but increases perceptions of dominance and aggressiveness only when the face is expressing anger, and increases perceptions of friendliness only when the face is expressing happiness (Thorstenon & Pazda, 2021). Further, machine-learning models trained to categorize emotion from neutral faces perform best when both shape and complexion information is supplied, and their outputs can be used to obtain novel faces with specific shape and complexion features that reliably signal dominance and trustworthiness to human observers (Albohn & Adams, 2021a). Together, this body of work highlights the role of 2D complexion in social perception, with an emerging yet important focus on social trait judgements.

1.3.3 Changeable facial features

Beyond the stable characteristics of 3D shape and 2D complexion, human faces can momentarily display a wide range of dynamic characteristics, including facial expressions, eye gaze, and head movements. Given the attention that *resemblance* to facial expressions of emotion has received, it is perhaps unsurprising that a large body of work has also investigated social trait perception from *actual* facial expressions of emotion (e.g., Hareli et al., 2009; Hess et al., 2009a; Knutson, 1996). Specifically, observers infer dominance from facial expressions of both anger (e.g., Hareli et al., 2009; Knutson, 1996) and disgust (e.g., Ueda et al., 2017), while trustworthiness is typically inferred from happy facial expressions

(e.g., Calvo et al., 2017; Hess et al., 2009a). However, unlike dominance and anger, the association between trustworthiness and happiness is somewhat less clear. This is because facial expressions of happiness are often reduced to—or at least focused on—smiles. However, different types of smiles can not only communicate different emotions beyond happiness (e.g., embarrassment, politeness; Niedenthal et al., 2010) but can also be perceived as dominant (Rychlowska et al., 2017). Reflecting this ambiguity, some evidence shows that perceptions of trustworthiness are less influenced by facial expressions of emotion than perceptions of dominance (Olszanowski et al., 2019), while other work shows that both social trait judgements are most influenced by facial expressions of emotion (Sutherland et al., 2017). Finally, data-driven investigations have revealed that observers associate similar—yet not identical—facial expressions with trustworthiness and happiness (Hensel et al., 2021), and that such trustworthy facial expressions can reverse judgements of untrustworthiness inferred from static facial appearance (Gill et al., 2014). Thus, despite some contrasting results, this body of work highlights the role that facial expressions play in social trait perception.

Eye gaze is another changeable facial feature that has been shown to influence social trait judgements. Specifically, averted eye gaze decreases perceptions of trustworthiness (Abbott et al., 2018; Kaisler & Leder, 2016), as does incongruent gaze cuing—that is, directing eye gaze away from the target—in learning tasks (Manssuer et al., 2016). Similarly, directed gaze guides dominant judgements (Main et al., 2009) and potentiates perception of dominance-related emotions (i.e., anger; Adams & Kleck, 2003, 2005). Another changeable facial features that can guide social trait perception is head tilt, though results are somewhat contrasting. Specifically, evidence has shown that: both raised and bowed heads increase ratings of untrustworthiness and dominance (Zhang et al., 2020); only bowed (Torrance et al., 2020) or only raised heads (Marshall et al., 2020) increase perceptions of dominance and untrustworthiness; and bowed heads increase perceptions of submissiveness and avoidance-motivated emotions (e.g., fear; Mignault & Chaudhuri, 2003). This contrast in findings may be due to the fact that head tilt affects perceived fWHR, and does so most when the head is raised vs bowed (Hehman et al., 2013), though this explanation is also disputed (Zhang et al., 2020). Nevertheless, when paired with other dominance-related cues such as directed eye gaze, both bowed and raised heads are perceived as more dominant (Toscano et al., 2018). Together, though both eye gaze direction and head tilt seem to impact social trait perception, their specific role is more poorly understood than other invariant and changeable facial features.

1.3.4 Other facial features

Beyond its invariant and changeable features, the appearance of the human face can be further altered by makeup, body modifications, hairstyles, and other accessories—each of which has been shown to bias social trait perception. While wearing makeup increases ratings of the attractiveness of female faces (e.g., V. R. Mileva et al., 2016; Mulhern et al., 2003), the type of makeup used can influence which other social traits are perceived from

them. Specifically, natural makeup increases perceptions of both competence and trustworthiness (Etcoff et al., 2011), while dramatic makeup decreases perceptions of leadership ability (James et al., 2018) and increases perceptions of narcissism (Mafra et al., 2024). Further, relative to no makeup, female faces wearing both light and dramatic makeup are perceived as more likely to engage in dominant and manipulative tactics, though this effect is mediated by perceived attractiveness (Schneider & Morón, 2023). Other aesthetical choices, such as hair colour and styles, can also impact how faces are socially perceived, with brunettes being perceived as more competent (Kyle & Mahler, 1996; Takeda et al., 2006) and loose hair styles being perceived as warmer than braids (Klatt et al., 2016). Additionally, body modifications such as facial tattoos and piercings decrease ratings of trustworthiness (Timming & Perrett, 2016) as well as attractiveness and warmth (Resenhoeft et al., 2008), increase perceptions of the guilt of defendants (Brown et al., 2018; Funk & Todorov, 2013), and decrease ratings of hireability (Timming et al., 2017). Finally, accessories such as eyeglasses can increase perceptions of competence and intelligence (Peterson et al., 2022). Nevertheless, while these aesthetical modifications can influence how faces are socially perceived, their evaluations are likely to change over time and with culture (Lieberson, 2000). Given this, in this thesis I control for these facial features to constrain my investigation and avoid possible confounds.

1.3.5 Non-facial features

Although human faces are highly salient and rich sources of social information, non-facial features can also powerfully influence social trait perception. Amongst these, the voice has received much attention. Vocal chord length, and with it voice pitch, is sexually dimorphic (Titze & Martin, 1998). Because of this, lower pitch drives judgements of masculinity (e.g., Cartei et al., 2014). This association in turn influences ratings of attractiveness differently for male and female speakers, with lower vs higher pitch being perceived as attractive in men vs women (Borkowska & Pawlowski, 2011; Y. Xu et al., 2013). Relative to dominance, however, lower pitch drives perception regardless of the sex of the speaker (Borkowska & Pawlowski, 2011; Puts et al., 2007). When other vocal qualities beyond pitch are considered (e.g., timbre), the human voice can also guide perceptions of trustworthiness and likeability (McAlear et al., 2014), as well as reward and affiliation (Wood et al., 2017), and even emotions (Arias et al., 2018). Additionally, local accents increase perceptions of both the competence and warmth of ethnic out-groups (Hansen et al., 2017). Interestingly, much like social trait perception from faces, social trait judgements of human voices can also be synthesized according to two primary dimensions of dominance and trustworthiness (McAlear et al., 2014), underscoring the importance of these judgements for human social perception.

Perceptions of dominance from voice pitch are thought to be at least partly mediated by perceptions of body size (e.g., Armstrong et al., 2019; Y. Xu et al., 2013), which is another non-facial features that has been shown to impact social trait judgements. For example, higher body weight is associated with positively valenced traits (e.g., trustworthy), while

gender-stereotypic body shape (i.e., shapely women, muscular men) is associated with high agency traits (e.g., dominant; Hu et al., 2018). Further, while perceptions of trustworthiness are primarily driven by facial features, perceptions of dominance are best inferred from a combination of the face and body (Hu & O'Toole, 2023). Notably, much like face-based judgements, perceiving social traits from the body occurs spontaneously and automatically (Wildman & Ramsey, 2023), which further demonstrates the fundamental nature of social trait judgements. Beyond body shape, other bodily cues such as body language can influence social trait perception. For instance, gesture speed is associated with ratings of emotional stability (Koppensteiner, 2013), and perceptions of trustworthiness vs dominance are associated with low vs high body movement expansiveness and quantity (Koppensteiner et al., 2016). Finally, clothing can guide judgements of competence and social class (Oh, Shafir, & Todorov, 2020) as well as ethnic group membership (Freeman et al., 2011).

Group membership, whether based on ethnicity or other social categories such as age, gender, and social class, is a critical source of variance in human social perception. Specifically, a large body of work shows that different social traits are attributed to male vs female (e.g., Becker et al., 2007; Sutherland et al., 2013), young vs old (e.g., Hess et al., 2023; Jaeger et al., 2020), rich vs poor (e.g., Bjornsdottir & Rule, 2017; Bjornsdottir et al., 2024), and own-ethnicity vs other-ethnicity group members (e.g., Blair et al., 2002; Hutchings et al., 2024; Xie et al., 2021). Relative to the latter, most evidence shows that other-ethnicity faces—which most typically refers to faces that are not White (see section 1.4 for further discussion)—are judged in line with ethnic stereotypes (Eberhardt et al., 2006; Ponsi et al., 2016; Xie et al., 2021), and these judgements are potentiated when the face appears more prototypical (Blair et al., 2002; Hutchings et al., 2024; Kleider-Offutt et al., 2018). For example, Romanians are perceived as less competent and warm than Italian conationals (Ponsi et al., 2016), while both White and Black faces with more Afrocentric features (e.g., bigger lips, darker skin tone; see Maddox, 2004) are perceived as less trustworthy (Hutchings et al., 2024) as well as having more traits stereotypically associated with the category "Black" (e.g., lazy, aggressive; Blair et al., 2002). Notably, these negative evaluations occur in spite of the fact that Black and East Asian faces resemble positive emotions more than White faces (Adams et al., 2022; Zebrowitz et al., 2010). This in turn suggests not only that social trait perception of other-ethnicity faces is powerfully shaped by group membership and consequent out-group biases and stereotypes, but also that the facial features driving social trait perception from own- vs other-ethnicity faces may differ. Critically, the field's persistent over-reliance on White faces and White Western observers (see section 1.4), coupled with the traditional use of methods that do not enable causal explanations (see section 1.5), limits direct examinations of this.

1.4 Social trait perception knowledge is White- and Western-centric

Current understanding of social trait perception—including theoretical accounts (see section 1.2) and proposed drivers (see section 1.3)—is fundamentally constrained by a key bias in the field: a persistent over-reliance on White faces and White observers from a Western cultural background (Cook & Over, 2021; Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018). Disregarding this central limitation does not only fail to question and test the generalizability of current knowledge, but may also introduce bias. Illustrating this, the fundamental dimensions of dominance and trustworthiness were first obtained from White Western observers' ratings of White faces (Oosterhof & Todorov, 2008, see also subsection 1.2.1). Though subsequent investigations have not—to date—entirely disputed these findings, they have revealed variance according to diversity in both the face (e.g., face sex; Sutherland et al., 2015) and the observer (e.g., culture; B. C. Jones et al., 2021). Similarly, most empirical testing of the facial features driving social trait perception is based on White faces (Cook & Over, 2021). Yet, when ethnic diversity is considered, findings diverge (e.g., Adams et al., 2022; Hutchings et al., 2024; Zebrowitz et al., 2010). Despite this, majority of current research on social trait perception remains centred on White faces and White Western observers. In this section, I review the existing evidence from social trait perception research and related fields that shows how diversity in both the face and the observer can impact social judgements.

1.4.1 Ethnic diversity in the face

Human faces vary considerably in both shape and complexion according to ethnicity (e.g., see Farkas et al., 2005; Maddox, 2004; Y. F. Wen et al., 2015). From these variations, observers readily and automatically categorize others according to ethnicity (e.g., Ito & Urland, 2003; Kubota & Ito, 2017; Rhodes et al., 2009)—for example, observers categorize face ethnicity within 200ms even when other task demands are in place (Kubota & Ito, 2017). The automaticity of these perceptions is further reflected in well-reported and cross-culturally generalizable Other-Race Effects (ORE), where other-ethnicity vs same-ethnicity faces are categorized more efficiently but individuated and thus remembered more poorly (see Kawakami et al., 2022, for a review). Within the context of social trait perception, ethnicity categorizations can bias social judgements in line with such in-group vs out-group effects—for example, ethnically diverse faces manipulated to appear trustworthy are evaluated more positively when they also match the ethnicity of the observer (Birkás et al., 2014; Rostovtseva et al., 2024), suggesting that observers may be less able to recognize trustworthy features in other-ethnicity faces. Additionally, negative judgements of other-ethnicity faces may be influenced by the activation of ethnic stereotypes. Supporting this, White observers high in racial prejudice perceive other-ethnicity faces as more untrustworthy (Stanley et al., 2011; Valmori et al., 2023), cooperate less with other-ethnicity game partners (Kubota et al.,

2013), require more cues to perceive other- vs same-ethnicity faces as trustworthy (Charbonneau et al., 2020), and have more similar mental representations of other-ethnicity and untrustworthy faces (Dotsch et al., 2008). In a similar vein, prejudiced White observers perceive anger—a negatively valenced social message associated with dominance (Said et al., 2009)—using fewer cues from other- vs same-ethnicity faces (Hugenberg, 2005). Together, this work shows that categorizing others according to ethnicity considerably influences social trait judgements in line with in-group vs out-group effects and ethnic stereotypes.

Importantly, ethnicity categorization is not a prerequisite for the activation of ethnic stereotypes (Ito & Tomelleri, 2017). Rather, in line with modern theories that propose automatic associations between features and traits (e.g., Over et al., 2020; Stolier et al., 2018, see also subsections 1.2.5 and 1.2.6), facial features that are more prototypical of a given ethnicity can bypass categorization while still influencing social trait perception in stereotype-consistent ways (Maddox, 2004; Maddox et al., 2022). Illustrating this, both White European and Black African faces with more Afrocentric features (e.g., darker skin tone, fuller lips; Maddox, 2004) are judged as being less trustworthy (Hutchings et al., 2024), and as having other traits stereotypically associated with the category "Black African" (e.g., hostile; Blair et al., 2002), even when having access to more diagnostic behavioural evidence (Blair et al., 2005). Similarly, Black African faces with fewer Afrocentric features are evaluated less negatively (e.g., Deska et al., 2020; Eberhardt et al., 2004; Hagiwara et al., 2012; Kleider-Offutt et al., 2018) and experience fewer real-world negative outcomes (Blair et al., 2004; Eberhardt et al., 2006). Together, this work demonstrates that ethnic appearance, with or without correct ethnic categorization, impacts social trait perception. However, the White-centric bias of current causal facial feature models (e.g., Oosterhof & Todorov, 2008, see also subsection 1.2.1) limits representation of this important source of variance and thus understanding of *how* ethnic facial features are integrated in social trait perception.

A final limitation of current facial feature models is that many of the facial cues proposed to drive social trait judgements—including sexual dimorphism and resemblance to emotions (see section 1.3 for further details)—vary in their appearance and evaluation across ethnicities. Relative to sexual dimorphism, the facial features associated with perceptions of masculinity from White European male faces (e.g., fWHR) do not elicit the same judgements from Turkish (Özener, 2012) and Siberian male faces (Mezentseva et al., 2024). Similarly, facial features considered sexually dimorphic in White Europeans (e.g., see Fink et al., 2005) are not so in Mongolian populations (Rostovtseva et al., 2020). Indeed, recent evidence shows that sexual dimorphism varies considerably across ethnic groups (Kleisner et al., 2021). Similarly, resemblance to facial expressions of emotion does not only vary across face ethnicities—with Black and East Asian faces being overall more structurally similar to approach-related emotion expressions (e.g., happy, fear) than White faces, and vice versa for anger (Adams et al., 2022; Zebrowitz et al., 2010)—but also differentially affects social trait judgements, with happy-resembling features increasing positive judgements of White faces, while only lessening (but not reversing) negative judgements of Black and East Asian faces (Zebrowitz et al., 2010). Together, this work suggests that many of the commonly

reported social trait facial features may fail to easily generalize to non-White faces, and/or may differentially influence social trait perception.

1.4.2 Ethnic and cultural diversity in the observer

Like many other cognitive processes, social trait perception is influenced by culture. For instance, though both Western and East Asian observers infer social traits from faces (e.g., H. Lee et al., 2015; Shimizu, 2012), Western observers make more spontaneous social trait judgements (Shimizu et al., 2017), are more likely to ascribe higher intensity to these judgements (Maeng et al., 2022), and are more sensitive to priming effects of previously learned face-trait associations (Na & Kitayama, 2011) than East Asian observers. Further, the extent to which different social trait dimensions underpin spontaneous perceptions varies across cultural boundaries (B. C. Jones et al., 2021). For example, when East Asian observers are asked to make spontaneous social trait judgements of faces, additional dimensions including youthfulness/attractiveness and masculinity/femininity emerge (e.g., Lin et al., 2021; Sutherland et al., 2018). Additionally, while dominance is often associated with physical strength in a Western context (e.g., Oosterhof & Todorov, 2008; Toscano et al., 2016), in East Asian culture it can instead reflect intellectual superiority and capability in ways similar to the Western concept of competence (Sutherland et al., 2018; H. Wang et al., 2019). Thus, the cultural background of the observer influences both the extent and the underlying conceptual structure of social trait judgements.

Despite these known differences, evidence shows that ratings of dominance and trustworthiness of same- and other-ethnicity faces made by Western and East Asian observers are highly consistent across face ethnicities and cultural groups (e.g., Albright et al., 1997; A. Wang et al., 2024; Zebrowitz et al., 1993). However, this does not necessarily mean that diverse observers use the same facial features to arrive at the same perception (see Newell, 1998; Schyns et al., 2022, for further discussion). Rather, a growing body of work suggests that the facial features used for social trait perception vary according to both the ethnicity and culture of the observer. For example, the facial features used to perceive White faces as trustworthy overlap with those associated with attractiveness for East Asian but not White Western observers (Y. Xu et al., 2013). Similarly, trustworthy judgements of Buryat male faces made by Buryat observers show declining agreement with observers from other world regions as a function of geographic distance (Rostovtseva et al., 2024). Finally, White Western observers rely more heavily on lower Spatial Frequency (SF) information when judging the trustworthiness of same- vs other-ethnicity faces (Charbonneau et al., 2020; Silvestri et al., 2022), and both Western and East Asian observers require fewer cues to perceive own-ethnicity faces as trustworthy (Mo et al., 2022). Together, this work shows that, despite the reportedly high similarity of social trait ratings across ethnicities and cultures, the causal facial features underpinning social trait judgements can nevertheless vary. Moving beyond rating-level examinations to understand *how* this variance occurs is central to not only advancing current feature-based models of social trait perception, but also better under-

standing—and thus possibly tackling—differences in evaluation and consequent treatment of diverse individuals.

Finally, as with face diversity (see subsection 1.4.1), ethnic and cultural diversity in the observer influence the perception of social cues relevant for social trait perception, including emotions and ethnic prototypicality. Relative to emotion perception, abundant evidence now shows that culture considerably impacts how emotions are perceived (e.g., Elfenbein & Ambady, 2002; Fernández-Dols, 2013; Jack, 2013; Jack et al., 2016) and subsequently evaluated (e.g., Krys et al., 2014; Tsai et al., 2016). Despite this, most work examining the use of emotion cues in social trait perception relies on Western-centric prototypes of facial expressions of emotion (e.g., Montepare & Dobish, 2003; Said et al., 2009; Zebrowitz et al., 2010). Similarly, the facial features that are perceived as more prototypical for a given ethnicity vary according to ethnicity of the observer (Ma et al., 2018; Norman et al., 2024; Strom et al., 2012). For example, White observers are more sensitive to ethnic variations in 3D shape, while East Asian and Black African observers rely more heavily on changes in skin tone (Strom et al., 2012). Thus, while ethnic prototypicality is known to influence social trait judgements, it remains unclear whether and how this occurs for diverse observers. Together, this work highlights a lack of clarity in not only the facial features that are specifically used for social trait perception by diverse observers, but also of those features that reflect other social cues theorized to influence these judgements.

1.5 Data-driven methods to diversify knowledge of social trait perception

Diversifying knowledge of social trait perception requires not only the inclusion of diverse faces and observers, but also the adoption of methods that can objectively reveal, rather than mask, any cultural, ethnic, and/or individual differences. Critically, traditional hypothesis-driven methods tend to constrain investigations to few preselected effects guided by White- and Western-centric theory (Jack & Schyns, 2017; Jack et al., 2018). Additionally, the use of uncontrolled face stimuli (e.g., Peterson et al., 2022; Sutherland et al., 2018) forbids causal explanations of which facial features drive social trait perception, including those that are similar or vary according to diversity in the face and/or the observer. Finally, commonly used group-level analyses risk erasing potentially relevant individual differences (Jack & Schyns, 2017; Medin et al., 1993). In this section, I discuss how the data-driven psychophysical method of reverse correlation, coupled with a high-fidelity generative model of 3D human faces and a per-observer modelling approach can alleviate these limitations and reveal whether and how the facial features used for social trait perception vary according to diversity in both the face and the observer.

1.5.1 Reverse correlation

Reverse correlation is a data-driven, psychophysical method first used in auditory perception (Ahumada & Lovell, 1971) and now growingly popular in social face perception (e.g., Dotsch et al., 2011; Oliveira et al., 2019; Oosterhof & Todorov, 2008; Stolier et al., 2018; Sutherland et al., 2018). Broadly defined, reverse correlation seeks to quantify and reveal the stimulus variations that guide behaviour. To do so, observers are presented with a large number of variations of a given stimulus class (e.g., faces varying in 3D shape) and rate each according to a given category (e.g., trustworthiness). Importantly, the stimulus variations are sampled agnostically with few a priori assumptions, therefore enabling testing of the entire stimulus space covered by the variations rather than few preselected features as is common in hypothesis-driven investigations (Brinkman et al., 2017; Dotsch & Todorov, 2012; Jack & Schyns, 2017). This is particularly valuable for social face perception, given that the large natural variations of facial appearance could translate into a near-infinite number of hypotheses (Dotsch & Todorov, 2012; Jack & Schyns, 2017). Further, because each agnostically sampled stimulus may or may not contain the information needed to perceive a given category, it is each observer's subjective ratings—rather than the experimenter's assumptions—that parse out the stimulus variations that are category-irrelevant from those that are diagnostic—i.e., needed to perceive a given category (Brinkman et al., 2017; Jack & Schyns, 2017). Finally, by relying on subjective perception, such diagnostic information can serve as a proxy for the observer's mental representations of a given category (Brinkman et al., 2017; Jack & Schyns, 2017) such as the facial features they associate with different social traits (e.g., Oliveira et al., 2019; Oosterhof & Todorov, 2008; Sutherland et al., 2018).

Reverse correlation offers several advantages key to diversifying knowledge of social trait perception. First, by agnostically sampling information, reverse correlation does not constrain investigations to the facial features deemed relevant in White- and Western-centric theory (Jack et al., 2018). Second, by relying on each observer's subjective perception, reverse correlation unmask any potential cultural, ethnic, and/or individual differences in perception (Jack & Schyns, 2017). Finally, by objectively quantifying the diagnostic stimulus features, reverse correlation provides a common information space that enables direct comparisons across diverse faces and observers (Jack et al., 2018). Nevertheless, these advantages hinge on the information sampling approach taken. Common approaches include superimposing white noise (e.g., random greyscale pixels) onto a base image (e.g., a neutral face) to create perceptually different yet random face stimuli (e.g., Dotsch et al., 2011; Jack, Caldara, & Schyns, 2012), or more directly sampling large sets of uncontrolled face stimuli (e.g., Peterson et al., 2022; Sutherland et al., 2016, 2018). However, the former approach produces results that are difficult to formally analyse (see subsection 1.3.1 for further discussion), and the latter approach prohibits objective explanations of which specific stimulus features drive perception. To overcome these limitations, a fruitful alternative is to sample information directly from a higher-order, multivariate information space bound by the characteristics of the stimulus class of interest, such as the features of 3D shape and 2D complexion of human faces.

1.5.2 Generative model of 3D human faces

Generative models have a long history in the behavioural sciences, spanning ethology (e.g., Tinbergen, 1948), neuroscience (e.g., Hubel & Wiesel, 1962), and more recently psychology (e.g., Oosterhof & Todorov, 2008; Peterson et al., 2022; Salatiello et al., 2021; Sobieszek et al., 2024). Broadly defined, generative models learn patterns within a given data set (e.g., set of human faces) to generate novel stimuli that follow the same broad parameters as the learning set (e.g., a novel human face; Jack & Schyns, 2017). The Generative Model of 3D human Faces (GMF; Yu et al., 2012; Zhan, Garrod, et al., 2019) is one such generative model based on the high-resolution 3D captures of 402 real human faces varying in ethnicity, sex, and age (ethnicity: 245 White European, 149 East Asian, 8 Black African; sex: 232 female, 170 male; age: mean = 28.19 years, SD = 14.65). Specifically, the GMF represents facial feature variance, separately for 3D shape and 2D complexion, as two complementary components: the categorical sources of facial feature variance, including ethnicity, sex, age, and their interaction, captured by applying a General Linear Model (GLM) to the 3D vertices and pixel values that describe each face in the database; and the identity-related sources of facial feature variance, captured by applying a Principal Component Analysis (PCA) to the GLM residuals (see Zhan, Garrod, et al., 2019, for more details). Thus, each face represented in the GMF is mathematically described as the sum of the facial features that capture the average appearance of a face for a given ethnicity, sex, and age *plus* the facial features that distinguish each individual within the same categorical parameters (Yu et al., 2012; Zhan, Garrod, et al., 2019).

Unlike other generative approaches commonly used in the field including FaceGen (Oosterhof & Todorov, 2008) and, more recently, Deep Neural Network-based models (e.g., Peterson et al., 2022), the GMF offers high parametric control and precise feature quantification which enable causal explanations of which specific facial features drive social trait perception (Yu et al., 2012; Zhan, Garrod, et al., 2019; see also Jack & Schyns, 2017). Further, despite its unbalanced sample of face ethnicities, the GMF has high and comparable expressivity and fidelity for faces varying in ethnicity as well as sex and age (see Supplementary materials 7.1.1 for a formal demonstration; see also Zhan, Garrod, et al., 2019). Finally, by separating category-related and identity-related facial feature variations, the GMF can generate facial stimuli that comprise the same facial feature variations across ethnic categories and thus enable direct comparisons. Together, this makes the GMF a powerful platform to generate a high number of ethnically diverse yet readily comparable face stimuli that can be quantitatively described and thus formally analysed.

1.5.3 Per-observer modelling

Classic reverse correlation paradigms—and indeed most experimental approaches in psychology (see McManus et al., 2023, for further discussion)—rely on group-level averaging (e.g., Dotsch et al., 2011; Oosterhof & Todorov, 2008; Sutherland et al., 2018). This can be problematic for several reasons. Most broadly, group-level averages do not necessarily

reflect the behaviour of any one individual in a sample (Speelman & McGann, 2013, 2020). Therefore, while measures of central tendency can be informative, they can also misrepresent individual effects (e.g., see Fisher et al., 2018, for a demonstration). Additionally, averaging obscures potentially relevant individual differences (Jack et al., 2018; Medin et al., 1993). This is particularly problematic for investigations that focus on human diversity, as cultural expectations (e.g., Shimizu et al., 2017), social group membership (e.g., Kawakami et al., 2022), and stereotype knowledge (e.g., Fiske, 2017) are all likely to vary across individuals. Finally, within reverse correlation paradigms specifically, averaging not only inflates the rate of type I errors (Cone et al., 2021) but also prevents understanding of how prevalent any given effect may be—a key step in deriving generalizable, causal explanations of social perception (see Schyns et al., 2022, for further discussion).

Per-observer analyses can alleviate many of these concerns (see Ince et al., 2022, for further discussion). Specifically, rather than averaging across observers, this approach models behaviour within each individual and thus preserves individual differences (Medin et al., 1993). This in turn provides built-in replications—that is, each observer acts as an independent replication of the same experiment, and replications of any effect can therefore be demonstrated across N observers in the tested sample (Ince et al., 2022). Extending from this, per-observer results can be used to provide an estimate of the prevalence of these effects in the sampled population (Donhauser et al., 2018; Ince et al., 2021), thereby justifying claims (if any) about the pervasiveness of a given behaviour (Speelman & McGann, 2020). Because of these advantages, in this thesis I model the facial features driving social trait perception for each observer separately, and constrain my results using a frequentist measure of population prevalence throughout (Donhauser et al., 2018; Ince et al., 2021). Any subsequent group-level inferences are derived from these per-observer models, and I provide information about the underlying distribution of these group-level effects in the Supplementary materials wherever possible.

1.6 The present work

As discussed in this initial chapter, current knowledge of social trait face perception is limited by a persistent over-reliance on White faces and White Western observers. Specifically, though growing evidence suggests that diversity in both the face and the observer influences social trait judgements, no work to date has provided a comprehensive causal explanation for these effects. Consequently, it remains unknown *how* key diversity factors affect the causal facial features used to perceive social traits central to human social life. This in turn questions the generalizability of prominent theories and feature-based models.

This thesis aims to address this critical limitation by modelling and comparing the causal facial features driving social trait perception from diverse faces in diverse observers. Specifically, in Chapter 2 I use a reverse correlation approach to model the specific facial features of 3D shape and 2D complexion that drive perceptions of trustworthiness and dominance from Black African, East Asian, and White European faces in 60 individual White Western ob-

servers. Results show that White Western observers use a combination of facial features that are shared across face ethnicities and represented in current models of social trait perception, *plus* novel facial features that are specific to each face ethnicity. To test whether these results replicate in different observer populations, in Chapter 3 I use the same experimental and analytical approach to model the causal features used for social trait perception from ethnically diverse faces in 60 individual East Asian observers. Results again show that East Asian observers use a combination of shared and face ethnicity-specific features to perceive social traits from ethnically diverse faces. However, the specific features comprised in these feature sets vary between White Western and East Asian observers. Thus, in Chapter 4 and 5 I formally compare the results of the previous two chapters to identify any cross-cultural and cross-ethnic similarities and differences in the causal facial features driving social trait perception. Specifically, in Chapter 4 I examine how the culture of the observer affects social trait perception across face ethnicities, with a further focus on testing whether and how emotion cues theorized to be universal drive social trait judgements across cultures. Finally, in Chapter 5, I examine whether and how observer culture and face ethnicity combine and interact to shape the facial features used for social trait perception.

Chapter 2

Ethnic diversity in facial features challenges generalizability of social trait perception models

Chapter Abstract

Human faces vary considerably in shape and colour and are judged differently across ethnicities. Despite this, current models of the facial features driving social trait judgements remain centred on White faces. To address this lack of diversity, in this chapter I modelled the 3D facial features that drive judgments of trustworthiness and dominance from Black African, East Asian, and White European faces using reverse correlation. Using a high-fidelity 3D generative model of the face, I manipulated random variations of 3D face shape and complexion features, producing 2,400 different faces per ethnicity. Sixty Western, White European observers rated the faces on trustworthiness and dominance. To model the specific facial features that drive the observers' judgments, I regressed their ratings on to the manipulated 3D shape and complexion features. Comparisons across face ethnicities revealed consistent features, such as smiling and frowning, but also ethnicity-specific variations, where observers psychologically exaggerated or diminished the phenotypic features of different face ethnicities. These findings show that face ethnicity biases the features that observers use to make important social judgments, thus challenging the generalizability of current models and highlighting the need for ethnic diversity in psychological research, with direct implications for theories of social perception.

2.1 Introduction

Human faces, with their rich variations of shape and colour (e.g., see Farkas et al., 2005; Maddox, 2004; Y. F. Wen et al., 2015), provide a wealth of social information about others, such as their sex, age, and ethnicity (e.g., Komori et al., 2011; Rhodes et al., 2009; van Rijsbergen et al., 2014). From even a brief glance, humans also infer important personal characteristics about others, such as how trustworthy or dominant they are, from their facial appearance (e.g., Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008). Though rapid (e.g., Willis & Todorov, 2006), spontaneous (e.g., Klapper et al., 2016), and often implicit (e.g., Swe et al., 2020), such judgments can have significant consequences on individual lives, ranging from dating preferences (e.g., South Palomares & Young, 2018) and professional success (e.g., Menegatti et al., 2021), to voting choices (e.g., Joo et al., 2015). Given their central relevance to human societal functioning, a longstanding goal in the human behavioural sciences has therefore been to understand which specific facial features drive social trait judgments.

Face ethnicity is a critical factor in influencing social trait judgments (e.g., Blair et al., 2002; Eberhardt et al., 2004, 2006; Hutchings et al., 2024; Kleider-Offutt et al., 2018; Xie et al., 2021). Despite this, current causal models of the facial features driving fundamental social judgements, such as of trustworthiness and dominance (e.g., Oosterhof & Todorov, 2008; Sutherland et al., 2015; Todorov et al., 2011; Vernon et al., 2014), are predominantly based on White faces (Cook & Over, 2021). Specifically, trustworthy-looking (White) faces tend to be smaller with upturned mouth corners, arched eyebrows, and a lighter skin tone (e.g., Jaeger et al., 2020; Said et al., 2009; Todorov & Oosterhof, 2011; Vernon et al., 2014; Zebrowitz & McDonald, 1991), while dominant-looking (White) faces are often larger with a prominent brow ridge and jaw, and a darker skin tone (e.g., Albert et al., 2021; V. R. Mileva et al., 2014; Todorov & Oosterhof, 2011; Zebrowitz et al., 2003).

This White-centric focus creates two major and related shortcomings: First, it is unclear whether these features generalize across different face ethnicities, or if face ethnicity alters the facial features that observers use to make social trait judgments. Second, we lack understanding of how the features of face ethnicity might bias the features that drive social judgments, including discounting, diminishing, or exaggerating certain features (see also Zhan et al., 2021, for a related example in perceptions of attractiveness). These limitations underscore a considerable gap in understanding the causal mechanisms of social trait face perception (Schyns et al., 2022; see also Newell, 1998) and reflect a bias towards developing a deeper understanding of White face perception specifically (Cook & Over, 2021). With mounting evidence of the inherent problems of a White-centric psychological science (Henrich et al., 2010; Rad et al., 2018; see also B. C. Jones et al., 2021), representing ethnic diversity is crucial to advance theoretical accounts of human social perception and interventions that address and quantify ethnic bias in human interactions.

In this chapter, I address this critical knowledge gap by modelling the specific facial features that drive the perception of two key social traits—trustworthiness and dominance—from

three face ethnicities—Black African, East Asian, and White European. I include these three broad ethnic groups because they are anthropometrically distinct in terms of skin tone and/or facial structure (e.g., Farkas et al., 2005; Y. F. Wen et al., 2015) and implicated in cross-ethnicity social trait perception differences (e.g., Xie et al., 2021). To enable direct comparison with current prominent models of social trait face perception (e.g., Oosterhof & Todorov, 2008), which are not only based on White faces but also derived from White Western observers (see e.g., Cook & Over, 2021), in this chapter I model these perceptions in 60 individual White Western observers.

2.2 Methods

To objectively identify the specific facial features of 3D shape and 2D complexion that drive social trait perception from faces of different ethnicities, I used a high-fidelity generative model of 3D human faces (GMF; Yu et al., 2012; Zhan, Garrod, et al., 2019, see also subsection 1.5.2) combined with the classic psychophysical method of reverse correlation used in ethology (e.g., Tinbergen, 1948), vision science (e.g., Mangini & Biederman, 2004), neuroscience (e.g., Hubel & Wiesel, 1962; Nestor et al., 2016; Zhan, Ince, et al., 2019), engineering (e.g., Thompson et al., 1999), and human social trait perception (e.g., Jack & Schyns, 2017). Figure 2.1 illustrates the approach.

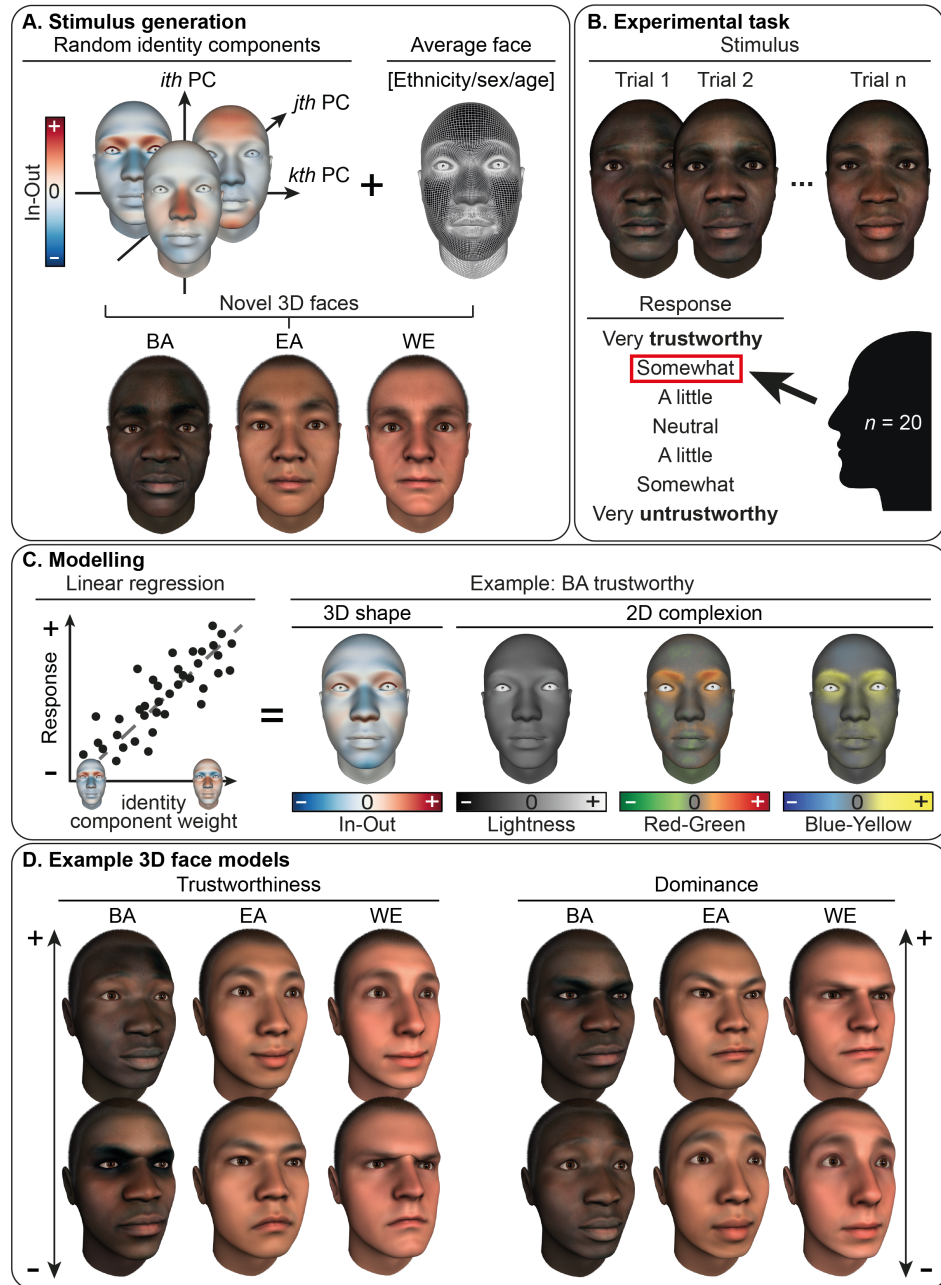


Figure 2.1: **A. Stimulus generation.** On each experimental trial, the generative model generated a novel 3D face identity by adding randomly sampled identity components to the average face for a given ethnicity, sex, and age. Three faces below show an example Black African (BA), East Asian (EA) and White European (WE) male aged 25 years, using the same randomly sampled identity components. **B. Experimental task.** In a between-subjects design, 60 naïve observers ($n = 20$ per face ethnicity) rated each face stimulus according to perceived trustworthiness or dominance on a 7-point bipolar scale ‘very trustworthy’ to ‘very untrustworthy’ in separate randomized blocks. **C. Modelling procedure.** For each face ethnicity, I modelled the facial features that drive the perception of trustworthiness and dominance by linearly regressing the identity component weights sampled on each trial onto each observer’s corresponding social trait ratings, for 3D shape and 2D complexion separately. Colour-coded faces (see colorbars below) show the results for one observer rating Black African faces according to trustworthiness, displayed on an ethnically neutral average male face. **D. Example 3D face models.** Each face shows the facial features that drive trustworthiness (left) and dominance (right) judgements from male faces from one illustrative observer in each face ethnicity condition.

2.2.1 Observers

I recruited a total of 60 White Western observers (30 male, mean age = 21.43 years, SD = 2.81 years). Each observer was pseudo-randomly assigned to one of the three face ethnicity conditions (Black African, East Asian, White European, for a total of 20 observers (10 males, 10 females in each condition). To control for the potential effects of culture on social face perception (e.g., De Leersnyder et al., 2011; Jack, 2013; Sutherland et al., 2018), I only recruited observers from a Western culture background who had with minimal exposure to and experience with non-Western cultures, assessed by questionnaire (see Supplementary materials 7.2.1). To control for the potential effects of viewing same- vs other-ethnicity faces on perception (e.g., McKone et al., 2023) and to compare results with the existing literature, I only recruited observers of White European ethnicity, assessed by self-report. All observers had normal or corrected to normal vision with no history of synaesthesia nor psychological, psychiatric, or neurological conditions affecting visual processing or face perception (e.g., depression, ASD, prosopagnosia) as per self-report. All observers gave written informed consent prior to testing and received £6/hour for their participation, based on the University of Glasgow's standard participation rate at the time of testing. The University of Glasgow College of Science and Engineering Ethics Committee provided ethical approval (Ethics Approval Number: 300160203).

2.2.2 Stimulus generation

To generate novel 3D face identities on each experimental trial, I used a high-fidelity generative model of 3D human faces which is based on the high-resolution 3D captures of real human faces varying in ethnicity, sex, and age (GMF; Yu et al., 2012; Zhan, Garrod, et al., 2019, see also subsection 1.5.2). The GMF represents each novel 3D face identity as the sum of two complementary components:

1. The average 3D face shape and 2D complexion for a specified face ethnicity, sex, and age, which captures the constant sources of facial feature variance associated with each category.
2. Weights for 402 principal components for 3D shape and 402 principal components $\times$ 5 spatial frequency bands for 2D complexion that capture the naturalistic facial feature variance associated with individual identity rather than any categorical information.

Thus, to generate a novel 3D face identity, the GMF randomly samples, for 3D face shape and 2D complexion separately, weights for 402 principal components that capture and control the natural facial feature variance associated with individual identities (henceforth referred to as 'identity components'). For example, in Figure 2.1A, red colour-coding shows the identity-related 3D face shape features that deviate outward from the generative model average (e.g., a more prominent brow) and blue colour-coding shows features that deviate inward from the average (e.g., a smaller nose). These randomly weighted identity components

are then added to the average face for a given ethnicity, sex, and age (represented in 2.1A as a mesh face), resulting in a novel 3D face identity that deviates in specific, measurable, and controllable ways from the selected average.

Importantly, to enable direct comparisons across face ethnicities, I used identical facial feature variations across all ethnicity conditions. That is, I used the same exact 2,400 randomly sampled identity components (1,200 per stimulus sex) in each of the three face ethnicity conditions, for each observer and each social trait rating task. Thus, across all face ethnicity conditions, the face stimuli shared the same age, sex, and random identity components, and differed only according to the ethnicity of the average base face. To illustrate, the three faces in the lower section of Figure 2.1A show the results of adding the same exact identity components to the average face for a Black African (BA), East Asian (EA), or White European (WE) male aged 25 years. Following this approach, I generated a total of 7,200 face stimuli ($[1,200 \text{ identity component variations} \times 3 \text{ face ethnicities} \times 2 \text{ stimulus sex}]$).

2.2.3 Experimental procedure

On each experimental trial, observers viewed a randomly generated 3D face identity and rated it according to social trait (either trustworthiness or dominance on a 7-point bipolar scale ranging from 1 ‘very untrustworthy’ to 7 ‘very trustworthy’ with ‘neutral’ as the mid-point (see Figure 2.1B for an example). Both the face stimulus and response scale remained on the screen until observer response. I instructed observers to respond quickly using their first impressions, using a Graphic User Interface (GUI). After each response, a fixation cross appeared on screen during a 0.5s inter-stimulus interval. In a between-subjects design, each observer thus completed a total of 4,800 trials ($[1,200 \text{ face stimuli} \times 2 \text{ social traits} \times 2 \text{ stimulus sex}]$; see Supplementary materials 7.2.2 for estimation of minimum trial number) across 6 separate 1-hour sessions each comprising 4 blocks of 200 trial with short breaks in-between blocks. I blocked trials by social trait rating task and stimulus sex and randomized the order of the blocks across the experiment for each observer. At the start of each block, I displayed the stimulus sex and the social trait rating task on-screen. Observers completed no more than 3 sessions per day and had at least a 1-hour mandatory break in-between each session. I displayed all face stimuli (average height = 18cm, average width = 11cm) on a $1,930 \times 1,080$ resolution colour-calibrated flat panel monitor at a constant viewing distance of 70cm (ensured via the use of a chin rest during all experimental sessions). All face stimuli thereby subtended 14.87° (vertical) and 9.03° (horizontal) of visual angle, which reflects the average size of a human face (e.g., Ibrahimagić-Šeper et al., 2006) at a typical social distance (Hall, 1966).

2.2.4 Modelling procedure

To model the 3D facial features that drive the perception of trustworthiness and dominance for each individual observer, I linearly regressed the weights of each identity component used on each trial onto the observer’s corresponding social trait ratings 1 – ‘very

untrustworthy’ to 6 – ‘very trustworthy’, for 3D shape (402×1) and 2D complexion (402×5 spatial frequency bands) separately (schematized in Figure 2.1C; see Supplementary materials 7.2.3 for testing of linearity assumption). This produced a quantitative model of the specific facial features that are statistically associated with the observer’s perception of trustworthiness and dominance. I therefore obtained a total of 240 per-observer 3D face models ($[20 \text{ observers} \times 3 \text{ face ethnicities} \times 2 \text{ social traits} \times 2 \text{ stimulus sex}]$; see Figure 2.1D for illustrative examples from one observer in each face ethnicity condition).

2.2.5 Model validation

Prior to analysis of the resulting 3D face models, I computationally validated them using a leave-one-out cross-validation method, for 3D shape and 2D complexion separately. For each observer, face ethnicity, social trait, and stimulus sex, I mass-univariately regressed the ratings of 19 out of the 20 observers onto the identity component weights using ridge regression. At each iteration, I identified which identity components are statistically significantly associated with the 19 observers’ ratings using non-parametric permutation testing ($N = 1,000$). Specifically, I obtained a distribution of chance beta coefficients by randomly shuffling the 19 observers’ ratings and mass-univariately regressing them onto the identity component weights 1,000 times and used the 95th percentile of the resulting distribution as a threshold for statistical significance. I then trained a General Linear Model (GLM) on the 19 observers’ ratings and the statistically significant identity components and used the resulting model to predict the ratings of the left-out observer. Finally, I correlated the model’s predicted ratings with the true ratings of the left-out observer (Spearman’s rho, Bonferroni-Holm corrected, $p < .05$) to obtain an index of how accurately each individual observer’s face model could be predicted by the other observers’ models. I repeated this procedure until all 20 observers in each face ethnicity condition had been tested. Results (see Figure 7.5 in Supplementary materials 7.2.4) showed that most of the 240 face models are valid: 238 3D shape models (119 male: 20 BA/EA/WE dominance; 19 BA trustworthiness, 20 EA/WE trustworthiness; 119 female: 20 BA/WE dominance, 19 EA dominance; 20 BA/EA/WE trustworthiness) and 237 2D complexion models (119 male: 20 BA/EA/WE dominance; 19 BA trustworthiness, 20 EA/WE trustworthiness; 118 female: 19 BA/EA dominance, 20 WE dominance, 20 BA/EA/WE trustworthiness).

2.2.6 Model visualization

To visualize the facial features captured by the 3D face models, I compared the significant 3D shape and 2D complexion identity components of each per-observer model to the average face. For clearest comparisons, each model was compared to the ethnically neutral average male or female face. To do this, I first converted the predicted identity component weights for 3D shape and 2D complexion into vertex values in cartesian space and pixel values in L^*a^*b colour space respectively. I then computed the difference between these converted values and those representing the ethnically neutral average face. Specifically, for 3D shape,

I subtracted the average face's cartesian vertex values from those of each predicted face and calculated the cosine of the angle between the model's difference from the average (i.e., residual) and the vector vertical to the tangent of each vertex of the average face. For 2D complexion, I calculated the difference of the model from the average face for each L^*a^*b colour channel separately.

2.2.7 Population prevalence

To estimate how prevalent the statistically significant effects (i.e., modelled identity components) are in the sampled population, I used the measure of population prevalence (Donhauser et al., 2018; Ince et al., 2021). Specifically, I used my within-observer results to model the population from which the observers are sampled, using binary coding—that is, each possible observer can either use, or not use, a specific identity component to perceive a given social trait, with a given proportion of this population showing a true positive effect. I then performed an inference against the null hypothesis that no observers in the population shows any effect (i.e., the prevalence of the effect is 0; see Donhauser et al., 2018; Ince et al., 2021, for more details). Thus, with 20 observers in each face ethnicity condition and using a $p < .05$ threshold, I can reject the null hypothesis that there is no effect in any observer in the sampled population when significant within-observer results are detected in at least 4 observers.

Together, the data-driven modelling approach described above provides several advantages. First, by agnostically generating facial features from a high-fidelity model of the human face, I can model those that drive social trait perception in individual observers without constraints or biases imposed by prior assumptions (see Jack & Schyns, 2017, for further discussion). Second, by using the exact same randomly generated identity components in each face ethnicity condition, I can isolate how face ethnicity influences the specific facial features that observers use to make each social trait judgment. Third, adopting a per-observer analytical approach preserves individual variation rather than erasing it as traditional averaging approaches can do. This in turn enables any replications of effects to be demonstrated across N observers in the tested sample (Ince et al., 2022) and thus provides an estimate of the prevalence of these effects in the sampled population (Donhauser et al., 2018; Ince et al., 2021).

2.3 Results

Using the approach described above, I modelled the specific facial features that drive the perception of trustworthiness and dominance from three face ethnicities—Black African (BA), East Asian (EA) and White European (WE)—across 60 White Western individual observers. Figures 2.2 and 2.3 show the results for male and female faces respectively, with results aggregated across individual observers ($n = 20$ per face ethnicity condition). For

brevity, and because results for female faces showed a similar pattern of results to male faces throughout, I focus on results for male faces throughout this chapter before highlighting any relevant differences in the results for female faces.

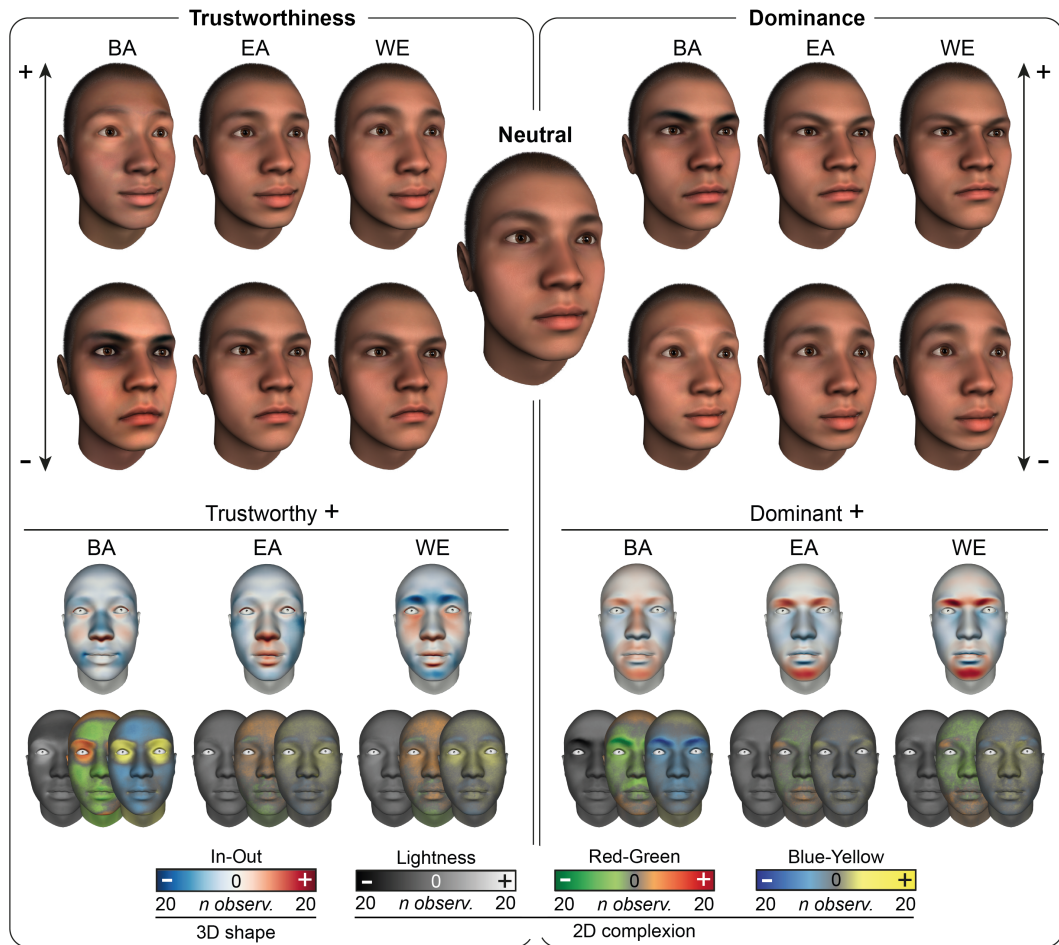


Figure 2.2: Each face shows the 3D models of male faces perceived as trustworthy, untrustworthy (left), dominant, and submissive (right) for each face ethnicity— Black African (BA), East Asian (EA) and White European (WE)—displayed on the same ethnically neutral face (see larger face in the centre for comparison). Results are aggregated across observers ($n = 20$ per face ethnicity condition) and include all identity components that meet or exceed the population prevalence threshold ($n = 4$ observers; faces extrapolated for clarity). Below, colour-coded faces show the results separately for 3D shape and 2D complexion. Colour-coding denotes the feature deviation direction from the average ethnically neutral face; colour saturation shows the number of observers with a statistically significant effect (see colorbars below).

In the top panel, each face shows the resulting male face models for trustworthiness vs untrustworthiness (left) and dominance vs submissiveness (right) for each face ethnicity, with shape and complexion combined and displayed on the same ethnically neutral face displayed in the centre for comparison (see Figure 2.1D for example models displayed on each ethnic average face). Colour-coded faces below show the results for 3D shape and 2D complexion separately, using the same format as in Figure 2.1. Colour saturation shows the number of observers ($n = 20$ per face ethnicity condition) with a statistically significant effect, for all identity components at or above the population prevalence threshold ($n = 4$ observers; Donhauser et al., 2018; Ince et al., 2021). For example, Black African male faces judged as

trustworthy typically have a shorter nose bridge, more upturned mouth corners, and a darker, cooler (greener, bluer) skin tone with a lighter eye socket region, while White European male faces judged as dominant typically have a more prominent brow ridge and chin, with a sallower (greener) skin tone compared to the average face.

Further inspection of the results reveals both clear similarities and differences across face ethnicities. For example, regardless of face ethnicity, male faces perceived as trustworthy typically have a narrower jaw, upturned mouth corners, and a warmer (redder, yellower) eye region than the average ethnically neutral face. Similarly, male faces perceived as dominant typically have a more prominent brow ridge and chin with a sallower skin tone. However, the modelled features also differ across face ethnicities. For example, for trustworthiness, Black African male faces typically have a shorter nose bridge and plumper cheeks with a darker, cooler skin tone and lighter eye socket region, whereas East Asian male faces are narrower with heavier upper eyelids, a larger mouth with fuller lips, and a warmer skin tone, and White European faces have more arched eyebrows, higher cheeks, fuller lips and a warmer skin tone. For dominance, differences across face ethnicities primarily reflect differences in the prevalence of effects—for example, the prominent brow ridge, a commonly reported feature (e.g., Oosterhof & Todorov, 2008), is more frequently associated with White European than East Asian or Black African male faces (see hue variation in the eyebrow region).

Results for female faces (see Figure 2.3) showed an overall similar pattern of results. For example, like for male faces, perceptions of trustworthiness are driven by upturned mouth corners regardless of face ethnicity plus ethnicity-specific differences similar to those for male faces (e.g., shorter nose bridge and lighter eye region for trustworthy-looking BA faces). However, the prevalence of the effects for female faces is on average slightly weaker than for male faces (see hue differences between Figure 2.2 and Figure 2.3).

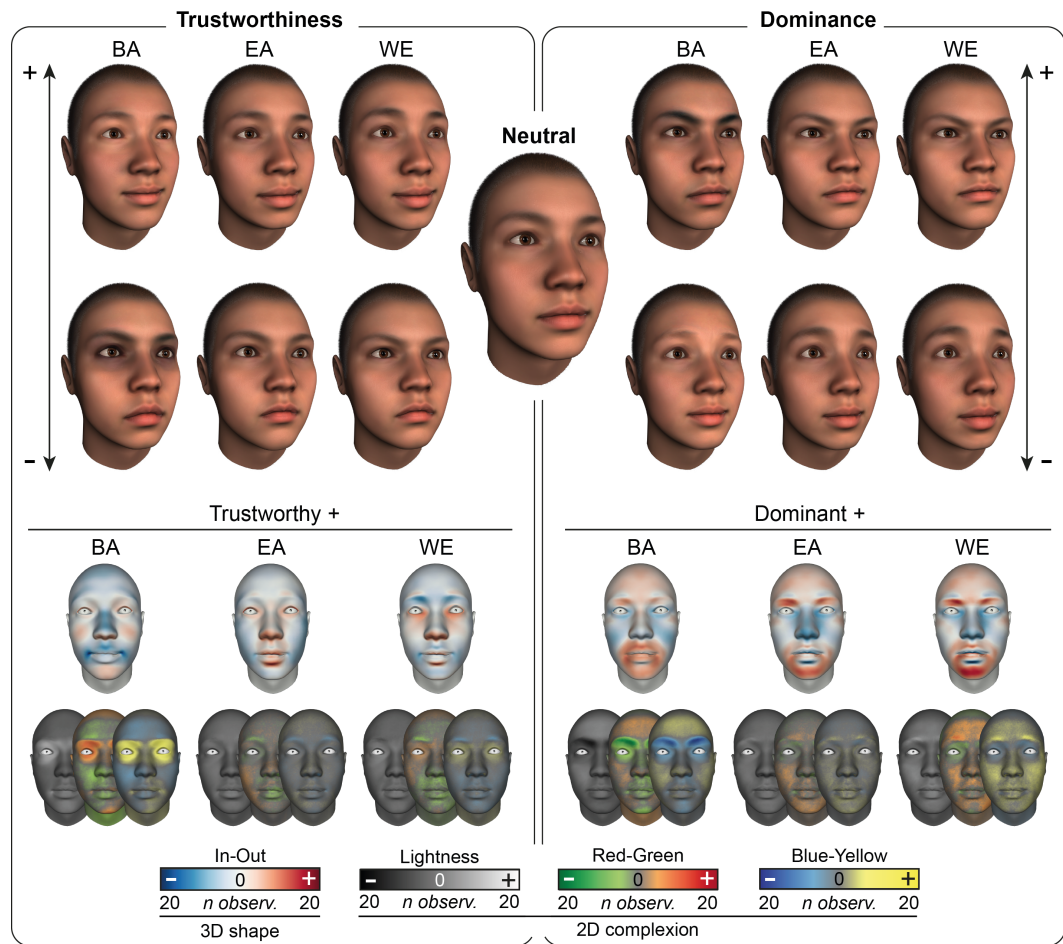


Figure 2.3: Following the same format as Figure 2.2, faces at the top show the 3D models of female faces perceived as trustworthy, untrustworthy, dominant, and submissive for each face ethnicity—Black African (BA), East Asian (EA) and White European (WE)—displayed on the same ethnically neutral face (see larger face in the centre) for comparison. Below, colour-coded faces show the results for trustworthy- and dominant-looking female faces separately for 3D shape and 2D complexion, with colour-coding following the same format as Figure 2.1. Results are aggregated across observers ($n = 20$), with all identity components meeting or exceeding population prevalence threshold of $n = 4$.

2.3.1 Social trait face perception is driven by shared plus specific facial features across face ethnicities

These results suggest that White Western observers' perceptions of trustworthiness and dominance are driven by a set of facial features that are shared across face ethnicities, plus those that are specific to each face ethnicity. To test this formally, I measured the specificity of each modelled facial feature to face ethnicity using the general measure of Mutual Information (MI; Cover & Thomas, 1991; Ince et al., 2017). Specifically, I computed MI between each modelled identity component at or above the population prevalence threshold ($n = 4$ observers) for at least one face ethnicity and each face ethnicity label. To do this, I first represented each identity component and each face ethnicity label as 1×60 ([20 observers $\times$ 3 face ethnicities]) binary-coded vectors as follows: each identity component is represented as statistically significant or not in each validated observer model, separately for

positive and negative identity component values; and each face ethnicity label is represented according to each face ethnicity condition of the validated observer models in turn "Black African" vs "not Black African". A high MI value indicates a strong relationship—i.e., the identity component is specific to a given face ethnicity; a low MI value indicates a weak relationship—i.e., the identity component is not specific to a given face ethnicity and is thus shared across face ethnicities, provided it is significant for at least 4 observers in each face ethnicity condition. I established statistical significance using non-parametric permutation testing as follows: for each identity component, I randomly shuffled the face ethnicity label of each 3D face model and recomputed MI 1,000 times to obtain a distribution of chance MI values. I then used the 95th percentile of the distribution as a threshold for statistical significance. Finally, I used Pearson correlation to discard any identity component with a high MI value but negative correlation with face ethnicity (i.e., the identity component has a strong relationship with face ethnicity because it is repeatedly absent in these 3D face models). I applied the same analysis to positive and negative identity components, for each social trait and stimulus sex separately. Figure 2.4 shows the results for male faces (see Figure 2.5 for female results).

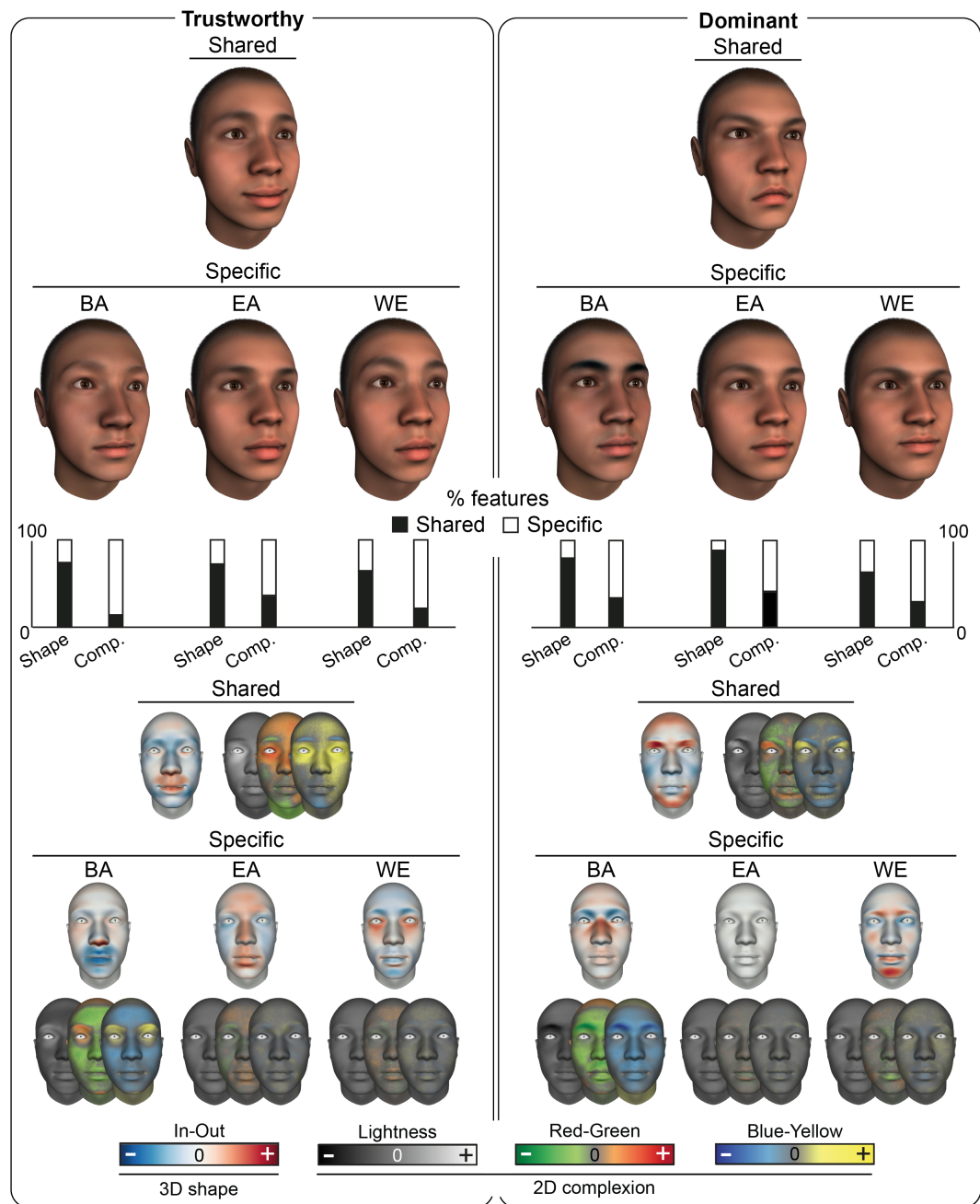


Figure 2.4: Each top panel (left: trustworthy; right: dominant), shows the facial features that are shared (top row) or specific (bottom row) to each face ethnicity (BA—Black African; EA—East Asian; WE—White European), for male faces only. Bar plots below show the proportion of shared (black) vs specific (white) 3D shape and 2D complexion features for each face ethnicity (see Supplementary materials 7.2.5 for further details). Each bottom panel shows the results separately for 3D shape and 2D complexion using the same color-coding as Figure 2.1 (see colorbars below), with values normalized per social trait for display purposes.

Faces in the top panel show, for both trustworthy (left panel) and dominant (right panel) judgements, the facial features of male faces that are shared across face ethnicities (top row) and those that are specific to a given face ethnicity (bottom row), with shape and complexion combined and displayed on an ethnically neutral face as in Figure 2.2. Bar plots below show the proportion of shared (black) vs specific (white) 3D shape and 2D complexion features for each corresponding face ethnicity (see Supplementary materials 7.2.5 for further details).

The bottom panel shows the results separately for 3D shape and 2D complexion using the same colour-coding as Figure 2.1 (see colorbars below), with values normalized per social trait for display purposes. Results confirmed that perceptions of trustworthiness and dominance are driven by a core set of shared facial features plus those that are specific to face ethnicity.

For both trustworthiness and dominance, the male shared features reflect those represented in current models—trustworthy-looking features comprise a narrower jaw, a bigger mouth with upturned mouth corners, and arched, darker eyebrows with warmer skin tones (e.g., Jaeger et al., 2020; Said et al., 2009; Todorov & Oosterhof, 2011; Vernon et al., 2014; Zebrowitz & McDonald, 1991); dominant-looking features comprise a prominent brow ridge with darker eyebrows and lighter brow ridges, a bigger head, a stronger chin, and a narrower face with darker, cooler skin tones (e.g., Albert et al., 2021; V. R. Mileva et al., 2014; Todorov & Oosterhof, 2011; Zebrowitz et al., 2003). Therefore, these results show that certain features are generalizable across male faces of different ethnicities. For the male ethnicity-specific features, the White European-specific features are also represented in current models, including bigger eyes, higher cheekbones, a fuller bottom lip and smaller chin with warmer skin tones for trustworthy judgments (e.g., Zebrowitz & McDonald, 1991), and a more prominent chin, angled eyebrows and smaller eyes with warmer skin tones and sallower lips for dominant judgments (e.g., Albert et al., 2021). However, unlike current models, these results demonstrate that these features are specific to White European faces rather than being generalizable across other face ethnicities. In contrast, Black African- and East Asian-specific features are not represented in current models (e.g., Oosterhof & Todorov, 2008). For trustworthy judgments, BA-specific features comprise a smaller mouth and pointier nose tip with lighter, warmer eye regions against darker, cooler skin tones; and EA-specific features comprise a narrower face and a bigger mouth, nose and forehead with warmer skin tones. For dominant judgments, BA-specific features comprise a longer nose bridge and arched eyebrows with darker, cooler skin tones and eyebrows; and EA-specific features comprise warmer skin tones and sallower lips (but no visible 3D shape features).

Results for female faces (see Figure 2.5; see also Supplementary materials 7.2.5 for further details) are generally comparable with, but not identical to, those for male faces. For trustworthy judgements, both the shared and face-ethnicity specific features are similar to those for male faces but vary in strength. Specifically, the shared features are stronger (i.e., deviate further from the average) than for male faces, while the face ethnicity-specific features are weaker (i.e., deviate less from the average face) than for male faces. For dominant judgements, the shared features are again similar to those for male faces but stronger (i.e., deviate more from the average), whereas the face ethnicity-specific features are much weaker.

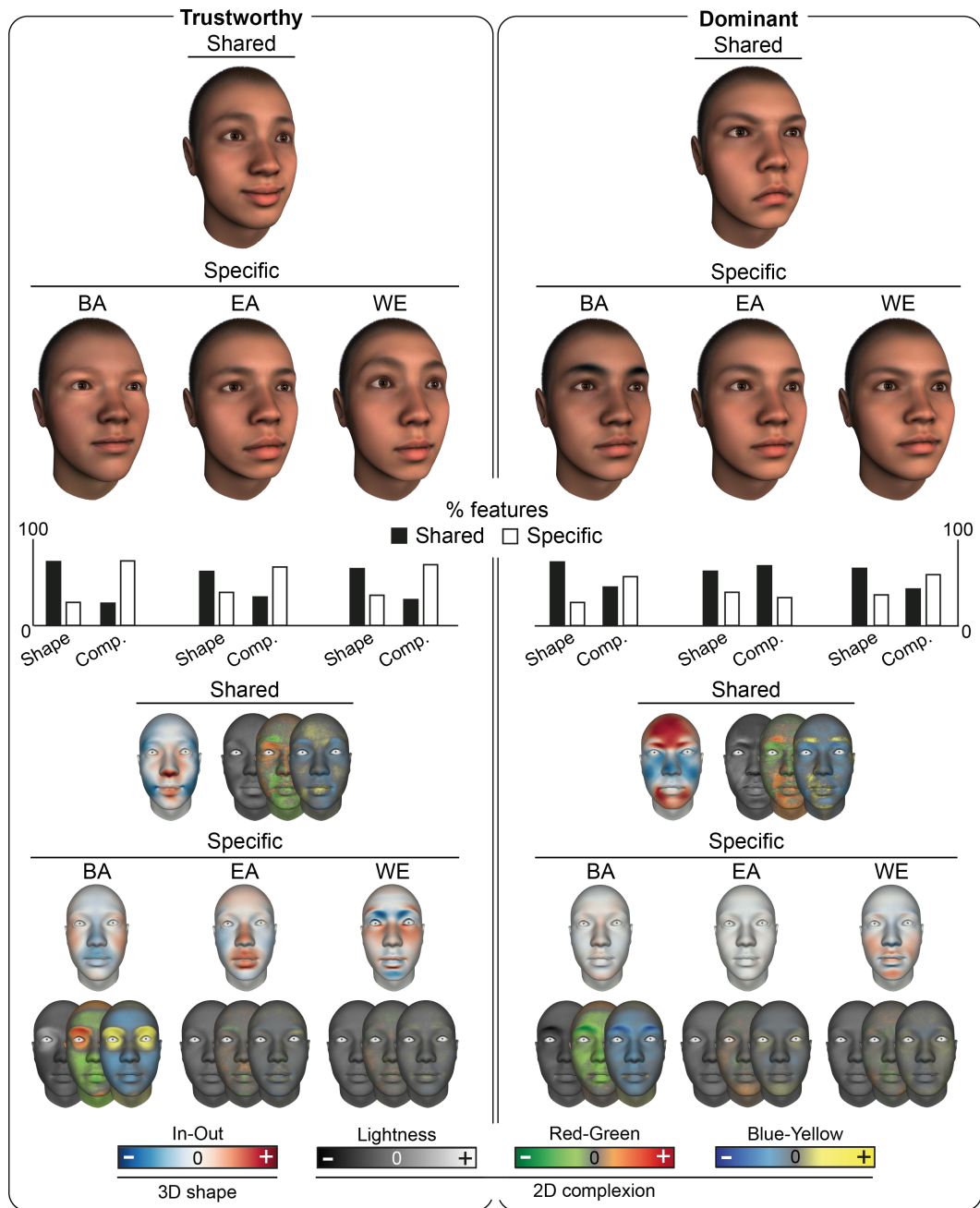


Figure 2.5: Following the same format as Figure 2.4, faces at the top show the facial features that are shared (top row) vs specific (bottom row) to each face ethnicity (BA—Black African; EA—East Asian; WE—White European), for female trustworthy-looking (left panel) and dominant-looking (right panel) faces. Bar plots below show the proportion of shared (black) vs face ethnicity-specific (white) 3D shape and 2D complexion features for each corresponding face ethnicity (see Supplementary materials 7.2.5 for further details). Colour-coded faces at the bottom show the results separately for 3D shape and 2D complexion. Colour-coding follows the same format as Figure 2.1 (see colorbars below), with values normalized per social trait for display purposes.

Together, these results show that additional and novel facial features drive White Western observers' perceptions of trustworthiness and dominance from faces of other ethnicities—a finding previously overlooked due to an over-reliance on White faces.

2.3.2 Validating ethnic variance in the facial features that drive social trait judgements in White Western observers

Finally, I validated the face ethnicity-specific features using two key predictions from the literature. For trustworthiness, existing work suggests that perceptions of trustworthiness is associated with diminished ethnic phenotypic features while untrustworthiness is associated with their exaggeration (e.g., Blair et al., 2002; Eberhardt et al., 2004; Hutchings et al., 2024; Kleider-Offutt et al., 2018). I tested this by correlating the ethnicity-specific features modelled here with the ethnic phenotype features (represented as the ethnic average in the GMF, see Supplementary materials 7.2.6 for details). Specifically, I first represented both the ethnicity-specific trustworthy-looking features and the ethnic phenotype features as 3D coordinates for 31,049 face vertices for 3D shape ($N = 14,319$ front-face vertices) and as 61,218 pixel values (downsampled to every third pixel) for each L*a*b channel of 2D complexion. I then computed the Pearson correlation between each set of face-ethnicity specific features and their corresponding ethnic phenotype features. To constrain results, I only considered statistically significant ($p < .05$, Bonferroni-Holm corrected) and above-chance correlations (see Supplementary materials 7.2.7 for details). Results for male faces confirmed that Black African- and East Asian-specific 3D shape trustworthy-looking features are negatively correlated with their ethnic phenotypic features ($r = -.69$, $p < .001$; $r = -.74$, $p < .001$ respectively). I found no significant results for White European-specific male trustworthy-looking features, nor for 2D complexion. Figure 2.6A shows the results for male faces (see Figure 2.7 for female faces).

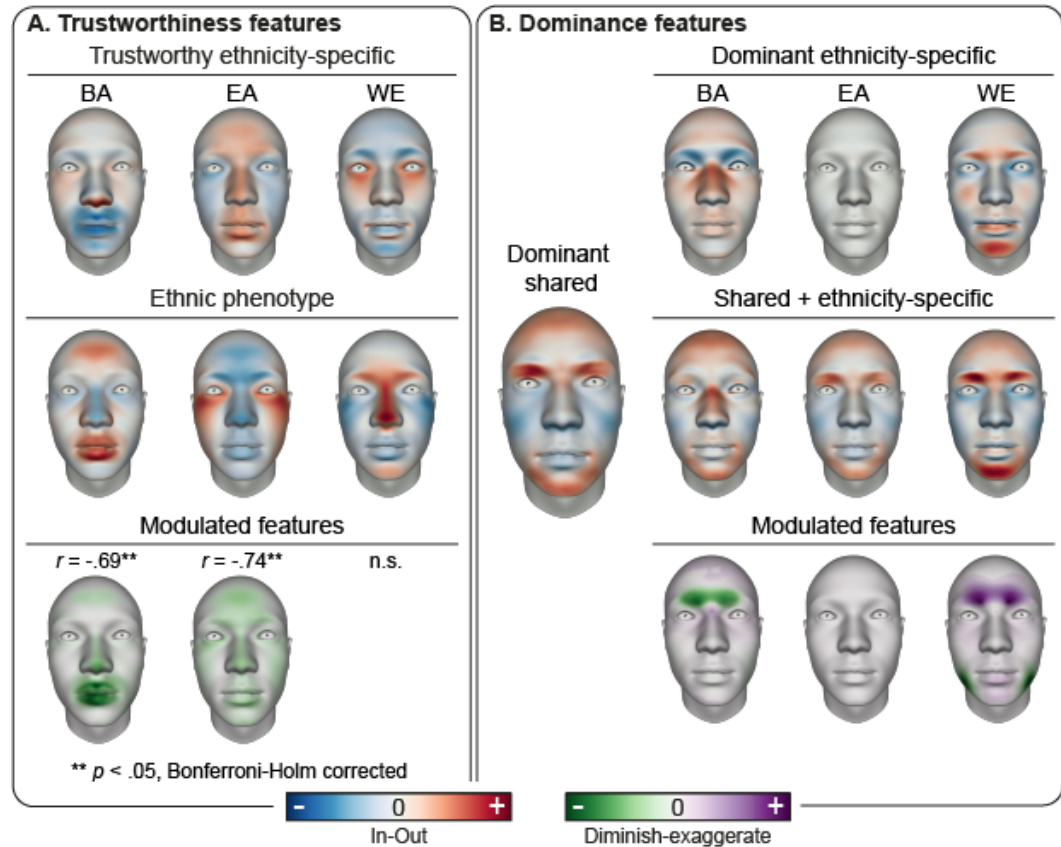


Figure 2.6: **A. Trustworthiness features.** For each face ethnicity, red-and-blue colour-coded faces show the face ethnicity-specific male trustworthy-looking features (top) and their corresponding ethnic phenotypes (centre), using the same colour-coding as Figure 2.1 (see colorbar below). Below, green colour-coded faces show the specific ethnic phenotypic facial features that are diminished for trustworthiness perception (see colorbar below; Pearson's r above each face). **B. Dominance features.** Red-and-blue colour-coded faces show the shared (left), face ethnicity-specific (top row), and shared plus face ethnicity-specific male dominant-looking features (centre row) for each face ethnicity separately, following the same colour-coding format as before (see colorbar below). Green-and-purple colour-coded faces below show the male shared dominant-looking facial features that are diminished vs exaggerated above-chance by the male face-ethnicity specific dominant-looking features, using the same colour-coding as A.

As shown by the red-and-blue colour-coded faces in Figure 2.6A, trustworthy-looking Black African male faces (see first row) have smaller mouths than the average Black African male face (see second row). Similarly, trustworthy-looking East Asian male faces have a more protruding forehead, nose and mouth than the average East Asian male face. To formally test whether these facial features drive the negative correlations (Bonferroni-Holm corrected, above-chance: $r < -.62$ and $r < -.63$ for BA and EA respectively), I computed the Euclidean distance between all pairs of non-zero vertex deviations from the ethnically neutral average face across the correlated sets of face ethnicity-specific trustworthy-looking facial features and ethnic phenotype facial features. Faces in the bottom row of Figure 2.6A show the result, with colour-coding indicating whether the ethnic phenotype features are diminished or exaggerated by the trustworthy-looking male features and colour saturation indicating the magnitude (see colorbar below). These results thus confirm that White West-

ern observers' perception of trustworthiness from other-ethnicity male faces is driven by features that reduce out-group ethnic phenotype appearance.

For dominance perception, existing work suggests that it could rely on fewer cues from other-ethnicity faces that are stereotypically associated with aggression (e.g., Hugenberg, 2005; Said et al., 2009). I tested this by examining whether the male face ethnicity-specific facial features of dominance diminish or exaggerate the male shared features above-chance. To do this, I first added the face-ethnicity specific dominant-looking features to the shared dominant-looking features and represented each resulting feature set (shared and shared + face ethnicity-specific) as 3D coordinates for 31,049 face vertices. I then computed the Euclidean distance of each set to the ethnically neutral average face, thereby obtaining a comparable measure of deviation for each feature set. I then computed the difference between all pairs of non-zero Euclidean distances across the shared and shared plus specific feature sets. Finally, I reverse-signed the results to obtain a $31,049 \times 1$ vector of Euclidean distance differences for each comparison, where negative values indicate that the shared feature is diminished, and positive values indicated that the shared feature is exaggerated. To ensure that the results were above-chance, I repeated the same steps but substituting the face ethnicity-specific dominant-looking features with the features of the randomly generated faces described in Supplementary materials 7.2.7. I then used the 5th (vs 95th) percentile of the distribution of random Euclidean distance differences for each vertex as thresholds for features that are diminished (vs exaggerated) above-chance.

Results confirmed that dominance features are diminished above-chance in dominant-looking Black African male faces and exaggerated above-chance in dominant-looking White European male faces. I found no significant effects for East Asian male faces. Figure 2.6B shows the results—the top row of faces shows the ethnicity-specific facial features; the second row shows how they modulate the underlying shared dominance features (see larger face to the left); the third row shows the specific features that are diminished or exaggerated above-chance (see colorbar below). Thus, these results show that the strength of the features White Western observers use to perceive dominance from male faces varies across face ethnicities.

Results for female faces were again similar, but not identical to, those for male faces. For trustworthiness (see Figure 2.7A), only Black African-specific trustworthy-looking 3D shape features were correlated with their corresponding ethnic phenotype ($r = -.57, p < .001$, Bonferroni-Holm corrected, above-chance). Like the results for male faces, this correlation is driven by the size of the mouth, where BA-specific female trustworthy-looking features diminish its phenotypically larger size. Notably, the results for East Asian female faces showed a similar pattern of effects to those for male faces ($r = -.39, p < .001$, Bonferroni-Holm corrected) but were not above-chance ($r < -.44$). Thus, similar to male faces, perceptions of trustworthiness from Black African female faces are driven by features which reduce ethnic phenotypic appearance. However, unlike male faces, this result is not robust for East Asian female faces. For dominance (see Figure 2.7B), similar to male faces, the female WE-specific features exaggerate the shared protrusion of the brow bone above-chance

and diminish the width of the jaw above-chance. Neither the female BA- nor EA-specific features produced any modulating effects. Thus, similar to male faces, perceptions of dominance from female faces rely on a stronger signal in same-ethnicity faces. However, unlike male faces, the strength of the signal is not weaker when judging Black African faces.

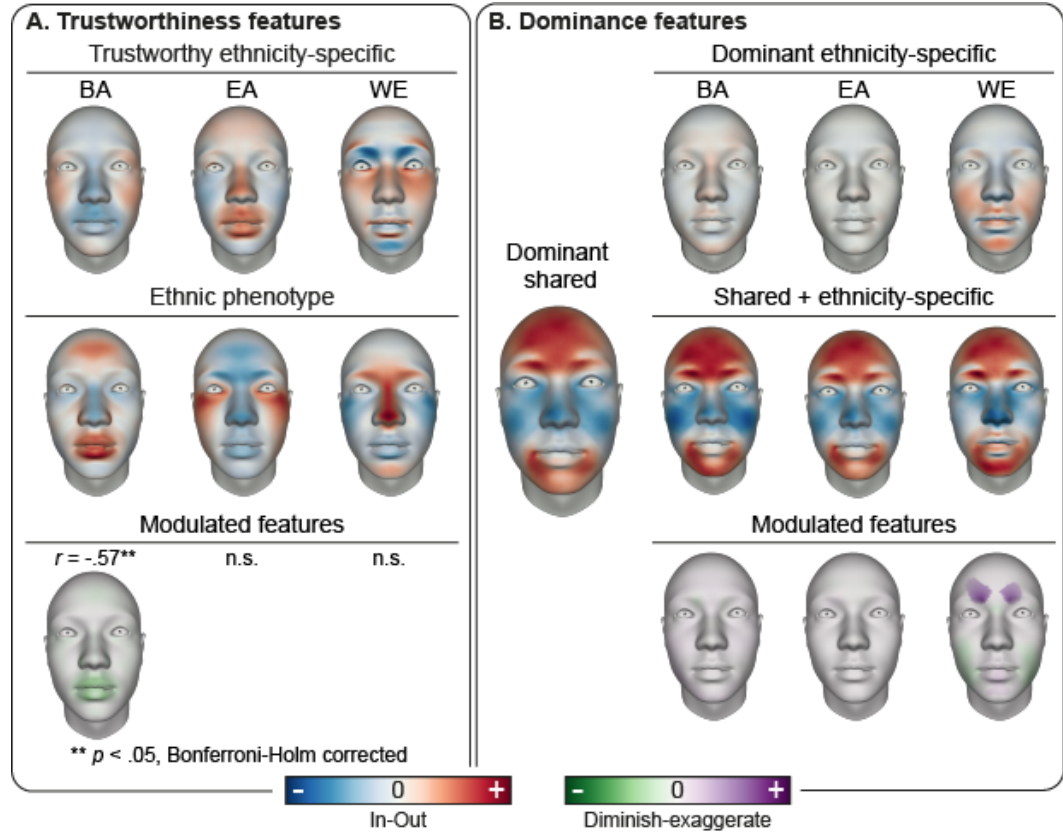


Figure 2.7: Following the same format as Figure 2.6, colour-coded faces in (A) show the face-ethnicity trustworthy-looking features (top), their corresponding ethnic phenotypes (centre), and the specific ethnic phenotypic facial features that are diminished for trustworthiness perception (see colorbar below; Pearson’s r above each face). (B) Colour-coded faces show the shared (left), face ethnicity-specific (top row), and shared plus face ethnicity-specific dominant-looking features (centre row). Green-and-purple colour-coded faces below show the shared dominant-looking facial features that are diminished vs exaggerated above-chance by the face-ethnicity specific dominant-looking features, using the same colour-coding as (A)

2.4 Discussion

In this chapter, I aimed to revisit current models of social trait face perception (e.g., Freeman & Ambady, 2011; Oosterhof & Todorov, 2008; Zebrowitz, 2017) by examining whether and how face ethnicity influences the facial features White Western observers use to make social trait judgments. Using a powerful perception-based data-driven approach (see Jack & Schyns, 2017, for more details) and a high-fidelity 3D generative model of the human face (Yu et al., 2012; Zhan, Garrod, et al., 2019), I modelled the specific facial features that drive the perception of two key social trait dimensions—trustworthiness and dominance—from three face ethnicities—Black African, East Asian, and White European—in 60 individual

white Western observers. Results revealed that, across face ethnicities, social trait face perception is driven by a combination of facial features that are shared across face ethnicities, plus facial features that are specific to each face ethnicity. Specifically, the shared facial features closely mirror those described in current models of social trait face perception (e.g., Oosterhof & Todorov, 2008), as do those that are specific to White European (i.e., same-ethnicity) faces. In contrast, the facial features that are specific to Black African and East Asian (i.e., other-ethnicity) faces are not represented in current models and instead closely mirror ethnic phenotypic features for trustworthiness judgments and modulate the strength of the features for dominance judgments. I discuss the implications of these results for current understanding of social trait face perception below.

I show that face ethnicity influences the specific facial features White Western observers use to make social trait judgments, with direct implications for current models of social trait face perception. Specifically, current influential models propose that social trait judgments are based on a specific set of facial features, such as prominent brows for dominance and upturned mouth corners for trustworthiness (e.g., Oosterhof & Todorov, 2008). However, as such models are primarily based on an ethnically limited set of White European faces (see Cook & Over, 2021), this constrains their ability to represent and therefore causally explain how face ethnicity influences which facial features drive social trait judgements (e.g., Blair et al., 2002; Eberhardt et al., 2004, 2006; Hutchings et al., 2024; Kleider-Offutt et al., 2018; Xie et al., 2021). My results directly demonstrate this limitation by showing that face ethnicity systematically influences the facial features White Western observers use to make social trait judgments. Further, these results suggest that, rather than representing a universal set of features, current models are specific to White faces because they include features that are specific to White European faces. Together, this directly demonstrates that face ethnicity is an influential source of variance in deriving causal explanations of social trait face perception, highlighting the limited generalizability of current models and the importance of including a more diverse range of faces in examining social perception.

My results also correspond with existing theories on the social perception of out-group faces. Specifically, I found that White Western observers' perception of untrustworthiness is driven by facial features that closely resemble ethnic phenotypic features (e.g., Farkas et al., 2005; Maddox, 2004). These results mirror existing findings on ethnic stereotyping, where other-ethnicity faces comprising ethnic phenotypic features tend to be judged according to ethnic stereotypic traits, such as aggressiveness for Black Africans and unfriendliness for East Asians (e.g., Blair et al., 2002; Eberhardt et al., 2004, 2006; Hutchings et al., 2024; Kleider-Offutt et al., 2018; Xie et al., 2021). Parallel to this, I found that White Western observers' perception of trustworthiness is driven by facial features that diminish (i.e., negatively correlate with) other-ethnicity phenotypic appearance. These results correspond with the existing literature on perceptual experience, where faces more similar to those in the observer's environment are perceived as more trustworthy (e.g., Dotsch et al., 2016; Sofer et al., 2015; see also e.g., Tanaka et al., 2004; T. Valentine, 1991). For dominance, I showed that White Western observers' perception of this trait is largely driven by features that are

shared across face ethnicities, but with specific variations across face ethnicities that either diminish or exaggerate those dominant-looking features. Such results correspond with previous work showing that perceptions of anger—a social message related to dominance (e.g., Cabral & de Almeida, 2019; Said et al., 2009)—rely on fewer cues from other-ethnicity vs same-ethnicity faces (Hugenberg, 2005).

Finally, results show that White Western observers' social trait face perception is also driven by a set of facial features that are shared across face ethnicities. This aligns with prominent theories such as the overgeneralization hypothesis (see Zebrowitz, 2017; Zebrowitz & Montepare, 2008, for reviews, see also subsection 1.2.2), which posit that the facial features driving social trait perception originate from an evolved attunement to adaptive social cues, such as age, sexual dimorphism, and facial expressions of emotion. These theories suggest that, because of their adaptive nature, such cues should drive social trait perception regardless of the ethnicity of the face. My result support this by showing that social trait perception is driven by a shared set of facial features that mirror previous findings on the role of adaptive cues—for example, perceptions of trustworthiness are associated with happiness-resembling (e.g., Montepare & Dobish, 2003; Said et al., 2009; Thorstenson et al., 2018) and youth-related facial features (e.g., Fink & Matts, 2008; Zebrowitz & McDonald, 1991), while perceptions of dominance are associated with anger-resembling (e.g., Adams et al., 2012; Hareli et al., 2009; Windmann et al., 2023) and masculine features (e.g., Albert et al., 2021; Batres et al., 2015; V. R. Mileva et al., 2014). By mapping onto these adaptive cues, these results therefore support the notion that social trait judgements central to human social interactions are, at least in part, underpinned by universal attunements to meaningful social cues which generalize across faces of different ethnicities.

Importantly, the current results show which specific facial features observers associate with different social traits, rather than reflecting actual behaviour or disposition. These results also reflect the perceptions of White Western observers perceiving same- and other-ethnicity faces. As both observer ethnicity (e.g., Hugenberg & Sacco, 2008; Tanaka et al., 2004; T. Valentine, 1991) and culture (e.g., De Leersnyder et al., 2011; Jack, 2013) can influence face perception, Chapter 3 will examine whether and how these observer characteristics could influence social trait perception from ethnically diverse faces. Finally, although pervasive (e.g., Todorov et al., 2014), social trait perceptions from static facial appearance can be influenced by other sources of information, including facial expressions (e.g., Gill et al., 2014), body shape (e.g., Hu et al., 2018), voice (e.g., McAleer et al., 2014), and concrete behavioural evidence (e.g., Van Dessel et al., 2019). Future research should explore whether and how these multiple sources of information influence social trait perception when viewing ethnically diverse persons.

In sum, the results presented in this chapter provide a causal demonstration of *how* face ethnicity influences social trait perception by showing that the facial features White Western observers use to perceive social traits psychologically diminish or exaggerate specific features in line with same- vs other-ethnicity stereotype knowledge. Specifically, I show that, across face ethnicities, White Western observers' social trait perception is driven by

facial features that resemble adaptive cues plus facial features that psychologically alter the phenotypic appearance of other-ethnicity faces for trustworthiness, and facial features that psychologically alter the intensity of commonly reported dominant-looking features. These results also highlight the White-centric bias of current prominent models (e.g., Oosterhof & Todorov, 2008) by showing that they include features that are specific to White European faces. Together, this chapter underscores the importance of diversity in providing generalizable accounts of the causal mechanisms underpinning fundamental social judgements.

Chapter 3

Ethnic variance in social trait facial features replicates in East Asian culture

Chapter abstract

Growing evidence suggests that variance in both the ethnicity of the face and the culture of the observer influence social trait perception. Despite this, most knowledge remains centred on White faces and Western observers. To address this bias, in this chapter I reverse correlate the 3D facial features that drive judgments of trustworthiness and dominance from Black African, East Asian, and White European faces in 60 individual East Asian observers. Comparisons of the resulting 3D models showed that, like White Western observers, East Asian observers perceive social traits using a combination of facial features that are shared across ethnicities plus ethnicity-specific variations. The shared features are similar, though not identical, to those used by White Western observers and largely align with previous work. In contrast, the face ethnicity specific features diverge both from those used by White Western observers and previous work. Together, these findings show that variance in both the observer and the face powerfully influence how social traits are perceived. This highlights the importance of moving away from Western-focused research, with direct implications for theoretical development.

3.1 Introduction

Social trait perception is a culturally ubiquitous aspect of social cognition (H. Lee et al., 2015; Shimizu, 2012). Across cultural boundaries, face-based social trait judgements—such as those of trustworthiness and dominance—influence both individuals (e.g., Bente et al., 2014; Y. Wang et al., 2024) and broader society (e.g., Na et al., 2015; Rule et al., 2010). Because of their impact, a large body of work has examined which facial features drive these judgements (e.g., Oosterhof & Todorov, 2008; Said et al., 2009; Sutherland et al., 2016; Vernon et al., 2014; Zebrowitz & Montepare, 2008, see also section 1.3). However, despite their considerable knowledge contribution, most of this work remains centred on White Western observers perceiving White faces (see Cook & Over, 2021, for further discussion; see also Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018). This cultural bias, combined with the White-centric focus discussed in Chapter 2, questions the generalizability of current knowledge of social trait perception.

Culture plays a pivotal role in shaping perceptual and cognitive processes (Kastanakis & Voyer, 2014; Q. Wang, 2016), including those involved in social trait perception (e.g., B. C. Jones et al., 2021; Maeng et al., 2022; Shimizu et al., 2017). For example, relative to Western observers, East Asian observers make fewer spontaneous social trait judgements (Shimizu et al., 2017), ascribe lower intensity to the perceived traits (Maeng et al., 2022), and conceptually structure social traits in different ways (e.g., B. C. Jones et al., 2021; Lin et al., 2021; Sutherland et al., 2018). Additionally, extant cross-cultural findings in visual perception show that East Asian and Western observers process faces differently (e.g., Blais et al., 2008; Caldara, 2017), particularly when extracting social information (e.g., Jack et al., 2009). Finally, culture shapes the norms, expectations, and stereotypes that observers bring to bear during social trait perception (Over et al., 2020). This is particularly relevant when examining the perception of ethnically diverse faces, as both the target and content of ethnic stereotypes is likely to shift with culture (Fiske, 2017).

Despite these known cultural differences and the centrality of culture in modern theories of social trait perception (Over & Cook, 2018; Over et al., 2020, see also subsection 1.2.6), relatively little empirical work has directly examined how non-Western observers perceive social trait, particularly from ethnically diverse faces. Of the available studies, many have argued that social trait judgements are largely universal (e.g., Sutherland et al., 2018; Walker et al., 2011; A. Wang et al., 2024; F. Xu et al., 2012; Zebrowitz et al., 2012). However, such conclusions may be premature. Specifically, if examined at the rating level (e.g., Walker et al., 2011; A. Wang et al., 2024; Zebrowitz et al., 2012), similarities could be inflated by the rating scale (Maeng et al., 2022). Similarly, feature-level similarities (e.g., Sutherland et al., 2018; H. Wang et al., 2019; F. Xu et al., 2012) could be biased by the face stimulus set, where the inclusion of highly salient cues (e.g., facial expressions) or the lack of ethnic variance could skew results. Indeed, when considering ethnically diverse neutral faces, both the cultural background of the observer and the ethnicity of the face influence how the same face is socially perceived (e.g., Charbonneau et al., 2020; Mo et al., 2022; Rostovtseva et

al., 2024; Silvestri et al., 2022). Finally, evidence from the stereotyping literature shows that, akin to prejudiced White observers (e.g., Hugenberg, 2005), both Asian and African observers high in racial bias perceive other-ethnicity faces more negatively (e.g., Qian et al., 2016; Q. Wang et al., 2014).

In sum, current knowledge of social trait perception is biased toward a White- and Western-centric understanding (Cook & Over, 2021; Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018). This leaves open questions about how non-Western observers perceive social traits from ethnically diverse faces, with direct implications for the generalizability of current theories and models of social trait perception (e.g., Oosterhof & Todorov, 2008; Vernon et al., 2014; Zebrowitz & Montepare, 2008). In this chapter, I address this gap by modelling the specific facial features of 3D shape and 2D complexion that drive the perceptions of trustworthiness and dominance from Black African, East Asian, and White European faces in 60 individual East Asian observers. As in Chapter 2, I include these three broad ethnic groups because of their documented differences in appearance (e.g., Farkas et al., 2005; Y. F. Wen et al., 2015) and implications for cross-ethnicity differences in social trait perception (e.g., Xie et al., 2021). Further, I focus on East Asian observers to enable comparisons with previous cross-cultural work (e.g., Maeng et al., 2022; Sutherland et al., 2018; Walker et al., 2011) and due to their wider availability as research participants at the University of Glasgow relative to other cultural groups.

3.2 Methods

To enable direct comparisons with the results presented in Chapter 2, I adopted the same experimental design as before. Below, I therefore only briefly summarize the methodological approach of the present study. For brevity, I do not repeat how the models were visualized nor how the population prevalence threshold was estimated (both procedures are identical to what is described in Chapter 2).

3.2.1 Observers

I recruited a total of 60 East Asian observers (30 male, mean age = 23.23 years, SD = 2.81 years). Each observer was pseudo-randomly assigned to one of the three face ethnicity conditions (Black African, East Asian, White European), for a total of 20 observers (10 males, 10 females) in each condition. As in Chapter 2, I controlled for the potential effects of culture (e.g., De Leersnyder et al., 2011; Jack, 2013) and ethnicity (e.g., McKone et al., 2023) by only recruiting observers of East Asian ethnicity and from an East Asian culture background who had minimal exposure to and experience with non-Eastern cultures, as assessed by self-report and questionnaire (see Supplementary materials 7.3.1) respectively. To further control for cross-cultural contact, I only recruited observers who had lived in the UK for less than three months at the time of testing. All observers had normal or corrected to normal vision with no history of synaesthesia nor psychological, psychiatric, or neurological condi-

tions affecting visual processing or face perception (e.g., depression, ASD, prosopagnosia), as per self-report. To ensure clear understanding of the experimental task, all observers had a minimum IELTS score of 6 (or equivalent). All observers gave written informed consent prior to testing and received £9/hour for their participation, based on the University of Glasgow's standard participation rate at the time of testing. The University of Glasgow College of Science and Engineering Ethics Committee provided ethical approval (Ethics Approval Number: 300160203).

3.2.2 Stimulus generation

This study used the same stimuli as Chapter 2, generated as described previously and illustrated in Figure 2.1A. Specifically, each observer viewed the same 2,400 Black African, East Asian or White European 3D face identities (1,200 female). Across face ethnicity conditions, the face stimuli shared the same sex, age, and randomly sampled identity components and only varied according to the ethnicity of the average base face.

3.2.3 Experimental procedure

This study used the same experimental design as Chapter 2. Briefly stated, each observer rated each 3D face identity according to perceived trustworthiness and dominance in separate tasks using a 7-point bipolar scale (e.g., 'very untrustworthy' to 'very trustworthy' with 'neutral' as the mid-point). All observers were instructed to respond quickly using their first impressions. In a between-subjects design, observers completed 4,800 trials ($[1,200 \text{ face stimuli} \times 2 \text{ social traits} \times 2 \text{ stimulus sex}]$) blocked by social trait rating task and stimulus sex across 6 separate 1-hour sessions, with no more than 3 sessions per day and at least one 1-hour mandatory break between sessions. All face stimuli (average height = 18cm, average width = 11cm) were displayed on a $1,930 \times 1,080$ resolution color-calibrated flat panel monitor at a constant viewing distance of 70 cm, thereby subtending 14.87° (vertical) and 9.03° (horizontal) of visual angle. Prior to taking part in the experiment, participants were provided with a copy of the task instructions translated in Mandarin Chinese by two native speakers.

3.2.4 Modelling procedure

Following the experiment, I modelled the facial features of 3D shape and 2D complexion that drive social trait perception in each individual observer using linear regression as described in Chapter 2 (see Supplementary materials 7.3.2 for testing of linearity assumption). However, unlike White Western observers, some East Asian observers did not provide social trait ratings using the entire rating scale—though they were instructed to do so—and instead avoided using extreme ratings (i.e., 1 = 'very untrustworthy/submissive' and 7 = 'very trustworthy/dominant'). Notably, these conservative response patterns reflect known cross-cultural differences in response style (e.g., C. Chen et al., 1995; T. Johnson et al., 2005;

Maeng et al., 2022). Therefore, prior to modelling, I normalized the social trait ratings of each observer, for each social trait, face ethnicity and stimulus sex separately.

3.2.5 Model validation

Prior to further analyses, I validated each observer model using a leave-one-out cross-validation approach as described in Chapter 2. Results (see Figure 7.11 in Supplementary materials 7.3.3) showed that most East Asian observer models were valid, though these were fewer than the White Western observer valid models: 213 shape models (105 male: 18 BA dominance, 17 EA dominance, 19 WE dominance; 17 BA/EA/WE trustworthiness; 108 female: 17 BA dominance, 19, EA dominance, 20 WE dominance; 18 BA/WE trustworthiness, 16 EA trustworthiness) and 210 complexion models (105 male: 17 BA/EA dominance, 19 WE dominance; 15 BA trustworthiness, 18 BA trustworthiness, 19 WE trustworthiness; 105 female: 17 BA dominance, 18 EA dominance, 20 WE dominance; 16 BA trustworthiness, 17 EA/WE trustworthiness).

3.3 Results

Using the approach summarized above and described in further detail in Chapter 2, I modelled the facial features that drive the perception of trustworthiness and dominance from three face ethnicities—Black African (BA), East Asian (EA) and White European (WE)—in 60 individual East Asian observers. Figures 3.1 and 3.2 show the results for male and female faces respectively, with results aggregated across individual observers ($n = 20$ per face ethnicity condition). As in Chapter 2, results for female faces were overall similar to those for male faces throughout. Therefore, in this chapter I focus on the results for male faces before highlighting any differences in the results for female faces.

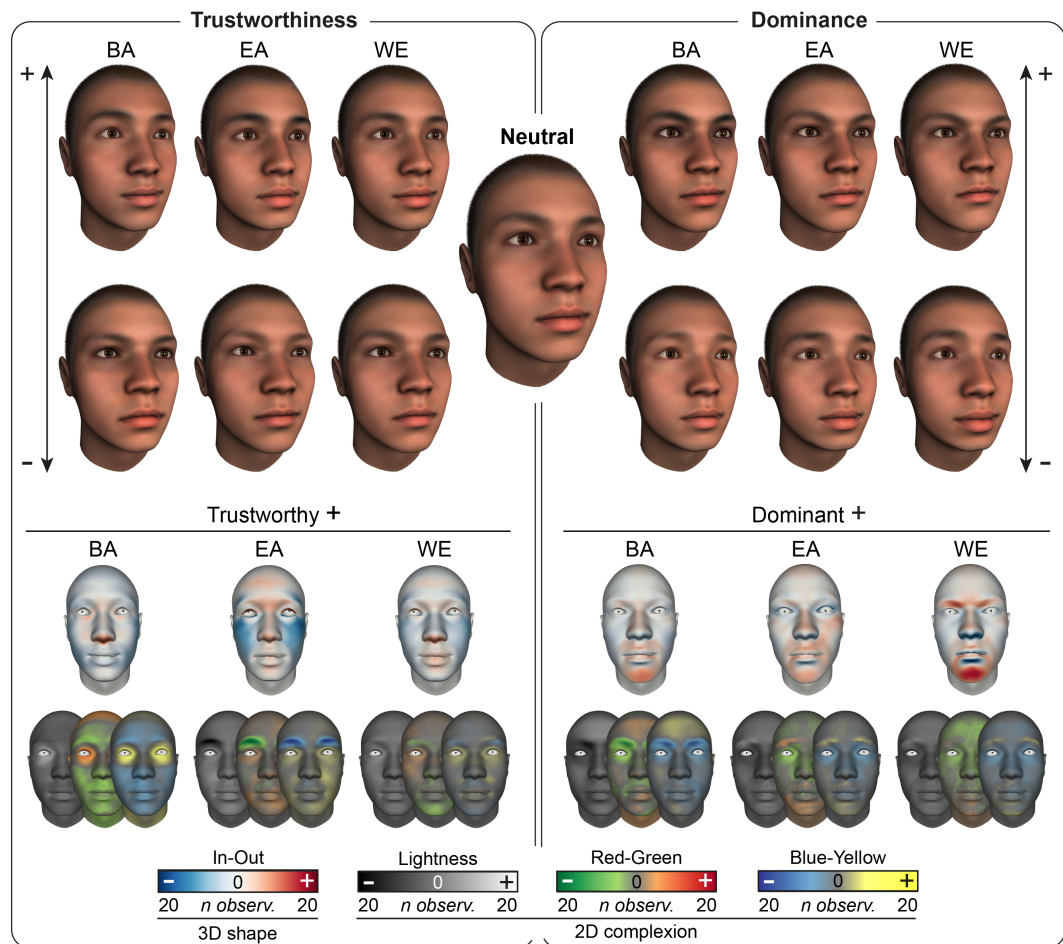


Figure 3.1: Each face shows the 3D models of male faces perceived as trustworthy, untrustworthy (left), dominant, and submissive (right) for each face ethnicity—Black African (BA), East Asian (EA) and White European (WE)—displayed on the same ethnically neutral face (see larger face in the centre) for comparison. Results are aggregated across observers ($n = 20$ per face ethnicity condition) and include all identity components that meet or exceed the population prevalence threshold ($n = 4$ observers; faces extrapolated for clarity). Below, colour-coded faces show the results for trustworthy- and dominant-looking male faces separately for 3D shape and 2D complexion. Colour-coding denotes the feature deviation direction from the average ethnically neutral face; colour saturation shows the number of observers with a statistically significant effect (see colorbars below).

Following the same format as Figure 2.2, faces at the top show the resulting male face models for trustworthiness vs untrustworthiness (left) and dominance vs submissiveness (right) for each face ethnicity, with shape and complexion combined and displayed on an ethnically neutral face (displayed in the centre for comparison). Colour-coded faces below show the results for 3D shape and 2D complexion separately, using the same format as in Figure 2.1. Colour saturation shows the number of observers ($n = 20$ per face ethnicity condition) with a statistically significant effect, for all identity components at or above the population prevalence threshold ($n = 4$ observers Donhauser et al., 2018; Ince et al., 2021).

Like the results for White Western observers (see Chapter 2), results for East Asian observers also show both similarities and differences across male faces of different ethnicities. Specifically, regardless of face ethnicity, male faces perceived as trustworthy are typically narrower with heavier upper eyelids and cooler (greener, bluer) eyebrows, while male faces

perceived as dominant typically have a higher, broader nose bridge, a more juttied jaw, and a prominent chin together together with a darker, cooler eye region and warmer (redder, yellower) chin. However, the prevalence of these similar effects varies by face ethnicity. Specifically, face width is most frequently associated with perceptions of trustworthiness from East Asian faces (see hue variation in the outer cheek and jaw regions), while a prominent chin and tight, thin lips are most frequently perceived as dominant from White European face (see hue variations in the relevant regions). Beyond these differences in prevalence, the modelled facial features also categorically vary according to face ethnicity. For trustworthiness, Black African male faces typically have fuller under eyes and upturned mouth corners together with lighter, warmer eye regions against darker, cooler skin tones; East Asian male faces typically have a more prominent brow bone, a bigger nose, and flatter cheekbones together with lighter, warmer skin tones; and White European male faces typically have upturned mouth corners and fuller lips together with warmer cheeks and cooler chins. For dominance, BA male faces typically have darker, cooler eyebrows against lighter, warmer skin tones; EA male faces typically have deeper-set, angled eyes together with lighter, warmer eyebrows; and WE male faces typically have angled, lighter, and warmer eyebrows.

Results for female faces (see Figure 3.2) are similar to those for male faces. However, the prevalence of these effects across observers is lower than for male faces (see hue differences between Figure 3.1 and Figure 3.2). Further, some of the features used for social trait perception from male faces are not used for female faces—for example, angled eyebrows are perceived as dominant from WE male, but not female, faces.

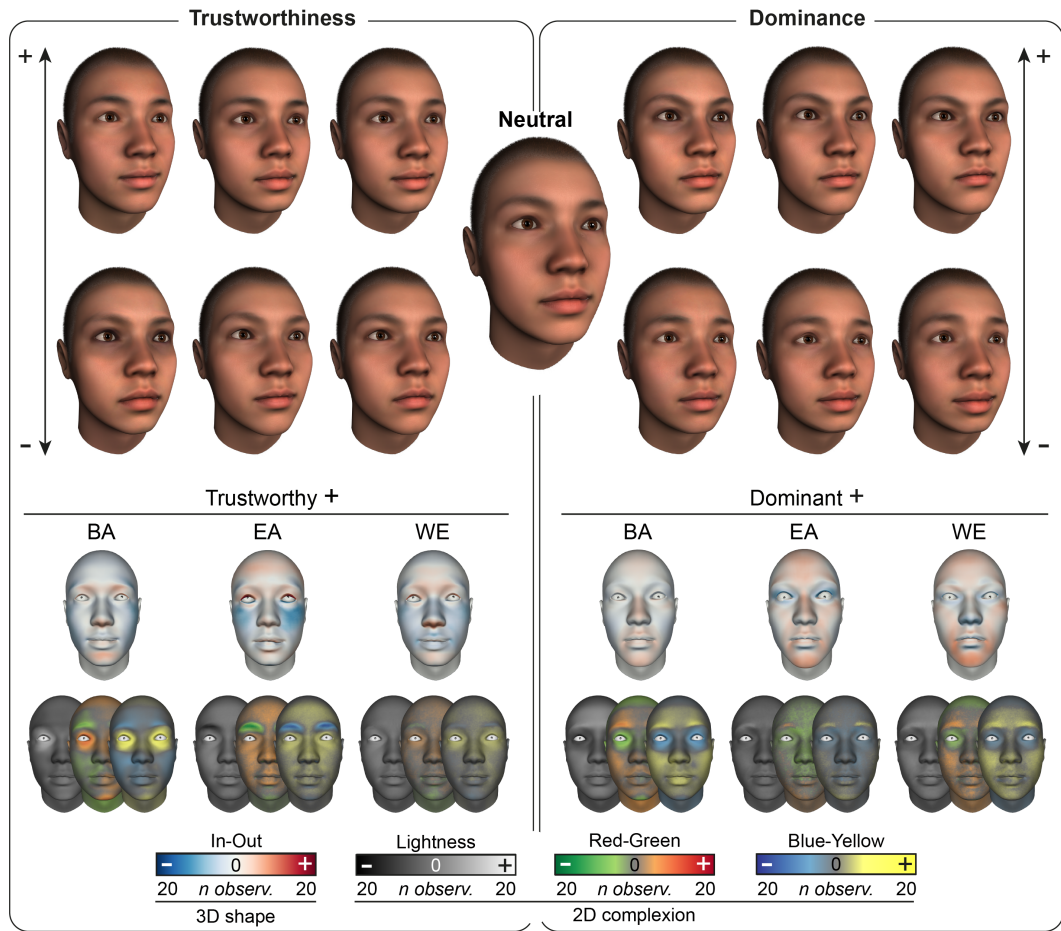


Figure 3.2: Following the same format as Figure 3.1, faces at the top show the 3D models of female faces perceived as trustworthy, untrustworthy, dominant, and submissive for each face ethnicity— Black African (BA), East Asian (EA) and White European (WE)—displayed on the same ethnically neutral face (see larger face in the centre) for comparison. Below, colour-coded faces show the results for trustworthy- and dominant-looking female faces separately for 3D shape and 2D complexion, with colour-coding following the same format as Figure 2.1. Results are aggregated across participants ($n = 20$), with all identity components meeting or exceeding population prevalence threshold of $n = 4$.

3.3.1 East Asian observers use shared plus face ethnicity-specific facial features to perceive social traits

These results suggests that, like White Western observers (see Chapter 2), East Asian observers perceive trustworthiness and dominance from a combination of facial features that are shared across face ethnicities plus those that are specific to each face ethnicity. To test this formally, I measured the specificity of each identity component to face ethnicity using Mutual Information (MI; Cover & Thomas, 1991; Ince et al., 2017) using the same approach described in detail in Chapter 2. Briefly stated, I computed MI between each modelled identity component at or above the population prevalence threshold ($n = 4$ observers) for at least one face ethnicity and each face ethnicity label in turn (e.g., "BA" vs "not BA"), separately for each identity component weight direction (positive vs negative), social trait, and stimulus sex. I established statistical significance using non-parametric permutation testing (n

= 1,000 permutations) and used Pearson correlation to discard any identity component with a high MI value but negative correlation with face ethnicity (i.e., the identity component has a strong relationship with face ethnicity because it is repeatedly absent in these 3D face models). Figure 3.3 shows the results for male faces (see Figure 3.4 for female faces).

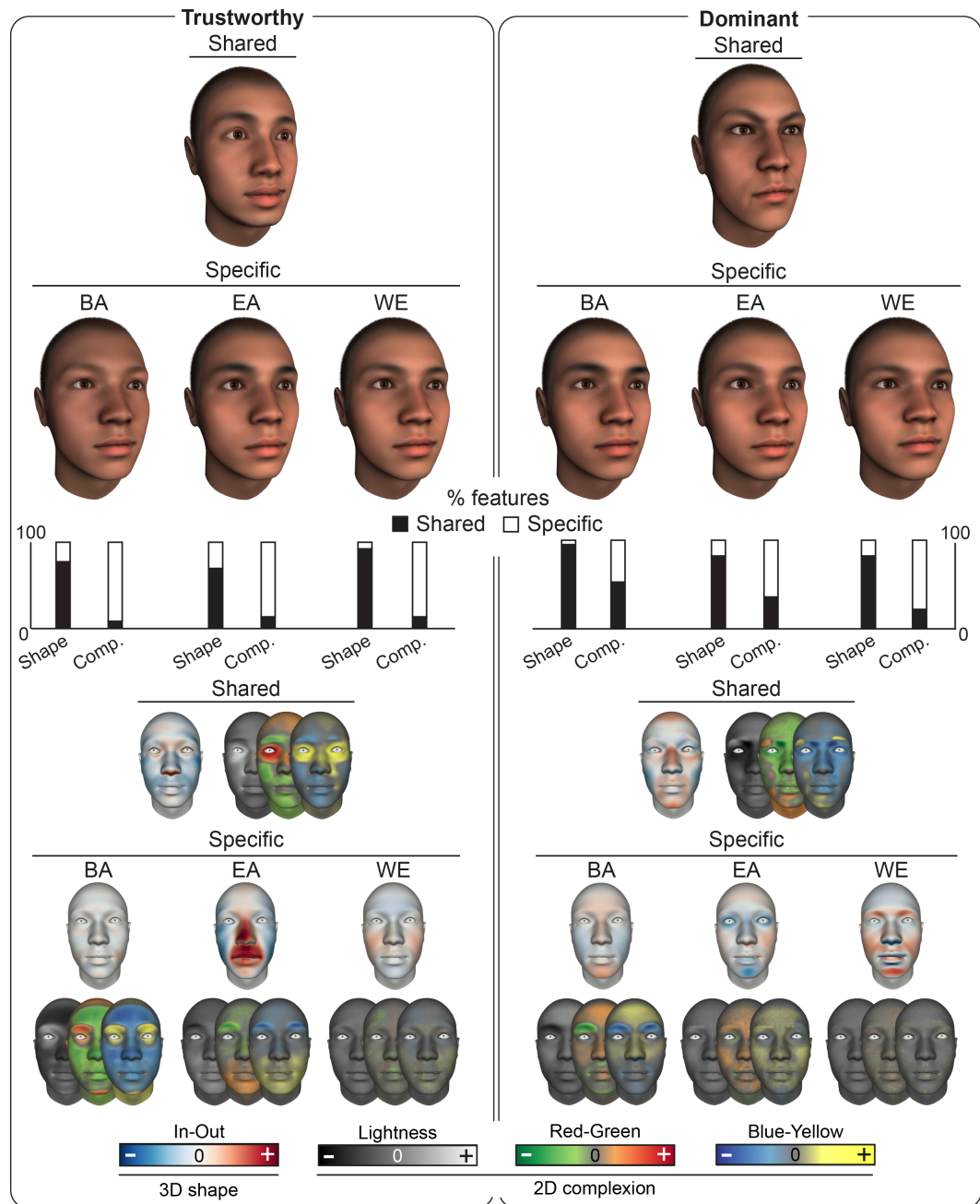


Figure 3.3: Each top panel (left: trustworthy; right: dominant), shows the facial features that are shared (top row) or specific (bottom row) to each face ethnicity (BA—Black African; EA—East Asian; WE—White European), for male faces only. Bar plots below show the proportion of shared (black) vs specific (white) 3D shape and 2D complexion features for each face ethnicity (see Supplementary materials 7.3.4 for further details). Each bottom panel shows the results separately for 3D shape and 2D complexion using the same colour-coding as Figure 2.1 (see colorbars below), with values normalized per social trait for display purposes.

Following the same format as Figure 2.4, faces in the top panel show the facial features

of male faces that are shared across face ethnicities (top row) and those that are specific to each face ethnicity (bottom row) for trustworthy (left panel) and dominant (right panel) judgements, with shape and complexion combined and displayed on the same ethnically neutral face as in Figure 3.1. Bar plots below show the proportion of shared (black) vs specific (white) 3D shape and 2D complexion features for each corresponding face ethnicity (see Supplementary materials 7.3.4 for further details). The bottom panel shows the results separately for 3D shape and 2D complexion using the same colour-coding format as before (see colorbars below), with values normalized per social trait for display purposes.

Results confirmed that perceptions of trustworthiness and dominance from ethnically diverse male faces are driven by a core set of facial features that are shared across ethnicities plus ethnicity-specific variations. For both social traits, most of the shared features are represented in current models. Specifically, the trustworthy-looking shared features comprise a smaller, rounder face with bigger eyes and slightly upturned mouth corners together with a lighter, warmer eye region (e.g., Said et al., 2009; Todorov & Oosterhof, 2011; Vernon et al., 2014; Zebrowitz & McDonald, 1991); and the dominant-looking shared features comprise a longer, narrower face with angled eyebrows, and tight lips with a darker, cooler skin tone (e.g., Albert et al., 2021; V. R. Mileva et al., 2014; Todorov & Oosterhof, 2011; Zebrowitz et al., 2003). However, the remaining shared features are not represented in current models—including heavier upper eyelids together with a darker, cooler skin tone for trustworthy judgements; and a higher nose bridge and angled eyes for dominant judgements. These results therefore show that the facial features that drive social trait perception in East Asian observers are partly generalizable across face ethnicities but are not fully represented in current models. Additionally, the facial features that are specific to each face ethnicity are mostly unrepresented in current models. For trustworthy perceptions, the BA-specific features comprise a slightly fuller jaw together with lighter, warmer eyebrows and chin tips; the EA-specific features comprise a narrower face with a bigger nose and a lower-set mouth together with darker, cooler eyebrows and a lighter, warmer lower face region; and the WE-specific features comprise fuller cheeks together with a slightly sallower (greener) skin tone. For dominant perceptions, the BA-specific features comprise a narrower face with a wider forehead, a bigger nose, and a wider chin together with darker, cooler eyebrows against, lighter warmer skin tones in the upper face region and cooler skin tones in the lower face region; the EA-specific features comprise deeper-set eyes, higher cheekbones, and a smaller chin together with warmer skin tones; and the WE-specific features comprise a more prominent brow ridge and chin, and a wider lower face area.

Results for female faces (see Figure 3.4; see also Supplementary materials 7.3.4 for further details) are similar to those for male faces, with some key differences. For trustworthy judgments, the female shared features largely align with those for male faces but additionally comprise a larger mouth and a lower brow ridge. Additionally, the face ethnicity-specific features differ from those for male faces in specific ways: the BA-specific features comprise sallower—rather than cooler—foreheads and cheeks; the EA-specific features comprise a fuller—rather than narrower—jaw, as well as a flatter brow ridge and a shorter nose; and the

WE-specific features also comprise a pointier nose tip together with a warmer brow bone and lips against cooler skin tones. For dominant judgements, the female shared features resemble those for male faces but are more pronounced. On the other hand, the face ethnicity-specific features are also similar to those for male faces but overall weaker (see hue variations between Figure 3.3 and Figure 3.4). Additionally, some of the face ethnicity-specific features are unique to female faces—for example, the EA-specific trustworthy features further comprise a fuller, rounder jaw, and the WE-specific dominant features do not comprise a more prominent brow bone and instead include downturned mouth corners.

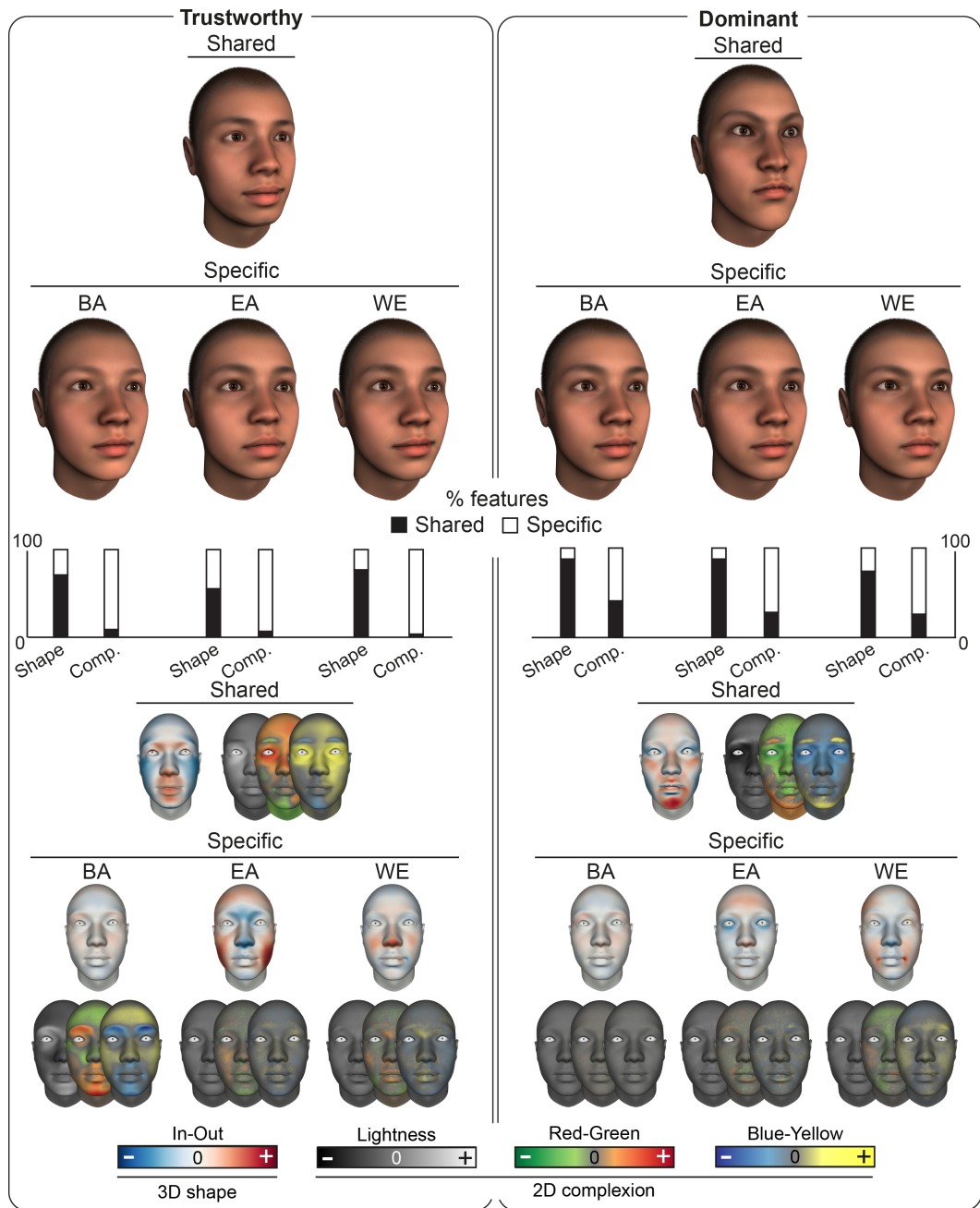


Figure 3.4: Following the same format as Figure 3.3, faces at the top show the facial features that are shared (top row) vs specific (bottom row) to each face ethnicity, for female trustworthy-looking (left panel) and dominant-looking (right panel) faces. Bar plots below show the proportion of shared (black) vs face ethnicity-specific (white) 3D shape and 2D complexion features for each corresponding face ethnicity (see Supplementary materials 7.3.4 for further details). Colour-coded faces at the bottom show the results separately for 3D shape and 2D complexion. Colour-coding follows the same format as Figure 2.1 (see colorbars below), with values normalized per social trait for display purposes.

Together, these results show that the facial features that drive perceptions of trustworthiness and dominance in East Asian observers are influenced by the ethnicity of the face. Further, while most of the facial features that are shared across ethnicities map onto previous findings (e.g., V. R. Mileva et al., 2014; Oosterhof & Todorov, 2008; Said et al., 2009; Todorov et al., 2011; Vernon et al., 2014), those that are specific to face ethnicity do not.

3.3.2 Comparing the shared facial features used by East Asian and White Western observers

The shared facial features used by East Asian observers appear similar, though not identical, to the shared facial features used by White Western observers (see Chapter 2). For example, regardless of face ethnicity, both East Asian and White Western observers perceive smaller faces with bigger eyes and a lighter, warmer eye region as trustworthy, and longer faces with angled eyebrows and a darker, cooler skin tone as dominant. However, only White Western observers perceive arched eyebrows as trustworthy across face ethnicities; and only East Asian observers perceive a higher nose bridge with angled eyes as dominant across face ethnicities. To formally test their (dis)similarity, I compared (Pearson r) the identity components of 3D shape and 2D complexion that are shared across face ethnicities for East Asian and White Western observers, separately for each stimulus sex. Figure 3.5 shows the results as colour-coded matrices (see colorbar below) with thick black outlines denoting statistically significant results ($p < 0.05$, Bonferroni-Holm corrected; others set to 0) for direct comparisons (see diagonal), separately for each stimulus sex (A: Male shared features; B: Female shared features) and for 3D shape (left) and 2D complexion (right; results aggregated across the 5 Spatial Frequency bands).

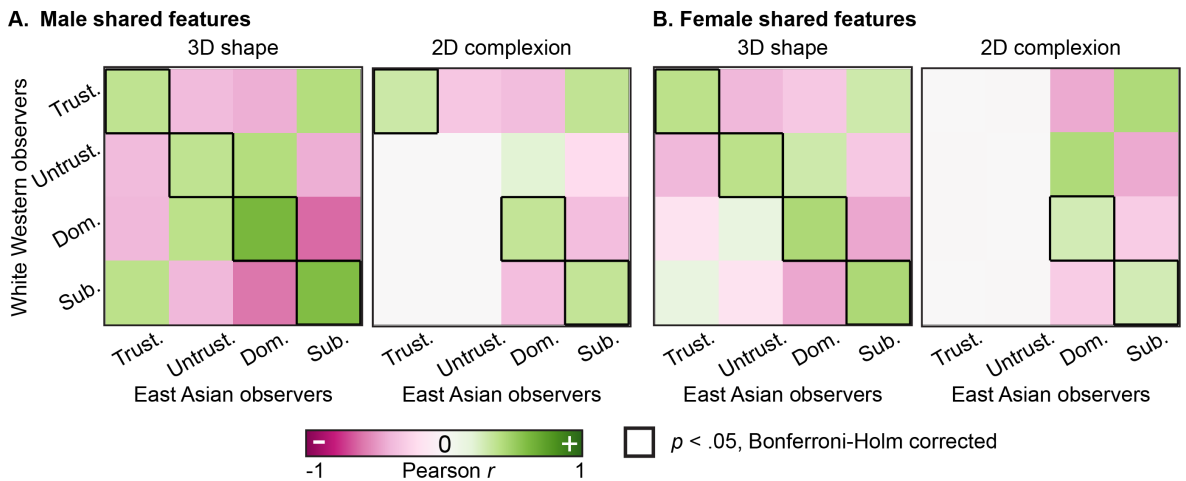


Figure 3.5: In each panel (A: Male shared features; B: Female shared features), colour-coded matrices show the Pearson correlation between the shared social trait features used by East Asian (x axis) and White Western observers (y axis), separately for 3D shape (left) and 2D complexion (right; results pooled across the 5 Spatial Frequency bands). Colour-coding indicates the magnitude and direction of the effect (see colorbar below). Thick black outlines indicate statistically significant correlations ($p < 0.05$, Bonferroni-Holm corrected) for all direct comparisons (see diagonal).

Results show that, for both stimulus sexes, most of the shared social trait features are statistically significantly ($p < 0.05$, Bonferroni-Holm corrected) positively correlated between observer groups. Specifically, the shared features of 3D shape are moderately positively correlated between East Asian and White Western observers for trustworthy (male: $r(400) = 0.37$; female: $r(400) = 0.39$), untrustworthy (male: $r(400) = 0.37$; female: $r(400) = 0.39$), dominant (male: $r(400) = 0.62$; female: $r(400) = 0.44$), and submissive (male: $r(400)$

= 0.59; female: $r(400) = 0.44$). Additionally, the shared features of 2D complexion are positively—although more weakly—correlated between observer groups in both stimulus sexes for dominant (male: $r(2,008) = 0.35$; female: $r(2,008) = 0.28$) and submissive (male: $r(2,008) = 0.35$; female: $r(2,008) = 0.28$), and only for male faces for trustworthy ($r(2,008) = 0.32$). Therefore, these results confirm that the facial features East Asian and White Western observers use to perceive social traits across face ethnicities are similar, but not identical.

3.3.3 Validating ethnic variance in the facial features that drive social trait judgements in East Asian observers

Finally, to better understand the face ethnicity-specific features used by East Asian observers, I tested them against key predictions from the literature as in Chapter 2. Specifically, I first tested whether the trustworthy-looking face ethnicity-specific features negatively correlate with their corresponding ethnic phenotypes (e.g., Blair et al., 2002; Hutchings et al., 2024). To do this, I computed the Pearson correlation between the trustworthy-looking ethnicity-specific features and the ethnic phenotype features represented in the GMF (see Supplementary materials 7.2.6), each represented as 3D coordinates for 31,049 face vertices for 3D shape ($N = 14,319$ front-face vertices) and as 61,218 pixel values (downsampled to every third pixel) for each L*a*b channel of 2D complexion. I only considered statistically significant ($p < 0.05$, Bonferroni-Holm corrected) and above-chance correlations (see Supplementary materials 7.2.7 for details). I found no above-chance significant results for any social trait, face ethnicity, or stimulus sex, neither for 3D shape nor 2D complexion. Thus, the face ethnicity-specific facial features East Asian observers use to perceive trustworthiness do not modulate ethnic phenotypic appearance.

Next, I tested whether the dominant-looking face ethnicity-specific features diminish or exaggerate the shared dominant features (e.g., Hugenberg, 2005; Said et al., 2009). To do this, I followed the same approach described in Chapter 2. Briefly stated, I first computed the Euclidean distance between the ethnicity-specific features and the ethnicity-neutral average face, and between the combination of the ethnicity-specific and shared features and the ethnicity-neutral average face. I then computed the difference between all pairs of non-zero Euclidean distances across the shared and shared plus specific feature sets and reverse-signed the results to obtain a measure of which shared features are diminished vs exaggerated by the addition of the face ethnicity-specific features. Finally, I repeated the same steps using the randomly generated faces described in 7.2.7 instead of the face ethnicity-specific features, and constrained the results using the 5th and 95th percentile of the resulting distributions. I again found no significant results for any social trait, face ethnicity, or stimulus sex, neither for 3D shape nor 2D complexion. Thus, the face ethnicity-specific facial features East Asian observers use to perceive dominance do not modulate the shared features.

Having found no significant results for key literature predictions, I finally tested the (dis)similarity of the face ethnicity-specific identity components used by East Asian and White Western observers using Pearson correlation as in subsection 3.3.2, separately for each

complexion, $r(2,008) = 0.08$, are positively correlated between observer groups. Therefore, while both East Asian and White Western observers use face ethnicity-specific features to perceive social traits from ethnically diverse faces, the specific nature of these features varies according to the culture of the observer.

3.4 Discussion

In this chapter, I aimed to extend prominent feature-based models of social trait face perception (e.g., Freeman & Ambady, 2011; Oosterhof & Todorov, 2008; Vernon et al., 2014; Zebrowitz, 2017) and my own results (see Chapter 2) by examining whether and how face ethnicity influences the facial features East Asian observers use to perceive trustworthiness and dominance. Specifically, as in Chapter 2, I combined a social perception-based psychophysical approach (see Jack & Schyns, 2017, for more details) with a high-fidelity 3D generative model of the human face (Yu et al., 2012; Zhan, Garrod, et al., 2019, see also Supplementary materials 7.1.1) to model the specific facial features of 3D shape and 2D complexion that drive perceptions of trustworthiness and dominance from Black African, East Asian, and White European faces in 60 individual East Asian observers. Comparisons of the resulting models revealed that, like White Western observers (see Chapter 2), East Asian observers perceive social traits from a combination of facial features that are shared across face ethnicities plus face ethnicity-specific variations. Further examination of these feature sets showed that the shared facial features largely align with both prominent models (e.g., Oosterhof & Todorov, 2008) and the previous results for White Western observers (see Chapter 2). In contrast, the face ethnicity-specific features diverged from previous work, including my own, and could not be explained in line with key predictions from the literature. Specifically, the face ethnicity-specific features did not modulate ethnic appearance for trustworthy judgements (Blair et al., 2002; Hutchings et al., 2024) nor did they modulate the strength of the shared features for dominant judgements (Hugenberg, 2005). I discuss the implications of these results below.

Growing evidence highlights the influence of face ethnicity on social perception (e.g., Blair et al., 2002; Eberhardt et al., 2004, 2006; Hutchings et al., 2024; Kleider-Offutt et al., 2018; Xie et al., 2021). However, as discussed in Chapter 2, current feature-based models of social trait judgements are predominantly based on White faces and are therefore limited in their ability to causally explain ethnic variance (see also Cook & Over, 2021, for further discussion). Further, despite mounting calls to move beyond WEIRD-centric examinations (see Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018), most work is based on the perceptions of White Western observers and is therefore ill equipped to capture known cultural differences in social perception (e.g., B. C. Jones et al., 2021; Maeng et al., 2022; Shimizu et al., 2017). My results underscore the limitations of this White- and Western-centric bias in two key ways. First, I show that face ethnicity causally alters the facial features East Asian observers use to perceive trustworthiness and dominance, thereby extending current prominent feature-based models (e.g., Oosterhof & Todorov, 2008), causally explaining previous

work on ethnic variance (e.g., Xie et al., 2021), and replicating my own results (see Chapter 2). Second, my results highlight the importance of considering diversity in both the face and the observer by showing that, although both East Asian and White Western observers perceive social traits from a combination of shared and face ethnicity-specific feature sets, the features that comprise these sets vary according to the culture of the observer. Together, this demonstrates the importance of shifting the White- and Western-centric focus of current research toward a more inclusive, and therefore nuanced, understanding of social trait perception.

My results also highlight a possible dearth of cross-cultural data on ethnic stereotyping. Specifically, I show that the face ethnicity-specific features East Asian observers use to perceive trustworthiness do not alter ethnic appearance, contrasting previous work including my own (e.g., Blair et al., 2002; Eberhardt et al., 2006; Hutchings et al., 2024; Kleider-Offutt et al., 2018, see also Chapter 2). However, such work has been conducted using a White viewpoint. That is, although both White and Black observers show a preference for White phenotypic appearance (e.g., Hutchings et al., 2024; Kleider-Offutt et al., 2018; see also Maddox, 2004, for a review), such work has been predominantly conducted in majority White societies or societies where Whiteness affords social power (e.g., USA). As both the target and content of ethnic stereotypes varies across national, ethnic, and cultural boundaries (Durante et al., 2017; Fiske, 2017), East Asian (or otherwise not White and Western) observers may not be as influenced by ethnic phenotypicity. Similarly, my results also show that, unlike previous work (e.g., Hugenberg, 2005), face ethnicity does not alter the intensity of the dominant-looking shared features. Again, however, such work captures the perception of White observers and may therefore not generalize to other observer groups. Notably, this is not to say that ethnic stereotypes do not exist in non-Western cultures, but rather that they may manifest and affect social perception in different ways.

Finally, as in Chapter 2, most—though not all—of the shared features mirror previous findings (e.g., V. R. Mileva et al., 2014; Oosterhof & Todorov, 2008; Said et al., 2009; Todorov et al., 2011; Vernon et al., 2014) and partly resemble (i.e., positively correlate, albeit moderately) the results for White Western observers (see Chapter 2). Specifically, across face ethnicities, both East Asian and White Western observers perceive smaller faces with bigger eyes and lighter eye regions as trustworthy (e.g., Oosterhof & Todorov, 2008; Thorstenson et al., 2018; Zebrowitz & McDonald, 1991), and longer faces with angled eyebrows and cooler skin tones as dominant (e.g., Todorov & Oosterhof, 2011; Windmann et al., 2023). These findings suggest that at least some of the facial features that drive social trait perception generalize across face ethnicities as well as observer cultures, in line with prominent evolutionary theories (Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008, see also subsections 1.2.1 and 1.2.2). Nevertheless, though similar, the facial features East Asian observer use to perceive social traits across face ethnicities are not identical to those used by White Western observers. Consistent with modern cultural perspectives on social trait perception (Cook et al., 2022; Over et al., 2020, see also subsection 1.2.6), this suggests that, in addition to face ethnicity, observer culture is also a causal source of variance in social trait

perception. To this end, Chapter 4 will extend these findings by formally examining whether and how the culture of the observer affects the facial features used to perceive trustworthiness and dominance across face ethnicities.

Although adding to growing efforts to diversify psychological science, the current work focuses on perceptions of trustworthiness and dominance—two social trait dimensions argued to be universal (Oosterhof & Todorov, 2008) but now growingly contested as Western-centric (e.g., B. C. Jones et al., 2021; Lin et al., 2021; Sutherland et al., 2018). Although this focus was an active choice to test key theories and enable direct comparisons between observer cultures, future work should refine investigations to more culturally sensitive social trait concepts (e.g., youthfulness/attractiveness; Lin et al., 2021; Sutherland et al., 2018). Additionally, this chapter only examines variance across three broad ethnic categories: Black African, East Asian, and White European. As emerging evidence suggests that different and/or more complex ethnic categories (e.g., multiracial) may differentially impact perception (e.g., Kang & Bodenhausen, 2015; Norman et al., 2024), future work should examine variance in facial feature use within and across additional face ethnicities. Finally, as throughout this thesis, these results only pertain to variance in the use of static facial features. Examining how ethnicity and culture influence social trait perception from additional sources of social information (e.g., facial expressions, voice; e.g., Gill et al., 2014; McAleer et al., 2014) is a key next step in furthering understanding of social trait perception.

Together, this chapter extends current knowledge of social trait perception by showing that face ethnicity causally alters the facial features East Asian observers use to perceive trustworthiness and dominance. In doing so, I conceptually replicate the results for White Western observers (see Chapter 2) and therefore demonstrate that face ethnicity is a central source of variance in deriving causal explanations of social trait perception across cultures. However, results also show that the specific facial features driving social trait perception within and across face ethnicities for East Asian observers differ from those used by White Western observers and do not reflect key predictions from the current literature on social trait perception and ethnic stereotyping. This suggests that, on top of face ethnicity, the culture of the observer also plays a significant role in shaping the causal facial features for social trait perception. Together, these findings highlight the importance of considering diversity in both the face and the observer when testing key theories in social perception.

Chapter 4

Cultural variance questions the universality of emotion overgeneralization

Chapter abstract

Social trait perception is largely considered to be universal and to be driven by evolutionarily adaptive cues. However, few studies have moved beyond correlational approaches to directly compare the causal facial features that drive social trait judgements across ethnic and cultural boundaries. This lack of causal findings questions previous claims of universality and, more broadly, the generalizability of current knowledge. In this chapter, I address this gap by formally comparing the 3D face models that capture perceptions of trustworthiness and dominance from ethnically diverse faces in East Asian and White Western observers. While social trait ratings were positively correlated across cultures, the facial features guiding those ratings differed. Specifically, information theoretic analyses revealed that perceptions of trustworthiness and dominance are driven by culturally shared plus culture-specific facial features across face ethnicities. Comparisons of the culture-specific features revealed that while White Western observers use features that resemble emotion cues (e.g., smiling, frowning) to socially perceive ethnically diverse faces in line with prominent theories, East Asian observers do not. These findings challenge previous claims of universality, question the generalizability of current theories, and highlight the need to move beyond rating-level comparisons toward causal models that can capture cultural variance.

4.1 Introduction

Prominent theories of social trait perception argue, at least in part, for an evolutionary basis of social judgements. Specifically, the trustworthiness-dominance model (Oosterhof & Todorov, 2008, see also subsection 1.2.1) argues that the mapping of all social trait judgements along its two primary dimensions reflects the evolutionary need to distinguish friend from foe. In the same way, the overgeneralization hypothesis (Zebrowitz, 2017; Zebrowitz & Montepare, 2008, see also subsection 1.2.2) posits that social trait judgements are driven by the resemblance of neutral faces to evolutionarily adaptive cues. Among these, resemblance to facial expressions of emotion has received arguably the most attention, with a large body of both behavioural (e.g., Adams et al., 2012; Montepare & Dobish, 2003; Windmann et al., 2023) and machine learning-based evidence (e.g., Albohn & Adams, 2021b; Said et al., 2009; Zebrowitz et al., 2010) highlighting its centrality in social trait perception. Specifically, resemblance to positive emotions (e.g., happy) elicits judgements of similarly positively valenced traits, such as trustworthiness (e.g., Albohn & Adams, 2021b; Said et al., 2009); and resemblance to negative emotions (e.g., anger) elicits judgements of negatively valenced traits such as dominance (e.g., Adams et al., 2022; Windmann et al., 2023).

If social trait judgements have an evolutionary origin, it stands to reason that social trait perception should be largely universal. Although most work on social trait perception remains centred on White Western observers perceiving White faces (e.g., Cook & Over, 2021; see also Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018), work that integrates diversity tends to support universality by reporting high cross-cultural similarities (e.g., Albright et al., 1997; Sutherland et al., 2018; Walker et al., 2011; A. Wang et al., 2024; Zebrowitz et al., 2012). However, recent evidence challenges these claims by showing that such similarities could be inflated by the rating scale (Maeng et al., 2022) or analyses used (B. C. Jones et al., 2021). Similarly, the inclusion of expressive faces in otherwise neutral face stimulus sets as in Sutherland et al. (2018) could mask differences. Finally—with the key exception of Sutherland et al. (2018)—most work supporting universality focuses on similarities in ratings rather than the facial features that elicit them. This is problematic because, in line with emerging evidence (e.g., Mo et al., 2022; Rostovtseva et al., 2024), culturally diverse observers could judge the same face as, for example, "trustworthy" but using different facial features (Medin et al., 1993; Schyns et al., 2022). Achieving a causal understanding of social trait perception is therefore critical to substantiating claims of universality.

Revealing the causal facial features diverse observers use for social trait perception from ethnically diverse faces is also central to testing the generalizability of key feature-based theories, such as emotion overgeneralization (e.g., Zebrowitz, 2017; Zebrowitz & Montepare, 2008). That is, if social trait judgements are driven by the resemblance of neutral faces to evolutionarily adaptive emotion cues, these features should be diagnostic of trustworthiness and dominance regardless of the ethnicity of the face and the culture of the observer. Critically, emotion overgeneralization has seldom been studied outside of a White and Western context. This is problematic for two separate but closely linked reasons. First, ethnic vari-

ance in the face influences how emotions are perceived (e.g., Hugenberg, 2005; Matsumoto, 1993) and subsequently integrated in social trait judgements (e.g., Adams et al., 2022; Zebrowitz et al., 2010; but see Jaeger & Jones, 2022, for an ethnicity-general account). Second, cultural variance in the observer influences which facial expressions are associated with different emotions (e.g., Elfenbein & Ambady, 2002; Jack, Garrod, et al., 2012; Jack et al., 2009) and how different emotions are subsequently evaluated (e.g., Park et al., 2016; Tsai, 2017; Tsai et al., 2016; Uchida & Kitayama, 2009). Together, this raises questions about whether and how observers from different cultural backgrounds use emotion cues to perceive social traits from ethnically diverse faces.

Thus, without a causal understanding of which facial features drive social trait judgements across diverse observers, claims of universality can neither be fully substantiated nor discarded. Additionally, it remains unknown whether and how adaptive emotion cues theorized to universally guide social trait judgements are used for social trait perception across ethnic and cultural boundaries. In this chapter, I address these gaps by formally comparing the previously derived 3D models of the facial features driving perceptions of trustworthiness and dominance from ethnically diverse faces in White Western (see Chapter 2) and East Asian observers (see Chapter 3). Specifically, I aim to examine whether the facial features used for social trait perception from ethnically diverse faces vary according to the culture of the observer, and whether and how they map onto emotion cues. This chapter therefore formally extends the initial cross-cultural comparisons presented in Chapter 3.

4.2 Methods

In this chapter, I formally compare the 3D face models of White Western and East Asian observers obtained in Chapters 2 and 3 respectively. As no additional data was collected to conduct the following analyses, I do not repeat here the methodological details used to obtain the models. The full methodological details are reported in the methods sections of Chapters 2 and 3.

4.3 Results

Despite known cultural differences in social trait perception (e.g., B. C. Jones et al., 2021; Maeng et al., 2022), evidence shows that judgements of trustworthiness and dominance are highly consistent across observer cultures (e.g., Albright et al., 1997; Walker et al., 2011; A. Wang et al., 2024; Zebrowitz et al., 2012). Here, I replicate these findings by correlating (Pearson r) the average rating (mean across 20 observers per face ethnicity and observer culture) given by East Asian and White Western observers to each of the 2,400 face stimuli in each face ethnicity condition ($[1,200 \text{ face stimuli} \times 2 \text{ stimulus sex}]$). Figure 4.1 shows the results as colour-coded scatter plots (blue = male faces, red = female faces), separately for each social trait rating task and face ethnicity.

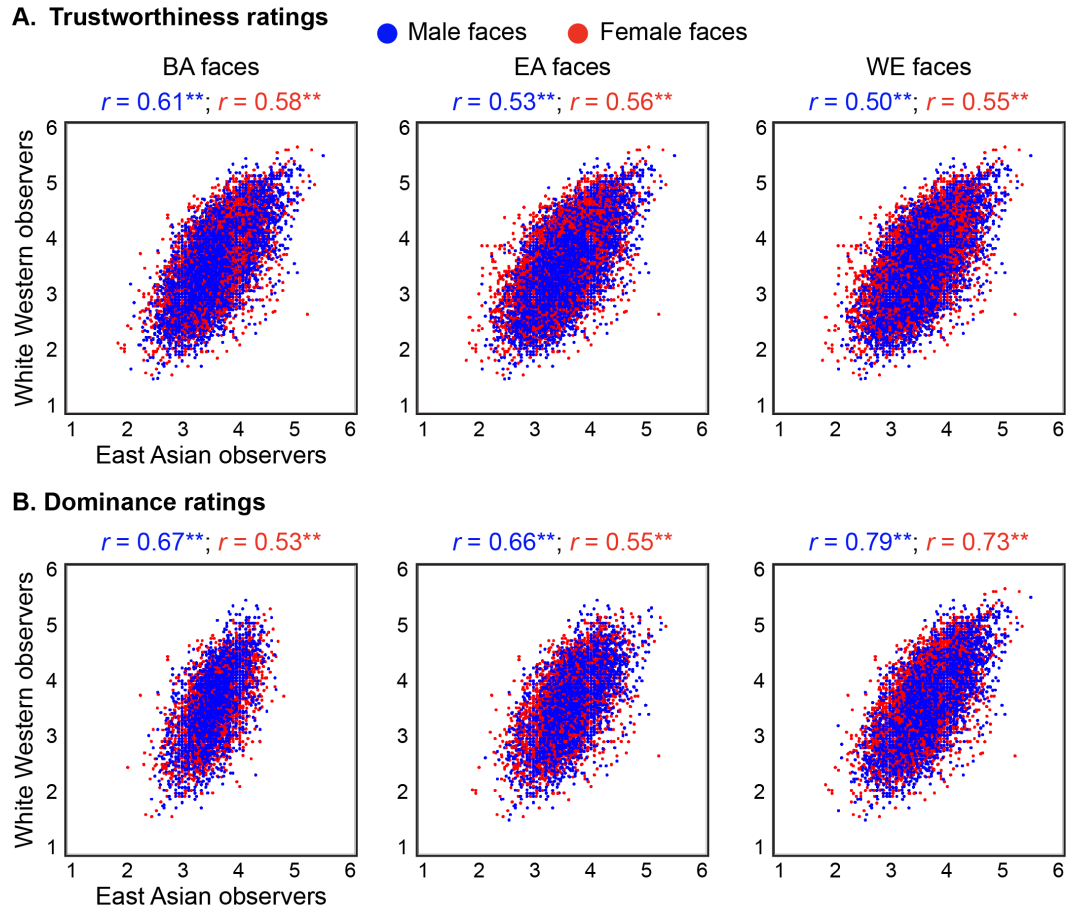


Figure 4.1: In each panel (A: Trustworthiness ratings; B: Dominance ratings), colour-coded dots in each subplot show the average rating given by East Asian observers (x axis) and White Western observers (y axis) for each of 1,200 male (blue colour-coded) and female face stimuli (red colour-coded), separately for each face ethnicity (see labels at top). Above each plot, colour-coded values report the Pearson correlation between the average ratings of the two observers groups for male and female faces (** $p < 0.01$). Results show that ratings of trustworthiness and dominance positively correlate between East Asian and White Western observers for each face ethnicity and stimulus sex.

Results show that ratings of dominance and trustworthiness are statistically significantly ($p < 0.01$) positively correlated between observer cultures for each face ethnicity. Specifically, trustworthiness ratings are positively correlated between White Western and East Asian observers for Black African (male: $r(1,198) = 0.61$; female: $r(1,198) = 0.58$), East Asian (male: $r(1,198) = 0.53$; female: $r(1,198) = 0.56$), and White European faces (male: $r(1,198) = 0.50$; female: $r(1,198) = 0.55$). Dominance ratings produce similar results (BA—male: $r(1,198) = 0.67$; female: $r(1,198) = 0.53$; EA—male: $r(1,198) = 0.66$; female: $r(1,198) = 0.55$; WE—male: $r(1,198) = 0.79$; female: $r(1,198) = 0.73$). Therefore, social trait judgements made by East Asian and White Western observers are consistent at the rating level—that is, faces judged as more trustworthy/dominant by East Asian observers tend to also be judged as more trustworthy/dominant by White Western observers.

Critically, this rating-level similarity does not necessarily mean that diverse observers use the same facial features to arrive at the same perception (Schyns et al., 2022; see also Medin et al., 1993; Newell, 1998). Rather, growing evidence suggests that observers from different

cultural backgrounds use distinct features to perceive social traits from ethnically diverse faces (e.g., Mo et al., 2022; Rostovtseva et al., 2024). To formally test whether the facial features that drive social trait perception from ethnically diverse faces vary as a function of the culture of the observer, I compared the previously obtained 3D models of the facial features that drive perceptions of trustworthiness and dominance from Black African (BA), East Asian (EA) and White European (WE) faces in 60 individual East Asian observers (see Chapter 3) and 60 individual White Western observers (see Chapter 2). Figure 4.2 shows the models for trustworthy-looking and dominant-looking male faces, with results aggregated across individual observers ($n = 20$ per face ethnicity condition in each observer culture; see Figure 4.3 for female faces).

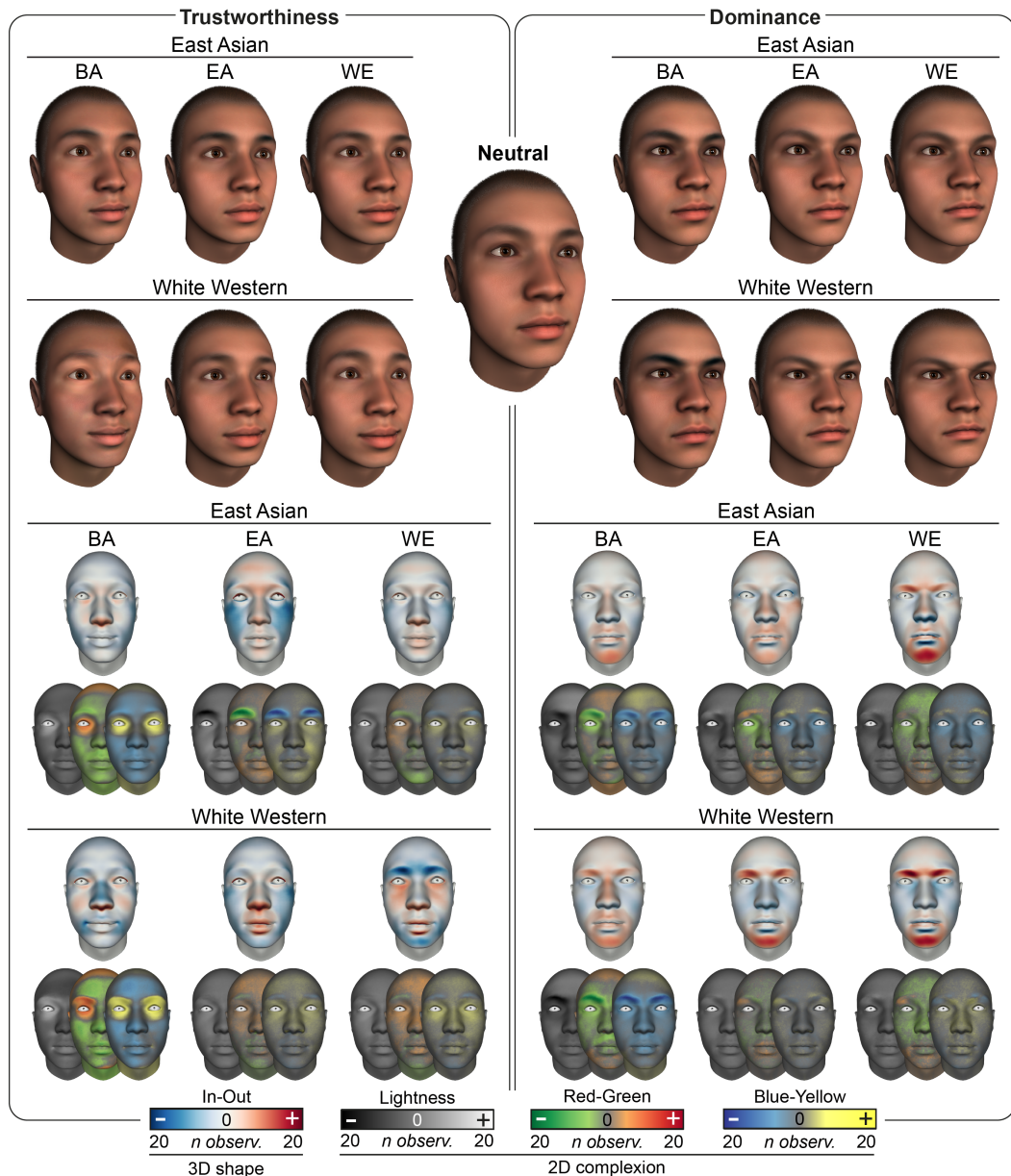


Figure 4.2: Faces at the top show the 3D models of male faces perceived as trustworthy (left) and dominant (right) by East Asian observers (top row) and White Western observers (bottom row) for each face ethnicity—Black African (BA), East Asian (EA) and White European (WE)—displayed on the same ethnically neutral face (see larger face in the centre) for comparison. Results are aggregated across observers ($n = 20$ per face ethnicity condition in each observer culture) and include all identity components that meet or exceed the population prevalence threshold ($n = 4$ observers; faces extrapolated for clarity). Below, colour-coded faces show the results separately for 3D shape and 2D complexion. Colour-coding denotes the feature deviation direction from the average ethnically neutral face; colour saturation shows the number of observers with a statistically significant effect (see colorbars below).

In the top panel, each face shows the resulting 3D models of male faces perceived as trustworthy (left) and dominant (right) by East Asian observers (top row) and White Western observers (bottom row) for each face ethnicity, with shape and complexion combined and displayed on the same ethnically neutral face shown in the centre for comparison. Colour-coded faces below show the results for 3D shape and 2D complexion separately, using the same format as in previous chapters (see colorbars below). Colour saturation shows the

number of observers ($n = 20$ per face ethnicity condition in each observer culture) with a statistically significant effect, for all identity components at or above the population prevalence threshold ($n = 4$ observers; Donhauser et al., 2018; Ince et al., 2021).

Focusing on comparisons between observer cultures, visual inspection of the results reveals clear cross-cultural similarities across face ethnicities. Specifically, regardless of the ethnicity of the face, both East Asian and White Western observers perceive smaller faces with softer features as trustworthy, and larger faces with more angular features as dominant. The modelled complexion features also bear cross-cultural similarities, though these effects vary according to the ethnicity of the face. Specifically, in both cultures, BA faces perceived as trustworthy tend to have lighter, warmer eye regions against darker, cooler skin tones; while both EA and WE faces perceived as trustworthy tend to have cooler eyebrows against warmer skin tones. Similarly, in both cultures, BA faces perceived as dominant tend to have darker, cooler eyebrows and eye regions together with lighter, warmer chins and foreheads; while both EA and WE faces then to have sallower skin tones. However, in addition to these similarities, the results also show cross-cultural differences, particularly in the use of emotion-resembling shape features. Specifically, across face ethnicities, fewer East Asian than White Western observers perceive trustworthiness from happy-resembling features such as arched eyebrows and upturned mouth corners (e.g., Albohn & Adams, 2021b; Said et al., 2009), and dominance from anger-resembling features such as a lower brow bone with angled eyebrows (e.g., Adams et al., 2012; Windmann et al., 2023).

Results for female faces (see Figure 4.3) show a similar pattern of cross-cultural similarities and differences across face ethnicities. For example, across face ethnicities, both East Asian and White Western observers perceive smaller vs larger faces as trustworthy and dominant respectively. Additionally, the upturning of mouth corners and the protrusion of the the brow-bone are respectively perceived as trustworthy and dominant by more White Western than East Asian observers.

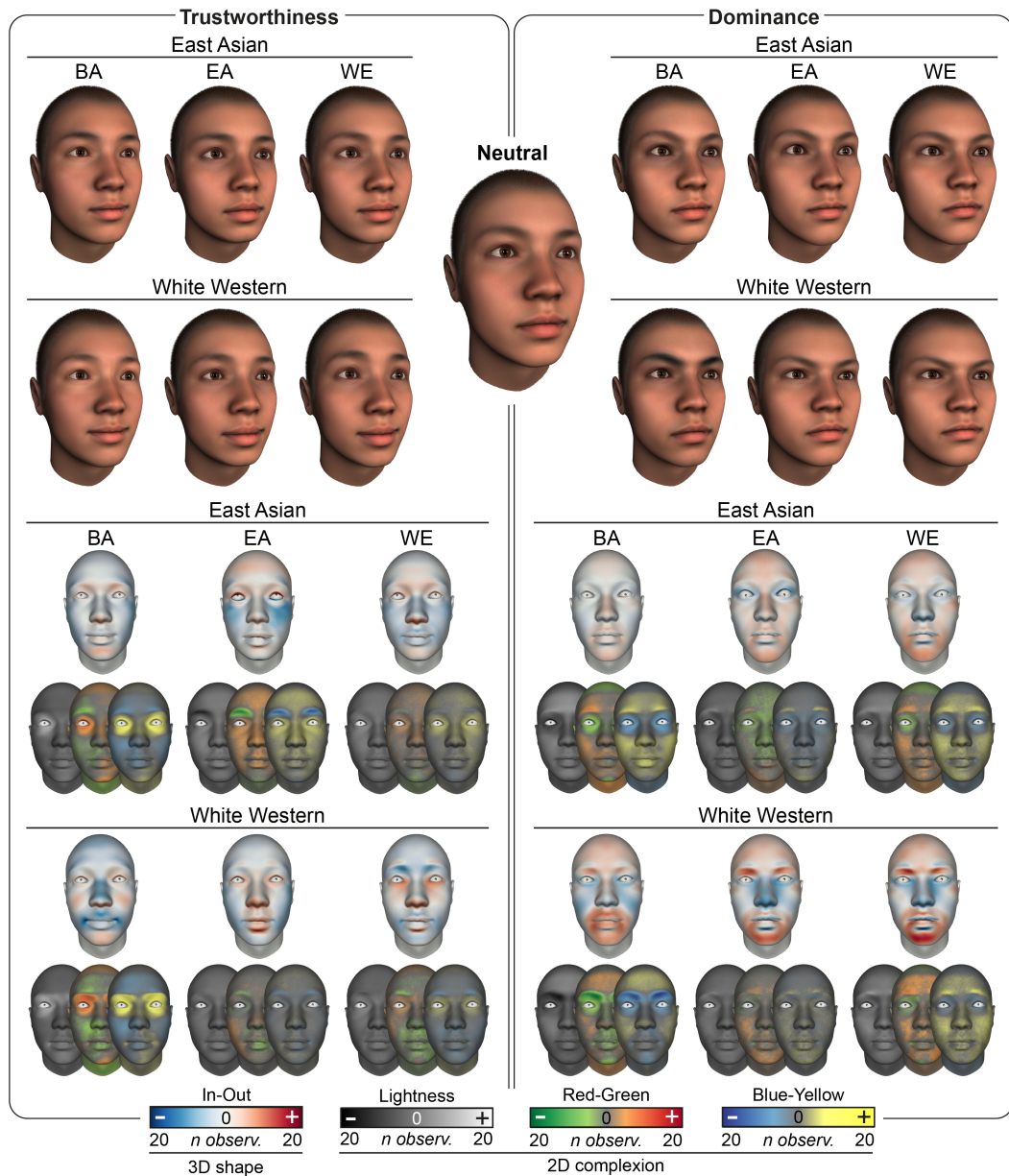


Figure 4.3: Following the same format as Figure 4.2, faces at the top show the 3D models of female faces perceived as trustworthy (left) and dominant (right) by East Asian observers (top row) and White Western observers (bottom row) for each face ethnicity, displayed on the same ethnically neutral face (see larger face in the centre) for comparison. Below, colour-coded faces show the results separately for 3D shape and 2D complexion (see colorbars below). Results are aggregated across observers ($n = 20$ per face ethnicity in each observer culture), with all identity components meeting or exceeding population prevalence threshold of $n = 4$.

4.3.1 Social trait perception is driven by culturally shared plus culture-specific facial features

Together, these results suggest that facial features used for social trait perception may depend not only on the ethnicity of the face—as tested in Chapters 2 and 3—but also on the culture of the observer. To test this formally, I measured the specificity of each modelled identity component to observer culture using Mutual Information (MI; Cover & Thomas,

1991; Ince et al., 2017). To do this, I first pooled the models of all 120 observers ([20 observers $\times$ 3 face ethnicities $\times$ 2 observer cultures]) for each social trait and stimulus sex separately. I then computed MI between each modelled identity component at or above population prevalence threshold ($n = 4$ observers) for at least one face ethnicity and observer culture and each observer culture label in turn (e.g., "East Asian observer" vs "not East Asian observer"). I established statistical significance using non-parametric permutation testing ($N = 1,000$ permutations) and used Pearson correlation to discard any identity components with a high MI value but negative correlation with observer culture (i.e., the identity component has a high MI with observer culture because it is repeatedly absent in these 3D face models). Finally, I discarded any identity components that had high MI with observer culture but were not above population prevalence threshold in all three face ethnicities for that observer culture. Parallel to this, I only retained those identity components that had low MI with observer culture if they were also above population prevalence threshold in all three ethnicities in both cultures. In other words, I only considered those identity components that, regardless of the ethnicity of the face, are either culture-specific or shared across cultures. I repeated this analysis for each identity component of 3D shape and 2D complexion, for positive and negative identity component weights, and for each social trait and stimulus sex separately. Figure 4.4 shows the results for male faces (see Figure 4.5 for female faces).

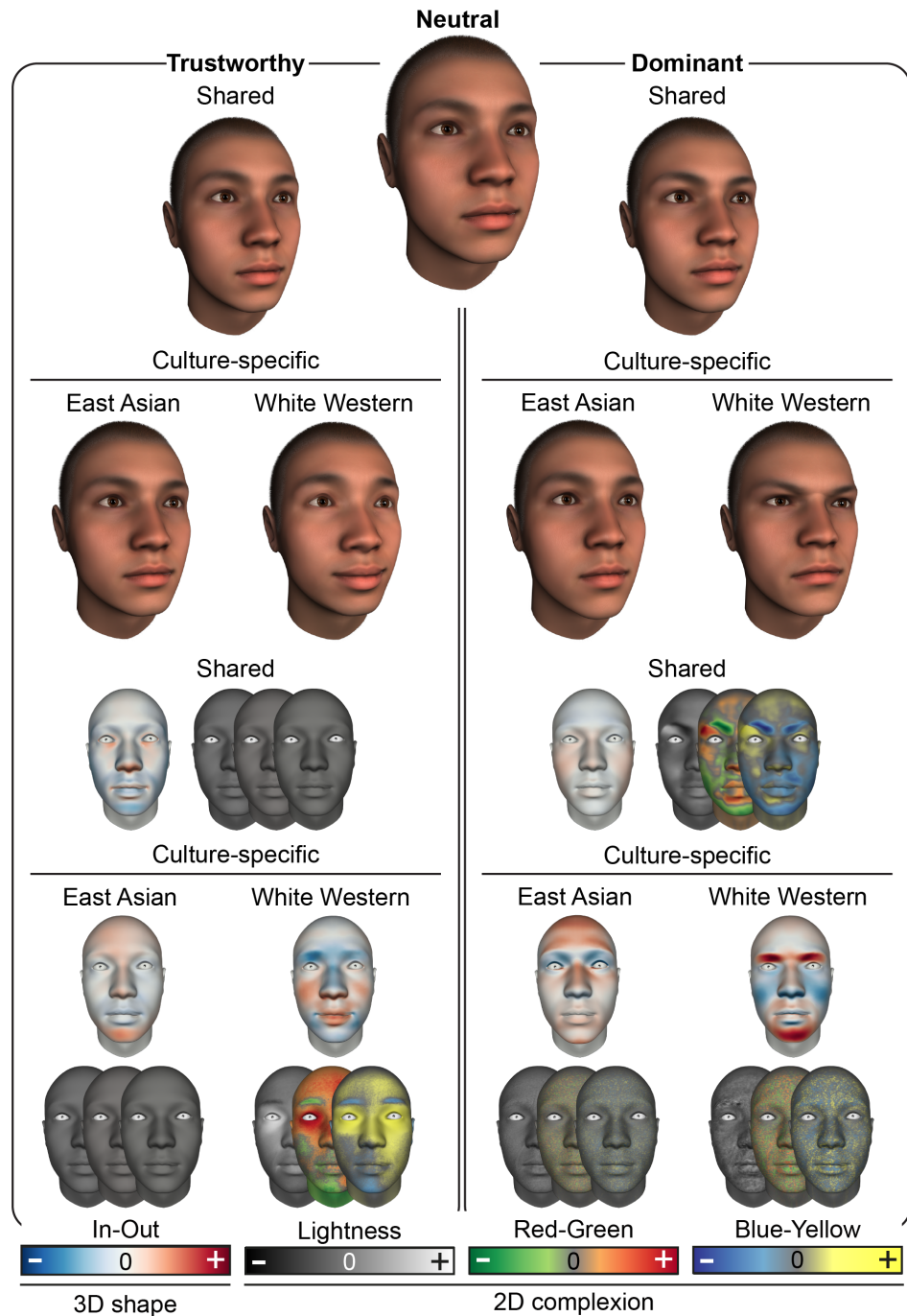


Figure 4.4: Faces at the top of each panel (left: trustworthy; right: dominant) show the facial features that are shared (top row) or specific (bottom row) to each observer culture (East Asian, White Western) across face ethnicities, for male faces only. Results are displayed on the same ethnically neutral face (see larger face at the top) for comparisons. Colour-coded faces at the bottom show the results separately for 3D shape and 2D complexion (see color-bars below), with values normalized per social trait for display purposes (see Supplementary Materials 7.4.1 for further details).

Faces at the top of each panel (left: trustworthy judgements; right: dominant judgements) show the facial features that are shared across observer cultures (top) and those that are culture-specific across face ethnicities (bottom), with shape and complexion combined and displayed on an ethnically neutral face for comparison (see larger face at the top). Colour-coded faces at the bottom show the effects separately for 3D shape and 2D complexion using

the same colour-coding format as before (see colorbars below), and with values normalized for each social trait for display purposes (see Supplementary Materials 7.4.1 for further details on the number and distribution of the shared and culture-specific features).

Results show that, across face ethnicities, perceptions of trustworthiness and dominance are guided by a combination of facial features that are shared across cultures plus those that are specific to each culture. Specifically, for both social traits, the shared facial features capture relatively subtle variations, including more protruding eyes with a narrower jaw and a smaller chin with no complexion effects for trustworthiness; and wider cheekbones with darker, cooler eyebrows and skin tones for dominance. Although subtle, these facial features map onto previous findings, including the role of baby-like, softer features in perceptions of trustworthiness (e.g., Oosterhof & Todorov, 2008; Zebrowitz et al., 2003) and of facial width and darker skin tones in perceptions of dominance (e.g., Albert et al., 2021; V. R. Mileva et al., 2014; Todorov & Oosterhof, 2011). Additionally, the Western culture-specific facial features also largely map onto previous findings. For trustworthy judgements, these features comprise a higher brow bone with arched eyebrows, fuller cheeks, fuller lips with upturned mouth corners and a softer jaw with darker, cooler eyebrows against lighter, warmer skin tones (e.g., Oosterhof & Todorov, 2008; Said et al., 2009; Todorov & Oosterhof, 2011; Vernon et al., 2014; Zebrowitz & McDonald, 1991). For dominant judgements, these features comprise a lower and more protruding brow bone with angled eyebrows, a narrower face, thinner lips and a stronger chin with darker, cooler skin tones (e.g., Adams et al., 2012; V. R. Mileva et al., 2014; Oosterhof & Todorov, 2008; Todorov & Oosterhof, 2011; Zebrowitz et al., 2003). In contrast, the features that are specific to East Asian culture are mostly unrepresented in current models of social trait perception and comprise a wider forehead, a higher nose bridge, and a larger chin with no complexion effects for trustworthy judgements; and a longer face with higher-set arched eyebrows, deeper-set eyes, and a higher nose bridge with slightly cooler skin tones with dominant judgments.

Results for female faces (see Figure 4.5) are similar, but not identical to, those for male faces. For trustworthy judgements, both the shared and culture-specific facial features are similar to those for male faces but include no complexion effects for either culture. For dominant judgments, the shared features further comprise a more protruding brow bone and a narrower jaw; the East Asian culture-specific features are similar to those for male faces; and the White Western culture-specific features further comprise a smaller nose and a wider face with a lighter, warmer mouth area and eyebrows against darker, cooler skin tones.

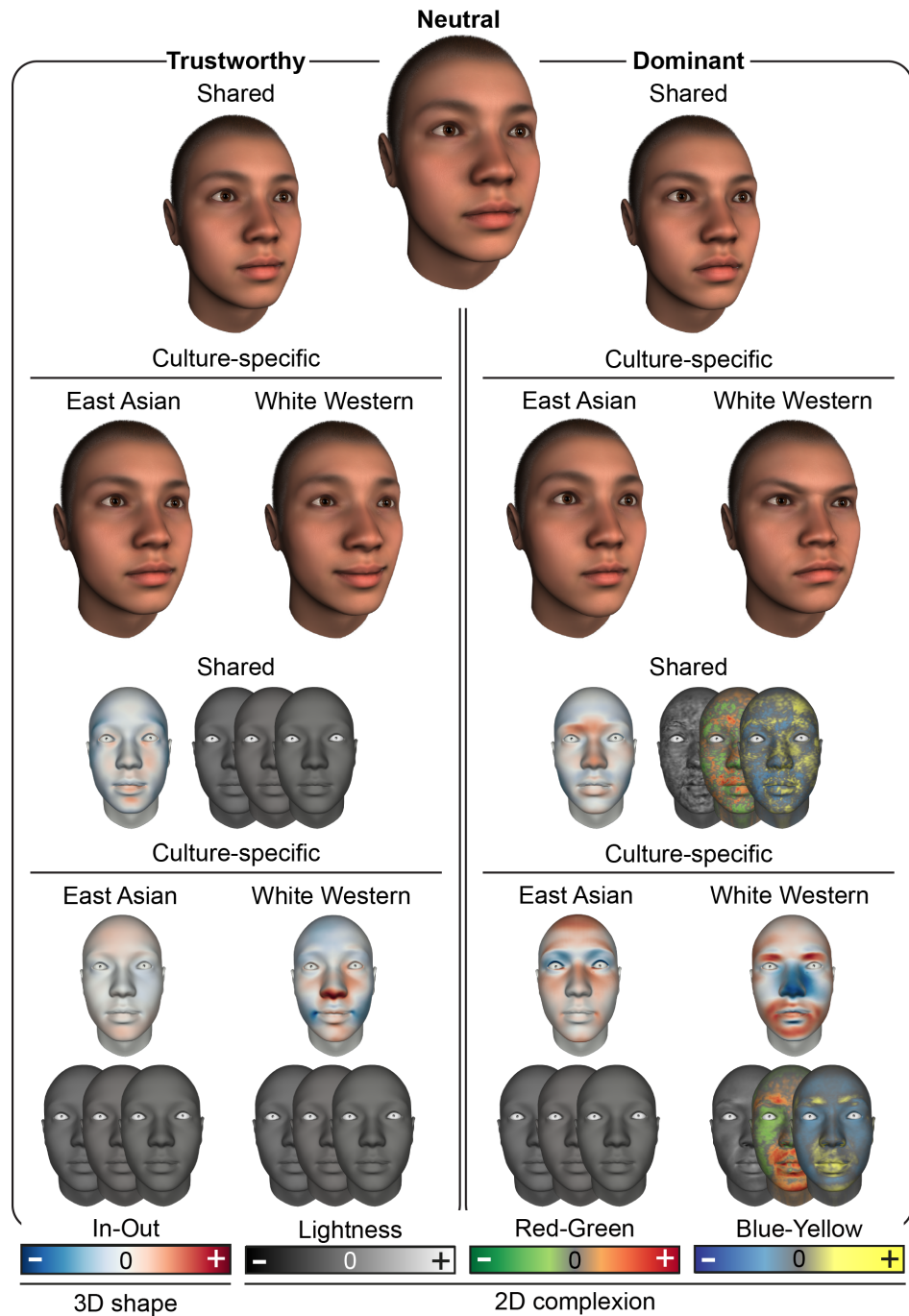


Figure 4.5: Following the same format as Figure 4.4, faces at the top show the trustworthy-looking (left) and dominant-looking (right) facial features of female faces that are shared (top row) or specific (bottom row) to each observer culture (East Asian, White Western) across face ethnicities. Colour-coded faces at the bottom show the results separately for 3D shape and 2D complexion (see colorbars below), with values normalized per social trait for display purposes (see Supplementary Materials 7.4.1 for further details).

Together, these results indicate that the facial features driving social trait perception from ethnically diverse faces vary as a function of the culture of the observer. Further, while both the shared and White Western culture-specific features largely map onto previous findings (e.g., Jaeger & Jones, 2022; V. R. Mileva et al., 2014; Oosterhof & Todorov, 2008; Said et al., 2009; Todorov et al., 2011), the East Asian culture-specific features do not.

4.3.2 The use of emotion cues in social trait perception is specific to Western culture

The social trait facial features that are specific to White Western culture seem to largely resemble emotion cues—including happiness-resembling features such as upturned mouth corners for trustworthy judgements and anger-resembling features such as a lower brow bone for dominant judgements. In contrast, the facial features that are specific to East Asian culture do not appear to. This stands in contrast to extensive theoretical and empirical work pointing to the primacy—and often implying the universality—of emotion cues in social trait perception (e.g., Albohn & Adams, [2021b](#); Jaeger & Jones, [2022](#); Montepare & Dobish, [2003](#); Oosterhof & Todorov, [2008](#); Said et al., [2009](#); Sutherland et al., [2018](#); Windmann et al., [2023](#); Zebrowitz & Montepare, [2008](#)).

To formally test whether the culture-specific features that East Asian and White Western observers use to perceive social traits across face ethnicities map onto emotion cues, I used a bespoke approach that measures the similarity of static facial features to dynamic facial movements—taxonomized as Action Units (AUs) in the Facial Action Coding System (Ekman & Friesen, [1978](#)). Due to the nature of the comparison, I did this only for the culture-specific features of 3D shape. Figure [4.6](#) illustrates the approach.

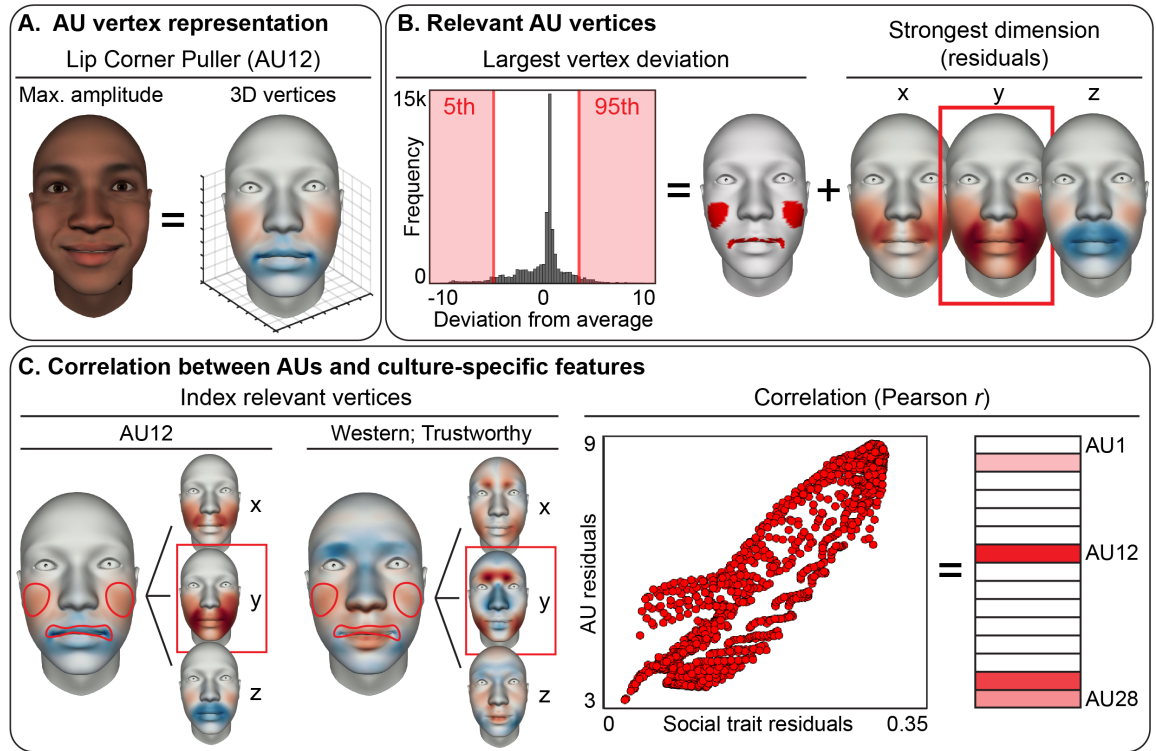


Figure 4.6: **A. AU vertex representation.** Each AU (set at maximum amplitude) is represented as 3D face vertices that deviate from the average ethnically neutral face. **B. Relevant AU vertices.** Vertex deviations exceeding the 5th or 95th percentile (or 10th and 90th percentile for small magnitude AUs, e.g., Upper Lid Raiser; see Figure 7.14 for details) are selected as most relevant. Of these vertices, the strongest cartesian dimension is identified using the highest absolute effect size of each dimension (i.e., mean of residuals from the average ethnically neutral face). **C. Correlation between AUs and culture-specific features.** The vertex representations of each AU and set of culture-specific social trait features are indexed using the relevant vertices and cartesian dimension identified in B. The resulting residuals are correlated (Pearson r), yielding a 28 (AUs) $\times$ 1 vector detailing the AU pattern that best represents the culture-specific social trait features.

As shown in Figure 4.6A, I first represented each individual AU (set at maximum movement amplitude) as 3D coordinates for 31,094 face vertices ($N = 14,319$ front-face vertices), and computed their deviation from the average ethnically neutral face. Next, as shown in Figure 4.6B, I identified the 3D front-face vertices that best capture each AU using either the 10th and 90th, or the 5th and 95th percentile of the distribution of its vertex deviations depending on the magnitude of the facial movement (e.g., Upper Lid Raiser—AU5 vs Lip Corner Puller—AU12). In other words, I only considered those vertices that make up the tail ends of the distribution (i.e., largest positive/negative deviations). Out of these vertices, I then identified the cartesian dimension (x, y, z axis) with the largest absolute effect size using the mean of the residuals from the average ethnically neutral face for each dimension. I did this to ensure that subsequent comparisons consider both the location of the AU vertices and the specific type of deviation (e.g., longer vs wider). To limit comparisons, I only retained bilateral versions (e.g., Lip Corner Puller) of lateralized AUs (e.g., Lip Corner Puller Left, Lip Corner Puller Right), for a total of 28 AUs (see Supplementary materials 7.4.2 for further details).

Having identified the 3D face vertices that capture each individual AU, I compared them to those that capture the culture-specific social trait facial features. To do this, I first recovered the weights of the culture-specific identity components for each individual observers ($n = 120$ [20 observers $\times$ 3 face ethnicity $\times$ 2 observer culture]), thereby obtaining an observer-level representation of each set of culture-specific social trait features. I then represented these results as 3D face vertices, computed their deviation from the average ethnically neutral face, and decomposed their deviation according to each cartesian dimension (i.e., residuals). Next, as shown in Figure 4.6C, I used the relevant AU vertices and strongest cartesian dimension of each AU to index each observer-level set of culture-specific facial features. Finally, I correlated (Pearson r) the indexed residuals between each AU and set of culture-specific features, separately for each observer, social trait and stimulus sex. This yielded a $28 \text{ (AU)} \times 1$ vector of Pearson correlation values reflecting the AUs that most resemble each set of culture-specific features for each observer. To establish statistical significance, I generated a chance distribution of Pearson correlation values by repeating the same procedure for each AU using 500 randomly generated, ethnically neutral faces. I then took the 95th percentile of the distribution as threshold for statistical significance. Finally, I only retained AUs that were above population prevalence threshold ($n = 4$ observers) in each face ethnicity condition within each observer culture (see Figure 7.15 for resulting AU patterns).

Having obtained the AU patterns that best represent the culture-specific facial features, I compared them (Pearson r) to the AU patterns that are associated with the six basic emotions—happiness, surprise, fear, disgust, anger, and sadness. As mounting evidence shows that the facial expressions associated with these emotions vary across cultures (e.g., Elfenbein & Ambady, 2002; Fernández-Dols, 2013; Jack, 2013), I used culture-specific models of these facial expressions previously derived by Jack, Garrod, et al. (2012) and C. Chen et al. (2024) in Western and East Asian culture respectively. Figure 4.7 shows the results for male faces (see Figure 4.8 for female faces).

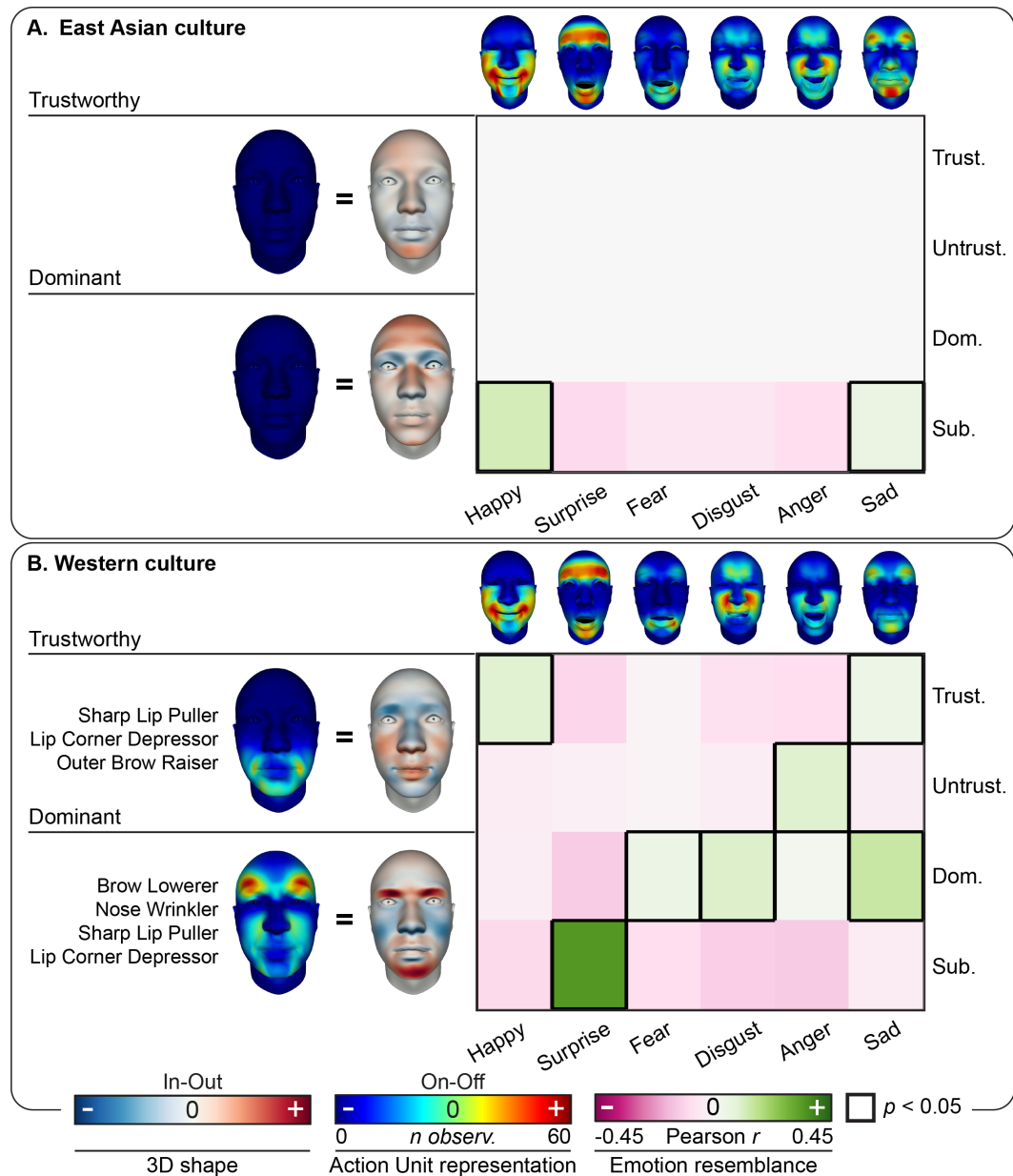


Figure 4.7: In each panel (A: East Asian culture; B: Western culture), colour-coded faces on the left show the Action Unit (AU) representation of the trustworthy-looking (top) and dominant-looking (bottom) culture-specific facial features. Colour-coding indicates the facial feature deviations as before, and the prevalence of the Action Units across observers (see colorbars below). Labels on the left report the most prevalent AUs across observers ($n = 60$; see Figure 7.15 for full AU patterns). Colour-coded matrices on the right show the Pearson correlation between the AU patterns of the culture-specific social trait features (y axis) and the six basic emotions in each culture (x axis; shown as colour-coded faces above each matrix). Matrix colour-coding indicates the correlation direction and magnitude (see colorbars below), and thick black outlines indicate statistically significant ($p < .05$) positive correlations.

In each panel (A: East Asian culture; B: Western culture), colour-coded faces on the left show the Action Unit (AU) representation of the culture-specific facial features associated with trustworthy (top) and dominant judgements (bottom). Colour-coding for the social trait facial features follows the same colour-coding format as before; colour-coding for their AU representations indicates the prevalence of the effects across observers ($n = 60$ observers

[20 observers $\times$ 3 face ethnicity]; see colorbars below). Labels on the left report the most prevalent AUs across observers for each social trait (ordered highest to lowest; see Figure 7.15 for full AU patterns). On the right, colour-coded matrices show the Pearson correlation between the AU patterns of the culture-specific social trait facial features (x axis) and the six basic emotions in each culture (y axis), also shown as colour-coded faces above each matrix. Colour-coding in the matrix indicates the correlation direction (green = positive, pink = negative), and colour saturation indicates the strength of the effect (see colorbar below). Thick black outlines indicate statistically significant ($p < .05$) positive correlations.

Results for male faces show that the Action Unit patterns that best represent each set of culture-specific facial features vary between cultures (see Supplementary materials 7.4.3 for full AU patterns). Further, these AU patterns statistically significantly ($p < .05$) positively correlate—albeit weakly—with different emotions across cultures. For East Asian culture (see Figure 4.7A), neither the trustworthy-looking nor the dominant-looking facial features resemble any of the 28 bilateral AUs, and therefore do not correlate with any basic emotions. Additionally, the untrustworthy-looking facial features also do not resemble any AUs; while the AU pattern of submissive-looking facial features resemble Sharp Lip Puller (AU13) and Lip Corner Depressor (AU15) and thus positively correlates with both happiness, $r = 0.12$, and sadness, $r = 0.06$ (see Figure 7.15 for full AU patterns). These findings stand in contrast with previous work on emotion overgeneralization, which reports that the facial features driving trustworthy and dominant judgements resemble similarly valenced emotion cues (e.g., Adams et al., 2012; Jaeger & Jones, 2022; Montepare & Dobish, 2003; Said et al., 2009).

For Western culture (see Figure 4.7B), the trustworthy-looking facial features resemble Sharp Lip Puller (AU13), Lip Corner Depressor (AU15), and Outer Brow Raiser (AU2); this AU pattern positively correlates with happiness, $r = 0.10$, and sadness, $r = 0.05$. Additionally, the dominant-looking facial features resemble Brow Lowerer (AU4), Nose Wrinkler (AU9), Sharp Lip Puller (AU13) and Lip Corner Depressor (AU15); this AU pattern positively correlates with fear, $r = 0.05$, disgust, $r = 0.10$, and sadness, $r = 0.15$. Additionally, the AU pattern of untrustworthy-looking facial features positively correlates with anger, $r = 0.10$; and the AU pattern of submissive-looking facial features positively correlates with surprise, $r = 0.35$ (see Figure 7.15 for full AU patterns). Thus, unlike the results for East Asian culture, the results for Western culture more closely align with previous work on emotion overgeneralization by showing that the facial features used for social traits perception by White Western observers map onto specific emotion cues which mostly mirror the valence of the social trait. Specifically, trustworthy-looking and dominant-looking facial features respectively map onto emotion cues of positive (e.g., happy; e.g., Jaeger & Jones, 2022; Said et al., 2009; Zebrowitz & Montepare, 2008) and negative emotions (e.g., disgust; e.g., Hensel et al., 2021; Montepare & Dobish, 2003; Ueda et al., 2017). However, in contrast to previous work (e.g., Adams et al., 2012), the trustworthy-looking features also correlate with sadness; and the dominant-looking features also correlate with sadness and fear.

Results for female faces (see Figure 4.8) follow a similar pattern to those for male faces,

with some key differences. For East Asian culture, the submissive-looking facial features only resemble Sharp Lip Puller (AU13) and thus only correlate with happiness, $r = 0.20$, but not sadness. Additionally, the dominant-looking facial features resemble brow-raising AUs, which positively correlate with surprise, $r = 0.25$. Notably, despite mapping onto emotion cues, their correlation with surprise contrasts previous findings (e.g., Said et al., 2009). For Western culture, the trustworthy-looking facial features positively correlate with happiness, $r = 0.12$, but not sadness. Additionally, the untrustworthy-looking facial features do not resemble any AUs and thus do not correlate with any basic emotions. Finally, the dominant-looking facial features also positively correlate with fear, $r = 0.18$, and sadness, $r = 0.21$, but not disgust.

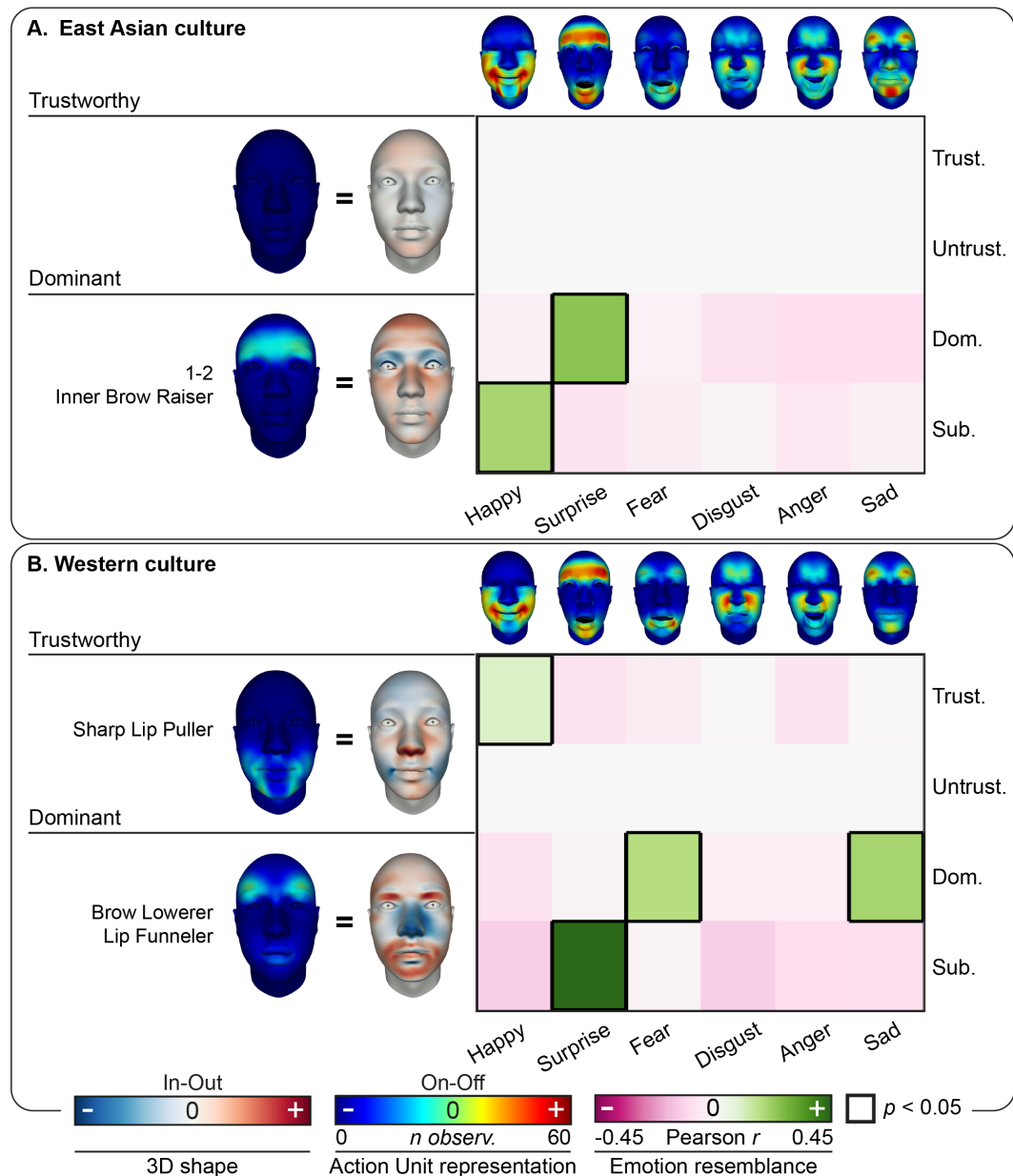


Figure 4.8: Following the same format as Figure 4.7, colour-coded faces on the left show the AU representation of the culture-specific facial features (A: East Asian culture; B: Western culture) of trustworthy-looking (top) and dominant-looking (bottom) female faces (see colorbars below). Labels on the left report the most prevalent AUs across observers ($n = 60$; see Figure 7.15 for full AU patterns). Colour-coded matrices on the right show the Pearson correlation (green = positive correlation; pink = negative correlation; see colorbar below) between the AU patterns of the culture-specific social trait features (y axis) and the six basic emotions in each culture (x axis; shown as colour-coded faces above). Thick black outlines indicate statistically significant ($p < .05$) positive correlations.

Together, these results show that the use of emotion cues in social trait perception varies across cultures. Specifically, while White Western observers use facial features that resemble emotion cues to perceive trustworthiness and dominance across face ethnicities in ways that largely align with previous work on emotion overgeneralization (e.g., Jaeger & Jones, 2022; Montepare & Dobish, 2003; Said et al., 2009; Zebrowitz & Montepare, 2008), East Asian observers do not.

4.4 Discussion

In this chapter, I aimed to test the universality of the facial features driving social trait perception implied by prominent theories in the field (e.g., Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008). Specifically, I formally compared the facial features that drive perceptions of trustworthiness and dominance from ethnically diverse faces in East Asian (see Chapter 3) and White Western observers (see Chapter 2) to address two key questions: are the facial features that drive social trait perception from ethnically diverse faces influenced by the culture of the observer, and do they map onto evolutionarily adaptive emotion cues across cultures. Results showed that, while perceptions of trustworthiness and dominance are similar between cultures at the rating level, the facial features that drive these perceptions across face ethnicities vary according to the culture of the observer. Using information theoretic analyses, I show that across face ethnicities social trait perception is driven by a combination of facial features that are culturally shared plus those that are culture specific. The culturally shared facial features, although subtle, reflect previous findings (Albert et al., 2021; Oosterhof & Todorov, 2008; Zebrowitz et al., 2003). Additionally, while the features specific to Western culture resemble emotion cues, such as smiling and frowning, the features specific to East Asian culture do not. Subsequent analyses confirmed that only the Western culture-specific facial features map onto emotion cues (i.e., Action Units) and correlate with basic emotions in line with previous work on emotion overgeneralization (e.g., Albohn & Adams, 2021b; Jaeger & Jones, 2022; Said et al., 2009; Zebrowitz & Montepare, 2008).

My results show that, although ratings of trustworthiness and dominance given by East Asian and White Western observers are positively correlated between cultures, they are driven by different facial features. This challenges previous claims of universality, which focused on rating-level similarities rather than the mechanisms underlying those ratings. Specifically, previous work has supported universality by showing that social trait judgments of the same face are positively correlated across cultures (e.g., Albright et al., 1997; Walker et al., 2011; A. Wang et al., 2024; Zebrowitz et al., 2012). Critically, such rating-level comparisons do not unpack *how* observers from different cultures arrive at the same perception and are therefore limited in their ability to substantiate claims of universality (Medin et al., 1993; Schyns et al., 2022). By modelling the specific facial features East Asian and White Western observers use to perceive social traits from ethnically diverse faces, my results reveal clear cross-cultural differences in the causal mechanisms underlying social trait perception. Specifically, in line with emerging evidence (Mo et al., 2022; Rostovtseva et al., 2024) and modern theories of social trait perception (Over & Cook, 2018; Over et al., 2020, see also subsection 1.2.6), I show that the cultural background of the observer causally alters the facial features used to perceive trustworthiness and dominance from ethnically diverse faces. Together, these findings challenge the cultural universality of social trait perception and motivate a shift from coarser measures of agreement toward causal examinations of social behaviour that can reveal, rather than obscure, cultural variance.

My results also question the generalizability of prominent theories such as emotion overgeneralization (Zebrowitz & Montepare, 2008, see also subsection 1.2.2). Specifically, I show that the facial features specific to Western culture resemble specific emotion cues—such as smiling (e.g., Sharp Lip Puller—AU13) and frowning (e.g., Brow Lowerer—AU4)—that in turn systematically correlate with basic emotions. In contrast, the features specific to East Asian culture do not map onto emotion cues or do so in ways that contrast previous work (e.g., dominant-surprise correlation for female faces; e.g., Said et al., 2009). This goes against previous work which has repeatedly identified emotion resemblance as a key driver for social trait judgements (e.g., Adams et al., 2012; Albohn & Adams, 2021b; Jaeger & Jones, 2022; Montepare & Dobish, 2003; Said et al., 2009; Windmann et al., 2023; Zebrowitz et al., 2010). However, such work has been predominantly conducted within a Western context, leaving open questions about its generalizability across cultures. By including both ethnic and cultural diversity, I show that emotion overgeneralization effects generalize across face ethnicities only within a Western cultural context. These findings may reflect cultural differences in emotion evaluation: East Asians tend to favour low intensity positive emotions (e.g., Park et al., 2016; Tsai et al., 2016, 2019) and are more accepting of negative emotions (e.g., Tsai, 2017; Uchida & Kitayama, 2009), possibly leading them to rely less on stereotypical emotion cues—like smiling and frowning—when perceiving trustworthiness and dominance from neutral faces. Finally, my findings diverge from Sutherland et al. (2018), who reported that both East Asian and Western observers associate smiling with trustworthiness. However, their study included both neutral and expressive faces, placing it outside the focus on neutral faces of emotion overgeneralization.

Finally, my results suggest that the emotion cues that guide social trait perception across face ethnicities in Western observers reflect overgeneralization to broader signals than specific emotions. Specifically, the trustworthy features resemble Action Units (AUs) typically included in happy expressions, such as Sharp Lip Puller (AU13), in line with previous work (e.g., Jaeger & Jones, 2022; Montepare & Dobish, 2003; Said et al., 2009). However, unlike previous work, they also resembled AUs typically associated with sadness (e.g., Lip Corner Depressor—AU15). Similarly, the dominant features map onto AUs involved with disgust (e.g., Nose Wrinkler—AU9) as in previous work (Albohn & Adams, 2021a; Montepare & Dobish, 2003) but also fear (e.g., Brow Lowerer—AU4) and sadness (e.g., Lip Corner Depressor—AU15). Notably, extant work on emotion overgeneralization has tended to dichotomize emotions according to valence and test only few emotions (e.g., happiness vs anger; Adams et al., 2012; Albohn & Adams, 2021b). My own and previous results may therefore more reflect broader signals such as negative vs positive affect or affiliation vs dominance, consistent with modern theories of social perception (e.g., Adams & Kveraga, 2015, see also subsection 1.2.4). Nevertheless, the correlation between dominant features and both sad and fearful facial expressions remains counter-intuitive and contrasts previous work (e.g., Adams et al., 2012; Albohn & Adams, 2021b). Notably, these results reflect similarities in individual AUs rather than perceived emotions; thus, while faces perceived as dominant may comprise features that are typically included in sad and fearful expressions,

dominant features as a whole are not necessarily perceived as such. Together, this highlights the difficulty of drawing clear comparisons between complex, multidimensional face signals.

Importantly, the current work is not without limitations. Firstly, the results only capture variance from two main cultures: East Asian and Western. With growing calls to move beyond the "East-West" dichotomy (e.g., Kitayama & Salvador, 2024; Vignoles et al., 2016), future work should examine how other cultural backgrounds may influence social trait perception. Secondly, I have tested universality using a relatively conservative framework, where only those facial features that are shared across face ethnicities and cultures are considered shared. As growing evidence suggests that different combinations of observer and face characteristics can interact in perception in different ways (e.g., Hehman et al., 2017; Xie et al., 2019), Chapter 5 will explore in more depth how observer culture and face ethnicity combine and interact to influence the facial features used for social trait perception. Lastly, as in previous chapters, the current work only explores variance in the static facial features driving social trait judgements. Future work should test the cultural universality of social trait judgements inferred from other sources of social information such as facial expressions (e.g., Gill et al., 2014), body shape (e.g., Hu et al., 2018), and the voice (e.g., McAleer et al., 2014).

In sum, the work reported in this chapter challenges the cultural universality of social trait perception implied by prominent theories (e.g., Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008) and supported by previous rating-level examinations (e.g., Albright et al., 1997; A. Wang et al., 2024; Zebrowitz et al., 2012) by showing that the culture of the observer causally alters the facial features East Asian and White Western observers use to perceive trustworthiness and dominance from ethnically diverse faces. Further, I show that while White Western observers use facial features that resemble emotion cues to perceive social traits across face ethnicities in line with previous work (e.g., Jaeger & Jones, 2022; Montepare & Dobish, 2003; Zebrowitz & Montepare, 2008), East Asian observers do not. Together, these results question the generalizability of current theories and highlight the importance of diversity in advancing knowledge of social perception.

Chapter 5

Face ethnicity and observer culture interact to causally influence social trait perception

Chapter abstract

Modern theories argue that social trait perception is influenced by the characteristics of the face, the observer, and their unique interaction. Recent efforts have quantified how these three sources of information explain variance in social trait ratings, yet it remains unknown whether and how they alter the causal facial features driving these ratings. In this final chapter, I address this gap by examining whether and how the facial features used by East Asian and White Western observers to perceive trustworthiness and dominance from ethnically diverse faces vary according to the ethnicity of the face, the culture of the observer, and their synergistic interactions. Results show that social trait perception is driven by four distinct sets of facial features: those that are shared across face ethnicities and cultures, those that are specific to each face ethnicity, those that are specific to each observer culture, and those that are specific to each combination. Subsequent analyses confirm that these feature sets closely mirror the appearance of the original modelled features, suggesting that they represent key sources of variance in social trait perception. Together, these findings extend previous ratings-based examinations and provide a mechanistic explanation for how the face, the observer, and their interaction influence social trait perception in line with modern theories.

5.1 Introduction

As shown in the previous chapters of this thesis, the facial features that drive perceptions of trustworthiness and dominance are influenced by both the ethnicity of the face within a given culture (see Chapters 2 and 3) as well as by the culture of the observer across face ethnicities (see Chapter 4). These findings highlight the role of two distinct, yet complementary, sources of variance in social perception: the face being perceived, and the observer perceiving it. Although typically studied (or at least focused on) in isolation, face and observer characteristics are known to perceptually interact. For example, cultural beauty standards influence which facial features are considered attractive (e.g., Han et al., 2018; Zhan et al., 2021), and racial prejudice differentially affects judgements of same- vs other-ethnicity faces (e.g., Hugenberg, 2005; Hutchings et al., 2024). These interactive effects are at the core of many modern theories of social perception (e.g., Freeman & Ambady, 2011; Over et al., 2020, see also subsections 1.2.5 and 1.2.6), and indeed broader models of social communication (e.g., see Jack & Schyns, 2015; Scott-Phillips, 2008, for reviews). Within such frameworks, social perceptions arise not only from the individual information that the face and the observer afford, but also their unique interaction. In line with this, emerging evidence now shows that face $\times$ observer interactions explain a considerable portion of variance in social trait judgements (e.g., Hehman et al., 2017; Hönekopp, 2006; Xie et al., 2019; see also Hehman et al., 2019, for a review).

Understanding how face and observer characteristics individually and jointly influence social trait perception is key to diversifying current knowledge, as cultural norms, ethnic stereotypes, and intergroup biases are likely to manifest and interact in different ways across ethnic and cultural boundaries (e.g., see Fiske, 2017; Kawakami et al., 2022). Reflecting this, the relative contributions of face and observer characteristics to social trait perception vary across face ethnicities (Xie et al., 2019) in line with each observer's stereotype knowledge (Xie et al., 2021). Similarly, perception of other social cues known to influence social trait judgements, such as emotion cues (e.g., Said et al., 2009; Zebrowitz & Montepare, 2008, see also Chapter 4) and ethnic prototypicality (e.g., Blair et al., 2002; Hutchings et al., 2024, see also Chapter 2), are influenced by both the face and the observer (Hehman et al., 2017; Norman et al., 2024). To quantify the relative contributions of face and observer characteristics as well as their interaction, previous work has either modelled the ratings variance explained at each level (i.e., between faces, between observers, and in face $\times$ observers interactions; e.g., Hehman et al., 2017; Xie et al., 2019), or compared (dis)similarity matrices derived from different levels (e.g., between faces vs between observers; e.g., Stolier et al., 2018; Xie et al., 2021). Although insightful, such work is not designed to reveal whether and how these effects influence the causal features driving social trait judgements—that is, which facial features are influenced by face ethnicity only, by observer culture only, and by the combination of face ethnicity and observer culture.

Achieving a feature-level understanding of how face and observer characteristics interact in social trait perception has the potential to directly complement and extend current efforts

(e.g., Hehman et al., 2017; Xie et al., 2019, 2021). Specifically, previous work has quantified the overall variance in *ratings* explained by the face, the observer, and face $\times$ observer interactions. However, it remains unknown whether and how these effects alter the *causal features* driving the ratings. Addressing this question is key to deepening knowledge of social trait perception beyond coarser measures of (dis)similarity to identify the specific mechanisms underlying these effects (see Medin et al., 1993; Schyns et al., 2022, for further discussion). Doing so also has direct implications for modern theoretical accounts (e.g., Freeman & Ambady, 2011; Over et al., 2020) which argue for—but to date do not causally explain—the interactive nature of the face and the observer. My own results support this potential by showing that ethnic variance in the face differentially affects the facial features used for social trait perception by White Western and East Asian observer (see subsection 3.3.3). In other words, the bottom-up influence of face ethnicity is modulated by the top-down influence of observer culture.

To this end, in this final chapter I continue to formally compare the 3D models presented in previous chapters (see Chapters 2 and 3) to reveal whether and how the culture of the observer and the ethnicity of the face individually and jointly influence the specific facial features used for social trait perception. Specifically, I aim to identify which (if any) social trait facial features are shared across observer cultures and face ethnicities, which are specific to each observer culture regardless of face ethnicity, which are specific to each face ethnicity regardless of observer culture, and which are specific to each combination of observer culture and face ethnicity.

5.2 Methods

As in Chapter 4, in this chapter I continue to formally compare the 3D face models of White Western and East Asian observers obtained in Chapters 2 and 3 respectively, without collecting any additional data. The full methodological details for obtaining the 3D face models can be reviewed in the methods sections of Chapters 2 and 3.

5.3 Results

To test whether and how the ethnicity of the face as well as the culture of the observer individually and jointly influence the facial features used for social trait perception, I continued to formally compare the 3D models of the facial features that drive perceptions of trustworthiness and dominance from Black African (BA), East Asian (EA) and White European (WE) faces in 60 individual East Asian observers (see Chapter 3) and 60 individual White Western observers (see Chapter 2). For clarity, Figure 5.1 shows the results aggregate across individual observers ($n = 20$ per face ethnicity condition in each observer culture) for trustworthy-looking and dominant-looking male faces in the same way as in Chapter 4 (see Figure 5.2 for female faces).

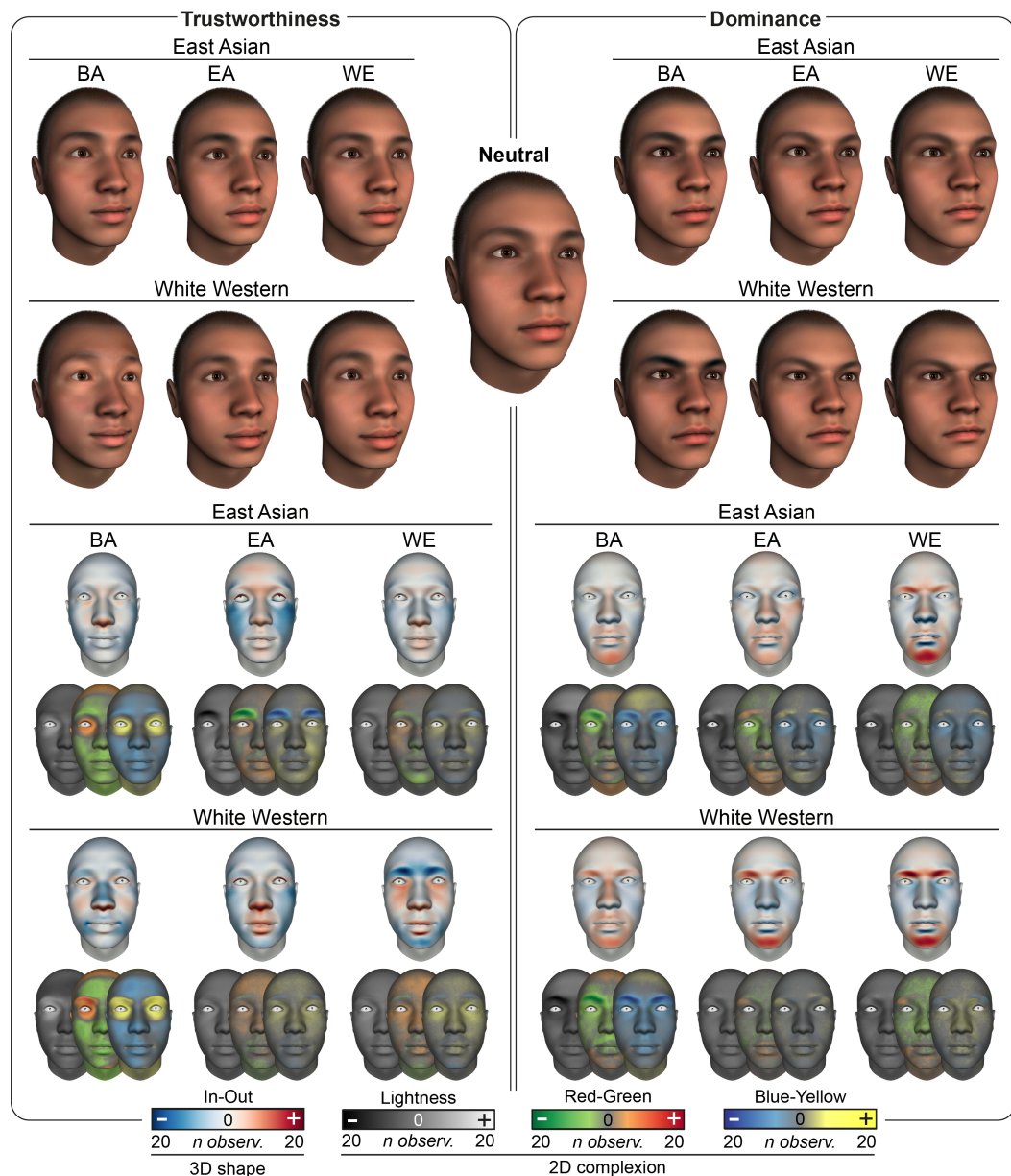


Figure 5.1: Faces at the top show the 3D models of male faces perceived as trustworthy (left) and dominant (right) by East Asian observers (top row) and White Western observers (bottom row) for each face ethnicity—Black African (BA), East Asian (EA) and White European (WE)—displayed on the same ethnically neutral face (see larger face in the centre) for comparison. Results are aggregated across observers ($n = 20$ per face ethnicity condition in each observer culture) and include all identity components that meet or exceed the population prevalence threshold ($n = 4$ observers; faces extrapolated for clarity). Below, colour-coded faces show the results separately for 3D shape and 2D complexion. Colour-coding denotes the feature deviation direction from the average ethnically neutral face; colour saturation shows the number of observers with a statistically significant effect (see colorbars below).

Faces at the top show the 3D face models of trustworthy-looking (left) and dominant-looking (right) male faces for East Asian (top row) and White Western observers (bottom row), displayed on the same ethnically neutral face (see larger face in the centre). Colour-coded faces below show the results for 3D shape and 2D complexion, using the same format as in previous chapters (see colorbars below). Colour saturation shows the number of observers ($n = 20$ per face ethnicity condition in each observer culture) with a statistically

significant effect, for all identity components at or above the population prevalence threshold ($n = 4$ observers; Donhauser et al., 2018; Ince et al., 2021).

As discussed in Chapter 4, visual inspection of the results across face ethnicities reveals both cross-cultural similarities—such as the association of smaller (vs larger) faces with trustworthy (vs dominant) judgements by both East Asian and White Western observers—and cross-cultural differences—such as the more frequent use of emotion-resembling features (e.g., upturned mouth corners, lower brow bone) by White Western relative to East Asian observers. In addition to these cross-cultural effects, the results also show differences between face ethnicities regardless of the culture of the observer. For trustworthy judgements, BA faces tend to have lighter, warmer eye regions against darker, cooler skin tones; EA faces tend to have a narrower face with heavier upper eyelids and lower cheekbones; and WE faces tend to have cooler chins against warmer skin tones. For dominant judgements, BA faces tend to have a larger nose with darker, cooler eyebrows and eye regions; EA faces tend to have slightly lighter, warmer eyebrows; and WE faces tend to have a lower brow bone and a more prominent, angular chin. Finally, the results also suggest that the facial features used to perceive trustworthiness and dominance vary according to specific combinations of face ethnicity and observer culture. For example, more East Asian than White Western observers perceive EA faces with darker, cooler eyebrows as trustworthy (see hue variations in eyebrow region); and East Asian, but not White Western, observers perceive EA faces with more angled eyes and higher cheekbones as dominant.

Results for female faces (see Figure 5.2) show a similar pattern of similarities and differences across face ethnicities and observer cultures. For example, across face ethnicities and observer cultures, smaller vs larger faces are perceived as trustworthy and dominant respectively. Additionally, regardless of the ethnicity of the face, White Western observers more frequently associate the protrusion of the brow bone with dominance, and the upturning of the mouth corners with trustworthiness regardless of the ethnicity of the face. Further, regardless of the culture of the observer, BA faces with lighter, warmer eye regions against darker, cooler skin tones are perceived as trustworthy; and WE faces with a wider jaw and cooler eye regions against warmer eyebrows and skin tones are perceived as dominant. Finally, the facial features that drive social trait perception from female faces also vary according to specific combinations of face ethnicity and observer culture. For example, East Asian, but not White Western, observers perceive BA faces with a lower brow bone as trustworthy, and EA faces with more angled eyes, a longer nose, and a cooler skin tone as dominant.

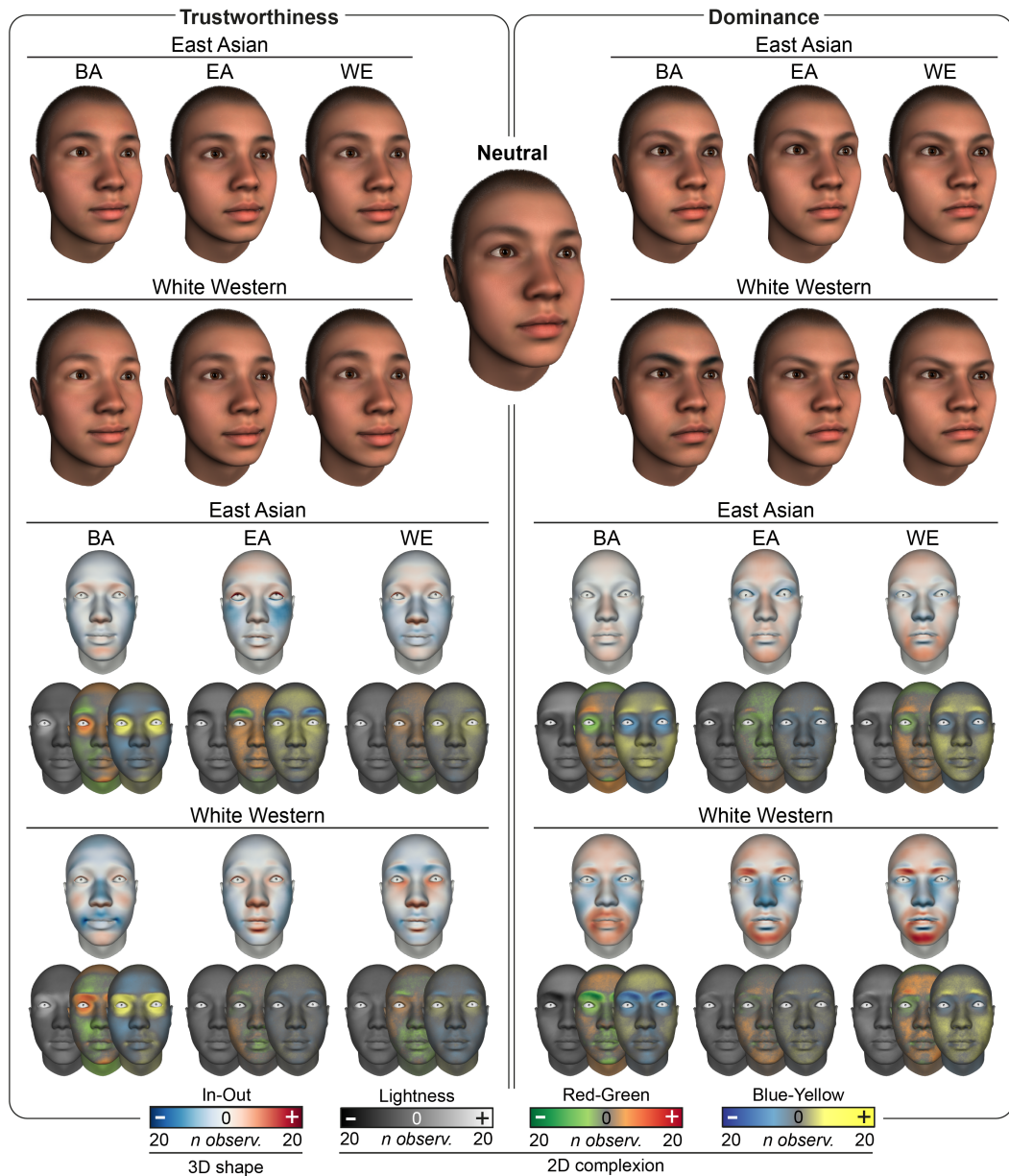


Figure 5.2: Following the same format as figure 5.1, faces at the top show the 3D models of female faces perceived as trustworthy (left) and dominant (right) by East Asian observers (top row) and White Western observers (bottom row) for each face ethnicity, displayed on the same ethnically neutral face (see larger face in the centre) for comparison. Below, colour-coded faces show the results separately for 3D shape and 2D complexion (see colorbars below). Results are aggregated across observers ($n = 20$ per face ethnicity in each observer culture), with all identity components meeting or exceeding population prevalence threshold of $n = 4$.

Together, these results suggest that the facial features driving social trait perception from ethnically diverse faces in East Asian and White Western observers vary according to the ethnicity of the face, the culture of the observer, and their combination.

5.3.1 Facial feature variance predicts differences in face ethnicity and observer culture

To formally test the distinctiveness of these visually inspected effects, I first obtained a global measure of (dis)similarity using a machine learning approach. Specifically, using a stratified k -fold cross-validation approach ($k = 10$), I trained a Support Vector Machine (SVM) to classify the 3D face models shown in Figures 5.1 and 5.2 according to both the ethnicity of the face and the culture of the observer, separately for 3D shape and 2D complexion and for each social trait and stimulus sex. In each fold, I optimized the hyperparameters for coding design, learner type, and lambda automatically, and trained the classifier to predict 6 class labels ([2 observer culture $\times$ 3 face ethnicities], e.g., East Asian observer perceiving Black African faces). High classification performance would indicate that the facial features East Asian and White Western observers use to perceive trustworthiness and dominance from ethnically diverse faces are sufficiently distinct; above-chance misclassification (i.e., $1 \div$ number of incorrect classification options) would indicate they are not. Across folds, the SVM had a mean classification error of 43.96% for the male 3D shape models, 41.46% for the male 2D complexion models, 46.67% for the female 3D shape models, and 40% for the female 2D complexion models. Figure 5.3 shows the results aggregated across folds for dominant and trustworthy male face models, separately for 3D shape and 2D complexion (see Figure 5.4 for results for female face models).

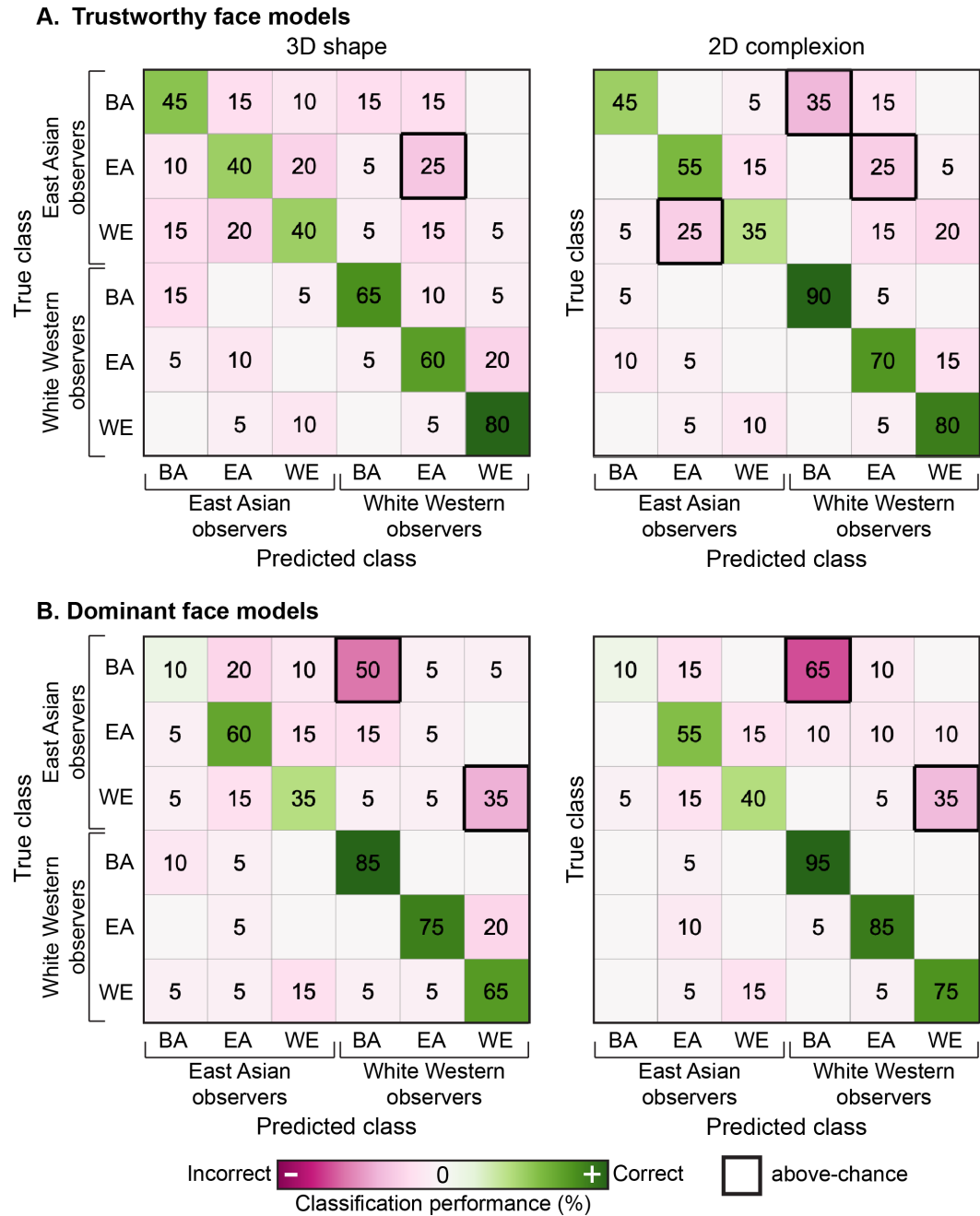


Figure 5.3: In each panel (A: Trustworthy face models, B: Dominant face models), colour-coded matrices show the SVM classification (x axis) of the 3D male face models (y axis), separately for 3D shape (left) and 2D complexion (right). Green colour-coding indicates correct classifications, pink colour-coding indicates incorrect classifications (see colorbar below). Colour saturation and numeric values in each cell report the percentage of (in)correctly classified 3D face models. Thick black outlines indicate above-chance misclassifications.

In each panel of Figure 5.3 (A: Trustworthy face models; B: Dominant face models), colour-coded matrices show the SVM classification according to face ethnicity and observer culture (x axis; face ethnicity: inner labels, observer culture: outer labels) of the 3D face models (y axis), separately for 3D shape (left) and 2D complexion (right). Green colour-coding indicates correct classifications, and pink colour-coding indicates incorrect classifications. Colour saturation and numeric values in each cell show the percentage of (in)correctly classified 3D face models ($n = 120$ per social trait and stimulus sex [20 observers $\times$ 3 face

ethnicities $\times$ 2 observer cultures]). Thick black outlines indicate above-chance misclassifications.

Results show moderate to high classification performance (see diagonal). However, for both trustworthy and dominant models, the mean classification performance across face ethnicities is higher for the models of White Western observers (Trustworthy—3D shape: 68%; 2D complexion: 80%; Dominant—3D shape: 75%; 2D complexion: 85%) than East Asian observers (Trustworthy—3D shape: 42%; 2D complexion: 45%; Dominant—3D shape: 35%; 2D complexion: 35%). Additionally, results also show some systematic errors, particularly for the 3D face models of East Asian observers. For trustworthy, the culture of the observer is misclassified above-chance for the East Asian observers' models of EA (3D shape: 25%; 2D complexion: 25%) and BA faces (2D complexion: 35%), and the ethnicity of the WE models of East Asian observers is misclassified above-chance as EA ethnicity (2D complexion: 25%). For dominant, the culture of the observer is misclassified above-chance for the East Asian observers' models of BA (3D shape: 50%; 2D complexion: 65%) and WE faces (3D shape: 35%; 2D complexion: 35%).

Results for female face models (see Figure 5.4) show similar performance, including poorer mean classification performance across face ethnicities for the models of the East Asian observers (Trustworthy—3D shape: 38%; 2D complexion: 43%; Dominant—3D shape: 35%; 2D complexion: 53%) vs White Western observers (Trustworthy—3D shape: 63%; 2D complexion: 70%; Dominant—3D shape: 65%; 2D complexion: 77%), but with some distinct errors. For example, the culture of the observer is misclassified above-chance for the East Asian observers' models of BA, but not EA, trustworthy faces (3D shape: 50%; 2D complexion: 50%). Additionally, the ethnicity of the White Western observer's models of EA faces is misclassified above-chance as WE for trustworthy models (3D shape: 25%; 2D complexion: 30%), and is misclassified above-chance as BA for dominant models (3D shape: 25%).

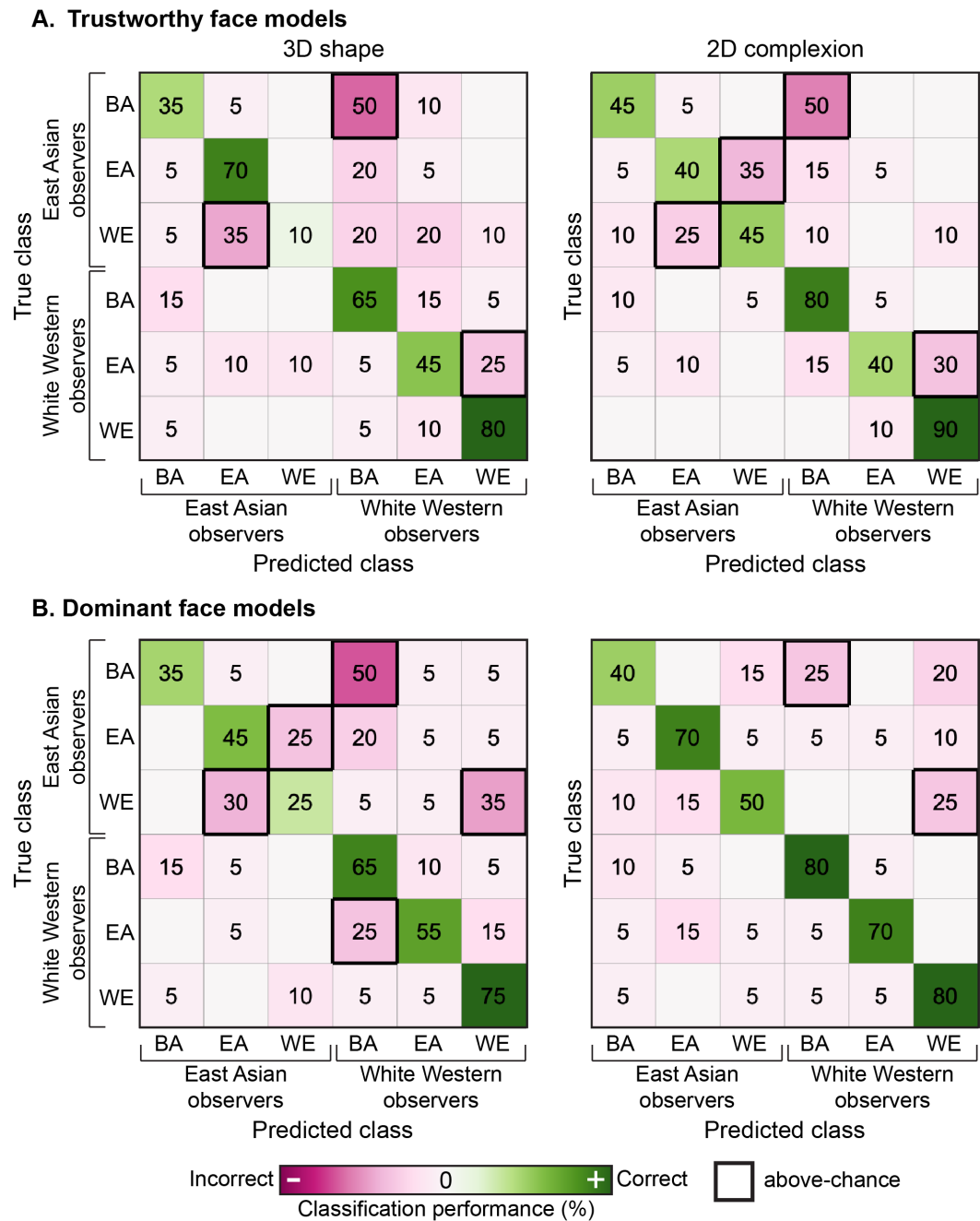


Figure 5.4: Following the same format as Figure 5.3, colour-coded matrices show the SVM classification (x axis) of the 3D female face models (y axis), separately for trustworthy (A) and dominant models (B) and for 3D shape (right) and 2D complexion (left). Colour-coding indicates correct (green) vs incorrect (pink) classifications; colour saturation and numeric values in each cell report the percentage of (in)correctly classified 3D face models (see colorbar below). Thick black outlines indicate above-chance misclassifications.

Together, these results indicate that, while there are some similarities across observer cultures and face ethnicities as evidenced by the above-chance misclassifications, the facial features that East Asian and White Western observers use to perceive social traits from ethnically diverse faces are sufficiently distinct to enable correct classification.

5.3.2 Measuring the individual and joint influence of face ethnicity and observer culture on the facial features used for social trait perception

To reveal which specific facial features differentiate the social trait models according to both the culture of the observer and the ethnicity of the face, I proceeded in three complementary steps. For clarity, I now detail the full approach before presenting the specific results for each step. Figure 5.5 illustrates the approach using trustworthy-looking male faces as an example.

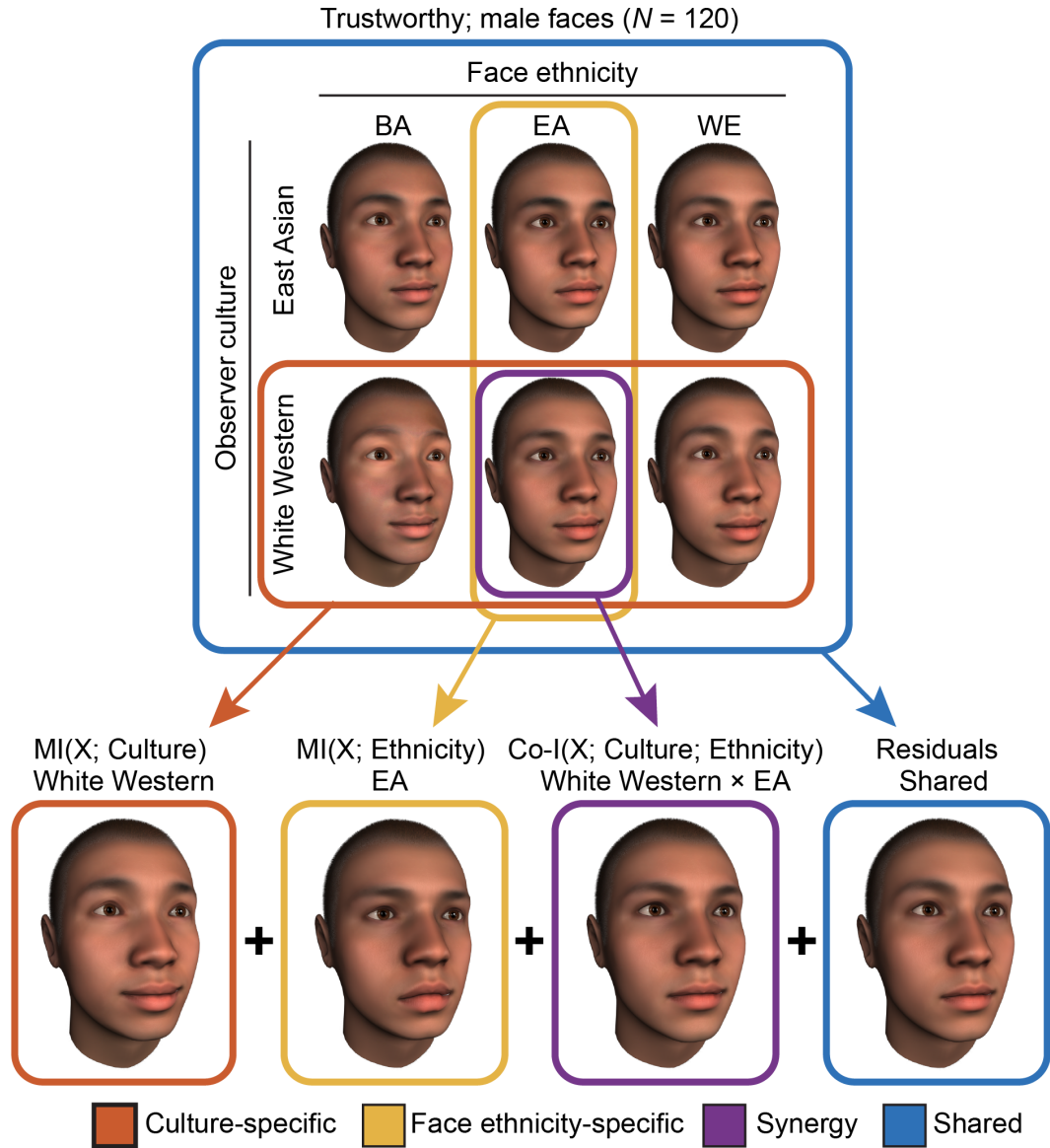


Figure 5.5: Faces at the top represent the 3D face models of trustworthy-looking male faces of all 120 observers, organized by observer culture (y axis) and face ethnicity (x axis). Larger faces below show the facial features that are specific to White Western culture (orange colour-coded); those that are specific to EA faces (yellow colour-coded); those that are synergistic for White Western $\times$ EA interaction; and those that are shared across all three ethnicities and both observer cultures (blue colour-coded; see legend below).

Prior to any analyses I first pooled the 3D shape and 2D complexion models of all 120

observers ($[20 \text{ observers} \times 3 \text{ face ethnicities} \times 2 \text{ observer cultures}]$), separately for each social trait and stimulus sex. The faces at the top of Figure 5.5 schematize this using the average models of trustworthy-looking male faces for each face ethnicity (x axis) and observer culture (y axis). I then represented each modelled identity component at or above population prevalence threshold ($n = 4$ observer) for at least one face ethnicity and observer culture as a 1×120 vector, binary coded as statistically significant or insignificant for each observer. Parallel to this, I obtained three vectors of the same length detailing the ethnicity of the face, the culture of each observer, and their combination (e.g., White Western observer perceiving East Asian faces). For example, at the top of Figure 5.5, all the models represented by the faces highlighted in orange ($n = 60$ [$20 \text{ observers} \times 3 \text{ face ethnicities}$]) would be coded as "White Western"; those represented by the faces highlighted in yellow ($n = 40$ [$20 \text{ observers} \times 2 \text{ observer cultures}$]) would be coded as "EA"; those represented by the face highlighted in purple ($n = 20$) would be coded as "White Western $\times$ EA".

Having thus coded the models, I first identified which facial features vary only according to the culture of the observer (e.g., White Western culture; orange colour-coded in Figure 5.5) and only according to the ethnicity of the face (e.g., EA faces; yellow colour-coded in Figure 5.5). To do this, I measured the specificity of each modelled identity component to observer culture across face ethnicities, and to face ethnicity across observer cultures using Mutual Information (MI; Cover & Thomas, 1991; Ince et al., 2017). Specifically, I complemented the culture-specific results presented in Chapter 4 by measuring the MI between each identity component and each face ethnicity label in turn (e.g., "EA face" vs "not EA face"). I established statistical significance using non-parametric permutation testing ($N = 1,000$ permutations) and used Pearson correlation to discard any identity components with a high MI value but negative correlation with face ethnicity (i.e., the identity component has a high MI with face ethnicity because it is repeatedly absent in these 3D face models). Finally, I discarded any identity components that had high MI with face ethnicity but were not at or above the population prevalence threshold ($n = 4$ observers) in both observer cultures for that face ethnicity. In Figure 5.5, the larger faces at the bottom show the resulting facial features of 3D shape and 3D complexion that are specific to White Western culture (orange colour-coded) and those that are specific to EA faces (yellow colour-coded).

Next, I identified those facial features that vary according to each specific combination of face ethnicity and observer culture (e.g., White Western observer perceiving East Asian faces; purple colour-coded in Figure 5.5). To do this, I measured the specificity of each identity component to the combination of face ethnicity and observer culture using Co-Information (Co-I; Bell, 2003; see also Timme et al., 2014). Broadly defined, Co-I is an extension of MI that measures the synergy vs redundancy of the information that two separate variables (here, face ethnicity and observer culture) provide about a third variable (here, modelled identity components). Here, I focus on synergy as it would indicate that the identity component is uniquely specific to a given combination of face ethnicity and culture (Bell, 2003; Timme et al., 2014). To measure Co-I, I first measured the MI between each modelled identity component and each combination of face ethnicity and observer culture (e.g., East

Asian observer perceiving Black African faces). I then computed the difference between this MI value and the sum of the MI between the same identity component and only face ethnicity and only observer culture derived as above; a negative resulting value would indicate synergy (see Bell, 2003; Timme et al., 2014, for mathematical details). To establish statistical significance, I used non-parametric permutation testing as follows: on each iteration ($N = 1,000$), I randomly shuffled the labels of face ethnicity and observer culture and re-computed all three sets of MI and the subsequent Co-I. I then used the 5th percentile of the distribution of chance Co-I values as threshold for statistical significance of synergistic results. In Figure 5.5, the purple colour-coded face shows the synergistic results for the combination of White Western culture and EA faces.

Finally, I identified those facial features that drive social trait perception regardless of the ethnicity of the face and the culture of the observer—that is, those facial features that are shared across face and observer characteristics (blue colour-coded in Figure 5.5). Specifically, I considered a facial feature as shared if the identity component was not specific to any observer culture, any face ethnicity, and any combination of observer culture and face ethnicity but was at or above population prevalence threshold ($n = 4$ observers) for all three face ethnicities and both observer cultures. In other words, I used the residuals of each set of analyses above, provided they still represent a true effect in each face ethnicity condition in both the populations sampled. The blue colour-coded face at the bottom of Figure 5.5 shows the shared facial features of trustworthy male faces.

I repeated these analyses for each modelled identity component of 3D shape and 2D complexion at or above population prevalence threshold ($n = 4$ observer) for at least one face ethnicity and one observer culture, separately for each identity component weight direction (positive, negative), social trait, and stimulus sex. Therefore, for each social trait and stimulus sex, I identified the facial features that are shared across observer cultures and face ethnicities, those that are specific to each observer culture regardless of face ethnicity, those that are specific to each face ethnicity regardless of observer culture, and those that are synergistic for each combination of observer culture and face ethnicity. In other words, these results capture how the characteristics of both the face and the observer individually and jointly influence the causal facial features used for social trait perception, mirroring the focus of previous work and complementing their findings (e.g., Hehman et al., 2017; Xie et al., 2019). In the following subsections, I present the full results for each feature set, before examining their relative contributions.

5.3.3 Face ethnicity and observer culture individually influence social trait perception

I first identified the facial features that vary according to only the culture of the observer, as in Chapter 4, and only the ethnicity of the face as described above (see subsection 5.3.2). Figure 5.6 shows these face ethnicity-specific results combined with the culture-specific results shown in Chapter 4 for trustworthy-looking and dominant-looking male faces (see Fig-

ure 5.7 for female faces).

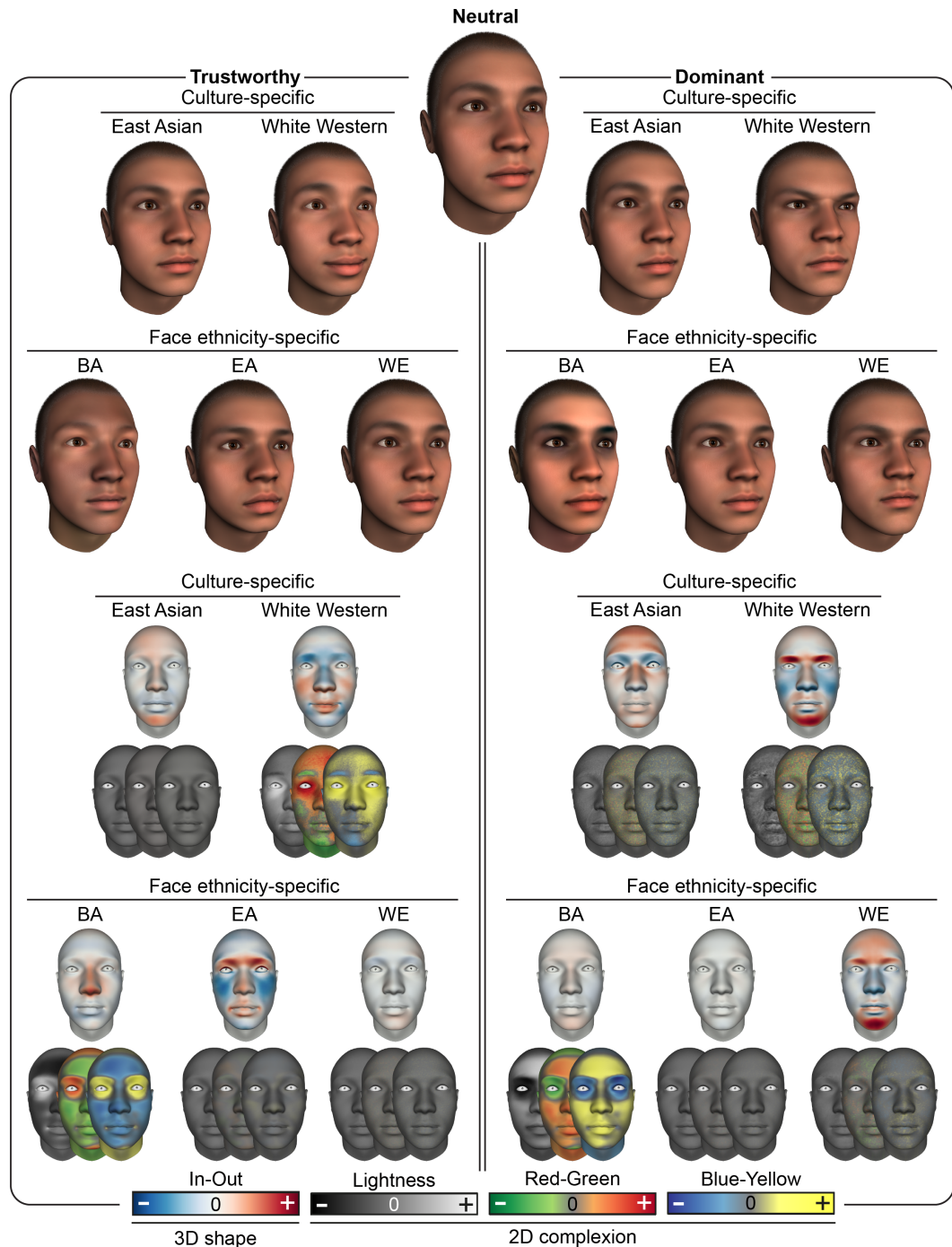


Figure 5.6: Faces at the top of each panel (Left: Trustworthy; Right: Dominant) show the facial features of male faces that are culture-specific (top row) and those that are face ethnicity-specific (bottom row), with shape and complexion combined and displayed on the same ethnically neutral face (see larger face at the top). Colour-coded faces below show the results separately for 3D shape and 2D complexion (see Supplementary materials 7.4.1 and 7.5.1 for further details on the culture-specific and face ethnicity-specific features respectively), with colour-coding following the same format as before (see colorbars below) and values normalized per social trait for display purposes.

Faces in the top panel show, for both trustworthy (left panel) and dominant (right panel) judgements, the facial features of male faces that are culture-specific regardless of face ethnicity (top row) and those that face ethnicity-specific regardless of observer culture (bottom

row), with shape and complexion combined and displayed on the same ethnically neutral face (see larger face at top). Colour-coded faces below show the results separately for 3D shape and 2D complexion using the same colour-coding in previous chapters (see colorbars below), with values normalized per social trait for display purposes (see Supplementary materials 7.4.1 and 7.5.1 for further details on the number and distribution of the culture-specific and face ethnicity-specific features respectively). Results show that both the culture of the observer and the ethnicity of the face individually influence which facial features are used for social trait perception.

As described in further detail in Chapter 4, the culture-specific facial features primarily reflect differences in the use of emotion-resembling facial features. Briefly stated, the Western-culture specific features largely align with previous work (e.g., Jaeger & Jones, 2022; Oosterhof & Todorov, 2008; Said et al., 2009) by mapping onto emotion cues such as smiling and frowning. In contrast, the East Asian culture-specific features do not map onto emotion cues and instead mostly capture variance in face width for trustworthy perceptions and face length for dominant perceptions (see Chapter 4 for further details).

In addition to these culture-specific effects, the results also show that the social trait facial features vary according to the ethnicity of the face regardless of the culture of the observer. For trustworthy perception, the BA-specific features comprise a bigger nose and thinner lips together with lighter, warmer eyebrows and eye regions against darker, cooler skin tones; the EA-specific features comprise a narrower face, a lower brow bone, flatter cheekbones, and fuller lips with downturned mouth corners together with a slightly warmer skin tone; and the WE-specific features comprise a higher brow bone with arched eyebrows together with no visible complexion effects. For dominant perceptions, the BA-specific features comprise a higher nose bridge with a narrower jaw and gaunter cheeks together with darker, cooler eyebrows and eye regions against lighter, warmer skin tones; the EA-specific features comprise no visible shape nor complexion effects; and the WE-specific features comprise a bigger forehead, a lower brow bone with angled eyebrows, a shorter nose, thinner lips, a narrower jaw, and a more prominent, angular chin together with slightly cooler skin tones.

Results for female faces (see Figure 5.7) largely resemble those for male faces, with some key differences in the face ethnicity-specific facial features. For trustworthy perceptions, the BA-specific features further comprise fuller cheeks and a smaller, rather than bigger, nose together with lighter, warmer chins; the EA-specific features comprise heavier upper eye lids, a higher nose bridge with a pointier nose tip, a bigger mouth and a smaller chin together with cooler eyebrows and warmer cheeks; and the WE-specific features comprise a wider forehead, arched eyebrows, and a lower mouth with fuller lips together with a warmer brow bone and mouth against cooler skin tones. For dominant perceptions, the BA-specific features comprise darker and cooler—rather than lighter and warmer—eye regions and skin tones; the EA-specific features again comprise no visible shape effects but do include lighter, warmer brow bones; and the WE-specific features further comprise a wider face with fuller cheeks.

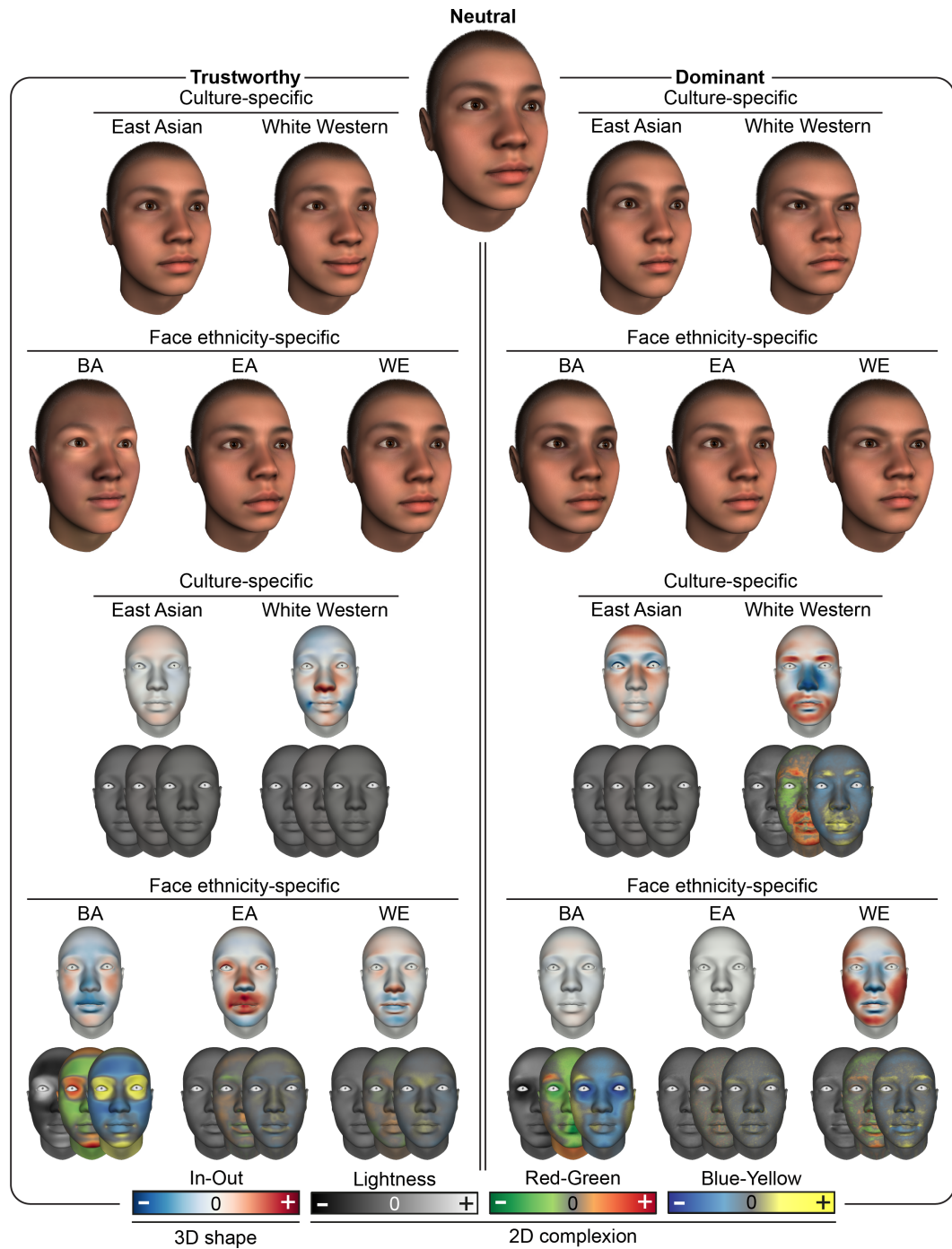


Figure 5.7: Following the same format as Figure 5.6, faces at the top show the culture-specific (top row) and face ethnicity-specific (bottom row) facial features of female faces perceived as trustworthy (left panel) and dominant (right panel). Colour-coded faces below show the results for 3D shape and 2D complexion (see colorbars below; see Supplementary materials 7.4.1 and 7.5.1 for further details).

5.3.4 Face ethnicity and observer culture synergistically influence social trait perception

I next measured the specificity of each modelled identity component to each combination of face ethnicity and observer culture using Co-Information (Co-I; Bell, 2003; see also Timme et al., 2014) as described in subsection 5.3.2. Figure 5.8 shows the synergistic results

for male faces (see Figure 5.9 for female faces).

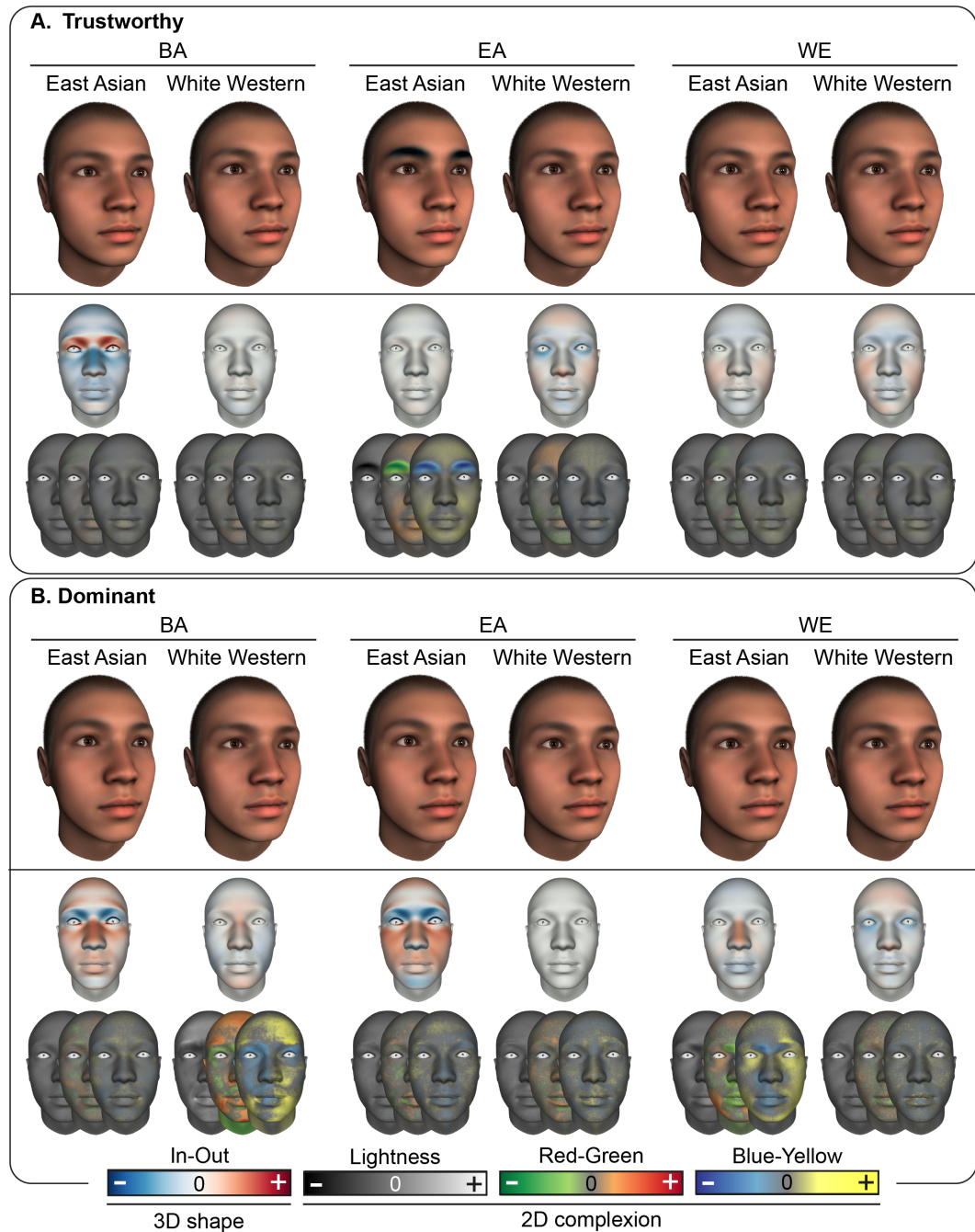


Figure 5.8: In each panel (A: Trustworthy; B: Dominant), faces at the top show the facial features of male faces that are synergistically influenced by each combination of face ethnicity and observer culture (see titles above), with shape and complexion combined and displayed on the same ethnically neutral face. Below, colour-coded faces show the results separately for 3D shape and 2D complexion, following the same colour-coding format as before (see colorbars below; see Supplementary materials 7.5.2 for further details).

In each panel (A: Trustworthy; B: Dominant), faces at the top show the facial features of male faces that are uniquely specific (i.e., synergistic) to each combination of face ethnicity and observer culture (see titles above), with shape and complexion combined and displayed on the same ethnically neutral face as in previous figures. Colour-coded faces below show the results separately for 3D shape and 2D complexion following the same colour-coding

format as before (see colorbars below; see also Supplementary materials [7.5.2](#) for further details on the number and distribution of synergistic identity components).

Results show that each combination of face ethnicity and observer culture determines the use of specific, though relatively subtle, facial features to perceive trustworthiness and dominance. For example, the facial features used to perceive EA faces as trustworthy comprise darker, cooler eyebrows against lighter, warmer skin tones for East Asian observers; while they comprise smaller, deeper-set eyes together with slightly warmer foreheads for White Western observers. Similarly, the facial features used to perceive BA faces as dominant comprise a longer face with a higher brow bone and a bigger nose together with a slightly sallower skin tone for East Asian observers; while they comprise a wider brow bone, a higher nose bridge, and a narrower jaw together with darker, cooler eyebrows and eye regions against lighter, warmer skin tones for White Western observers. Unlike previous results, however, these synergistic features do not appear to readily map onto cues known to influence social perception (e.g., emotions, sexual dimorphism, ethnic phenotypes).

Results for female faces (see Figure [5.9](#)) are similarly subtle, though not identical, to those for male faces. For example, the EA $\times$ East Asian trustworthy features further comprise wider cheeks and redder, but not yellower, skin tones. Similarly, the BA $\times$ White Western dominant features comprise no visible 3D shape features but much darker, cooler eyebrows and skin tones than for male faces.

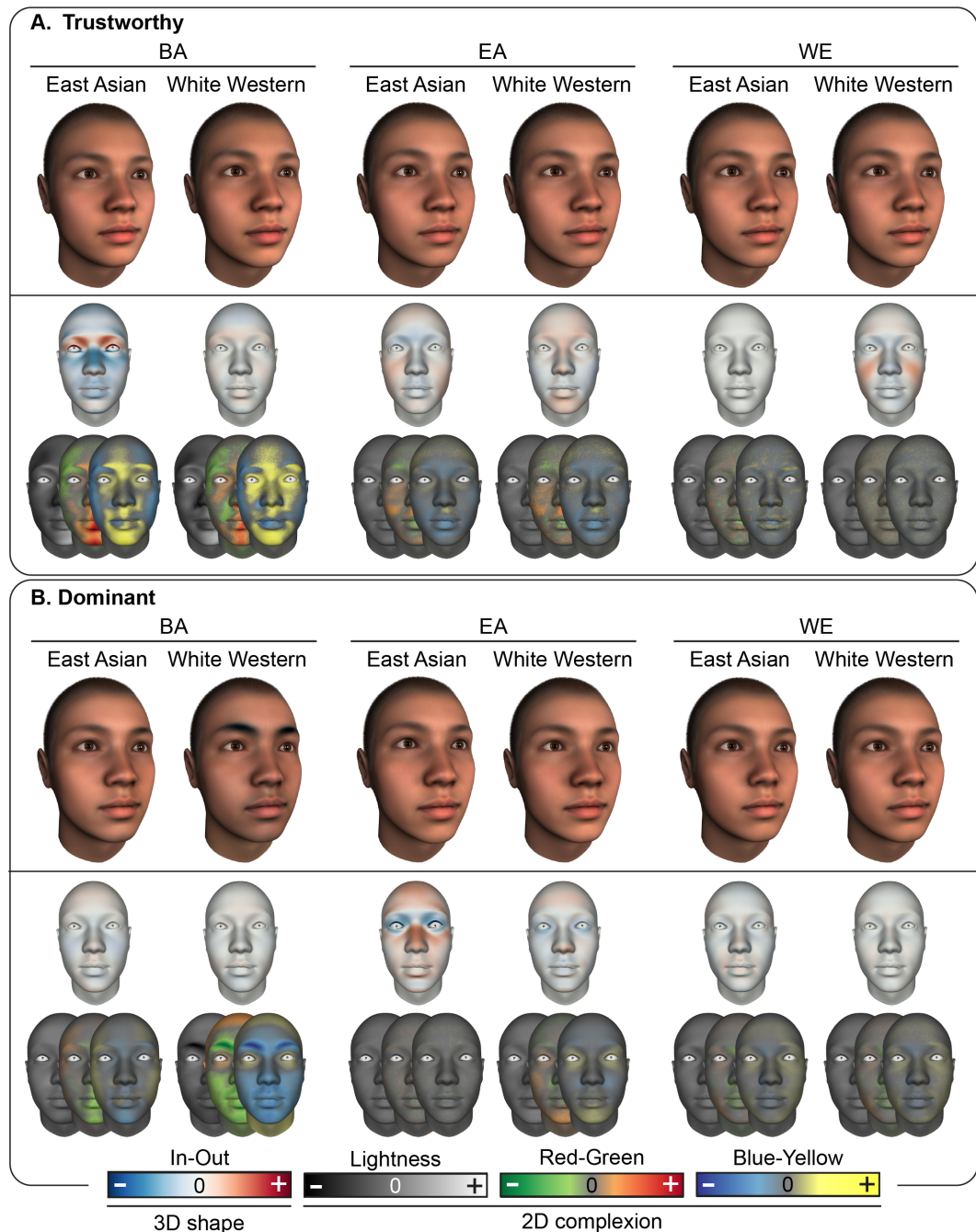


Figure 5.9: Following the same format as Figure 5.8, faces at the top of each panel (A: Trustworthy; B: Dominant) show the synergistic facial features of female faces for each combination of face ethnicity and culture (see titles at top). Colour-coded faces below show the results separately for 3D shape and 2D complexion (see colorbars below; see Supplementary materials 7.5.2 for further details).

Despite their subtlety, these findings nevertheless show that the facial features used for social trait perception are influenced by the synergistic combination of face ethnicity and observer culture.

5.3.5 Social trait perception is driven by features shared across face ethnicities and observer cultures

Finally, I used the residuals of the above analyses to identify those facial features that are shared across face ethnicities and observer cultures (see subsection 5.3.2 for details). Figure 5.10 shows the results for male and female faces perceived as trustworthy and dominant.

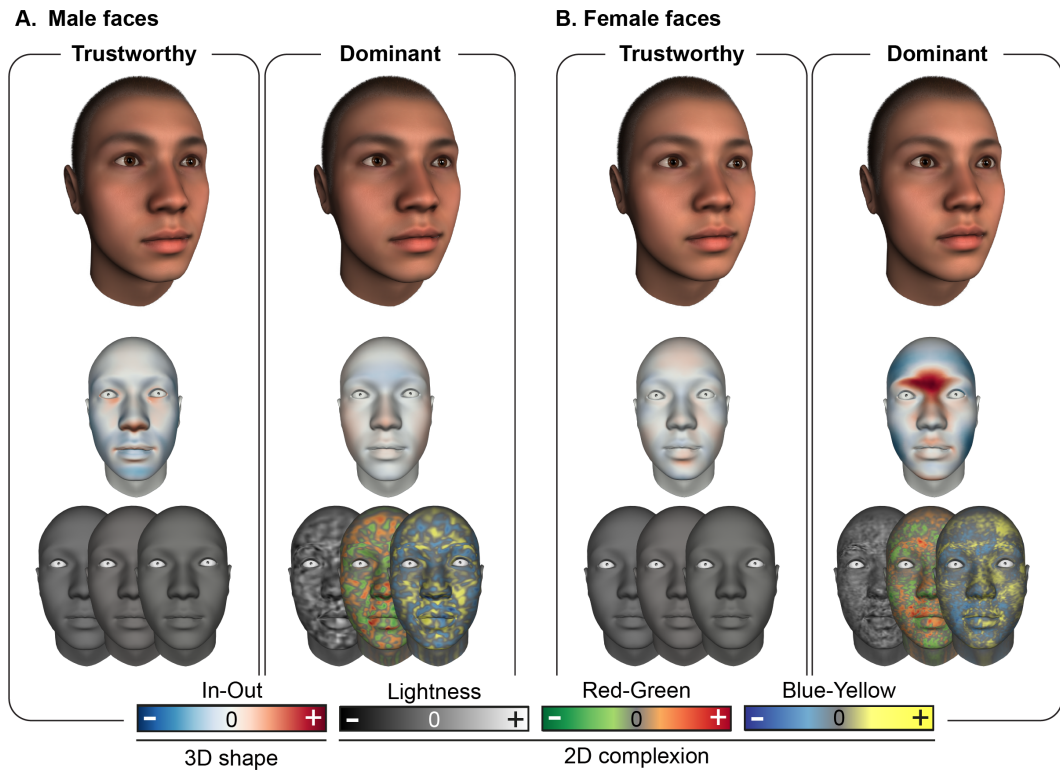


Figure 5.10: For each stimulus sex (A: Male faces; B: Female faces), faces at the top show the facial features associated with trustworthy (left) and dominant perceptions (right) that are shared across observer cultures and face ethnicities, with shape and complexion combined and displayed on the same ethnically neutral face. Below, colour-coded faces show the results separately for 3D shape (top) and 2D complexion (bottom), following the same colour-coding format as before (see colorbars below; see Supplementary materials 7.5.3 for further details).

Faces at the top show the facial features of male (A) and female faces (B) that are perceived as trustworthy (left) and dominant (right) regardless of the ethnicity of the face and the culture of the observer (i.e., shared). Colour-coded faces below show the results separately for 3D shape (top) and 2D complexion (bottom) using the same colour-coding format as before (see colorbars below; see Supplementary materials 7.5.3 for further details).

Results for male faces are relatively subtle. For trustworthy perceptions, the shared features comprise a narrower face with more protruding eyes and no complexion effects. For dominant perceptions, the shared features comprise a wider face, a flatter forehead, and a bigger nose with an uneven skin tone. Results for female faces are similar to those for male faces for trustworthy perceptions but differ for dominant perceptions, which are instead driven by a more protruding brow bone and a narrower jaw together with a similarly uneven skin tone. Together, these results show that some facial features, although subtle, drive perceptions of

trustworthiness and dominance regardless of the ethnicity of the face and the culture of the observer.

5.3.6 Examining the relative contributions of face ethnicity, observer culture, and their combination to social trait perception

Consistent with previous findings (e.g., Hehman et al., 2017; Xie et al., 2019), these results show that the characteristics of the face, the observer, and their combination all uniquely influence which facial features drive perceptions of trustworthiness and dominance. To better understand their contributions to social trait perception, I first measured the proportion of identity components within each feature set (i.e., culture-specific; face ethnicity-specific, synergistic, shared) out of all the modelled identity components for each social trait and stimulus sex, separately for 3D shape and 2D complexion (results aggregated across the 5 Spatial Frequency bands). Figure 5.11 shows the results as colour-coded barplots.

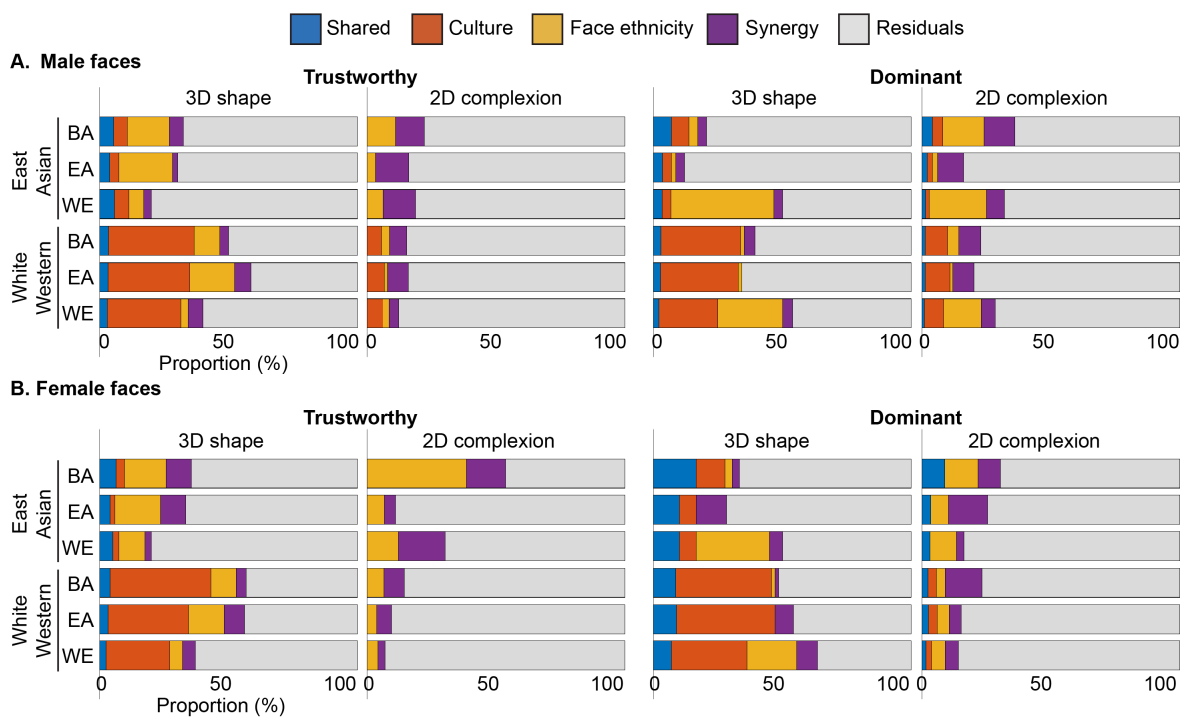


Figure 5.11: In each panel (A: Male faces; B: Female faces), colour-coded barplots show the proportion of trustworthy (left) and dominant (right) facial features of 3D shape and 2D complexion (see titles above) that are shared across observer cultures and face ethnicities (blue colour-coded), those that are culture-specific regardless of face ethnicity (orange colour-coded), those that are face ethnicity-specific regardless of observer culture (yellow colour-coded), and those are synergistic (purple colour-coded) for each combination of observer culture and face ethnicity (see labels on left). Modelled facial features at or above population prevalence threshold not comprised in any feature set (i.e., residuals) are colour-coded in grey.

In each panel (A: Male faces; B: Female faces), colour-coded barplots show the proportion (x axis) of modelled facial features (i.e., identity components) of 3D shape and 2D complexion perceived as trustworthy and dominant (see titles above) that are shared across

observer cultures and face ethnicities (blue colour-coded), those that are culture-specific regardless of face ethnicity (orange colour-coded), those that are face ethnicity-specific regardless of observer culture (yellow colour-coded), and those are synergistic (purple colour-coded) for each combination of observer culture and face ethnicity (see labels on y axis). Modelled facial features which are not part of any feature set but are nonetheless at or above population prevalence threshold ($n = 4$ observers) for a given face ethnicity and observer culture are colour-coded in grey (i.e., residuals).

Results show that the relative contributions (i.e., proportion) of each feature set vary across social traits and stimulus sexes for each combination of observer culture and face ethnicity. This is consistent with previous work (e.g., Hehman et al., 2017; Xie et al., 2019, 2021), which reports that the relative contributions of face, observer, and face $\times$ observer characteristics vary between both face/observer characteristics (e.g., male vs female faces) and social judgements (e.g., attractiveness vs trustworthiness). Nevertheless, results also show that, in total, the feature sets capture on average less than half of the modelled identity components. Specifically, for both stimulus sexes and social traits, the feature sets represent on average 40% (min. = 12%; max. = 64%) of the modelled identity components of 3D shape, and 21% (min. = 7%; max. = 54%) of the modelled identity components of 2D complexion. However, it is important to note that these results are based on the number of identity components that comprise each feature set and their residuals. Within the GMF (see subsection 1.5.2 and Supplementary materials 7.1.1 for details), identity components are Principal Components (PCs). Therefore, by their very nature, identity components are weighted and do not equally capture variance in facial appearance. In other words, while the proportions above provide insight about the distribution of each feature set within the models, it may obscure the extent to which they approximate the appearance of the original modelled features.

To this end, I next compared the vertex representation of the feature sets of 3D shape and the pixel representation of the feature set of 2D complexion to those of the original models. To do this, I first aggregated the identity components of each feature set (i.e., culture-specific; face ethnicity-specific, synergistic, shared) of 3D shape and 2D complexion and weighted each by the proportion of observers for whom it was significant, separately for each combination of face ethnicity and observer culture for each social trait and stimulus sex. In other words, I sought to reconstruct the original models using the feature sets derived above. Figure 5.12 shows a step-by-step example of this reconstruction process for EA male faces perceived as trustworthy by White Western observers as in Figure 5.5.

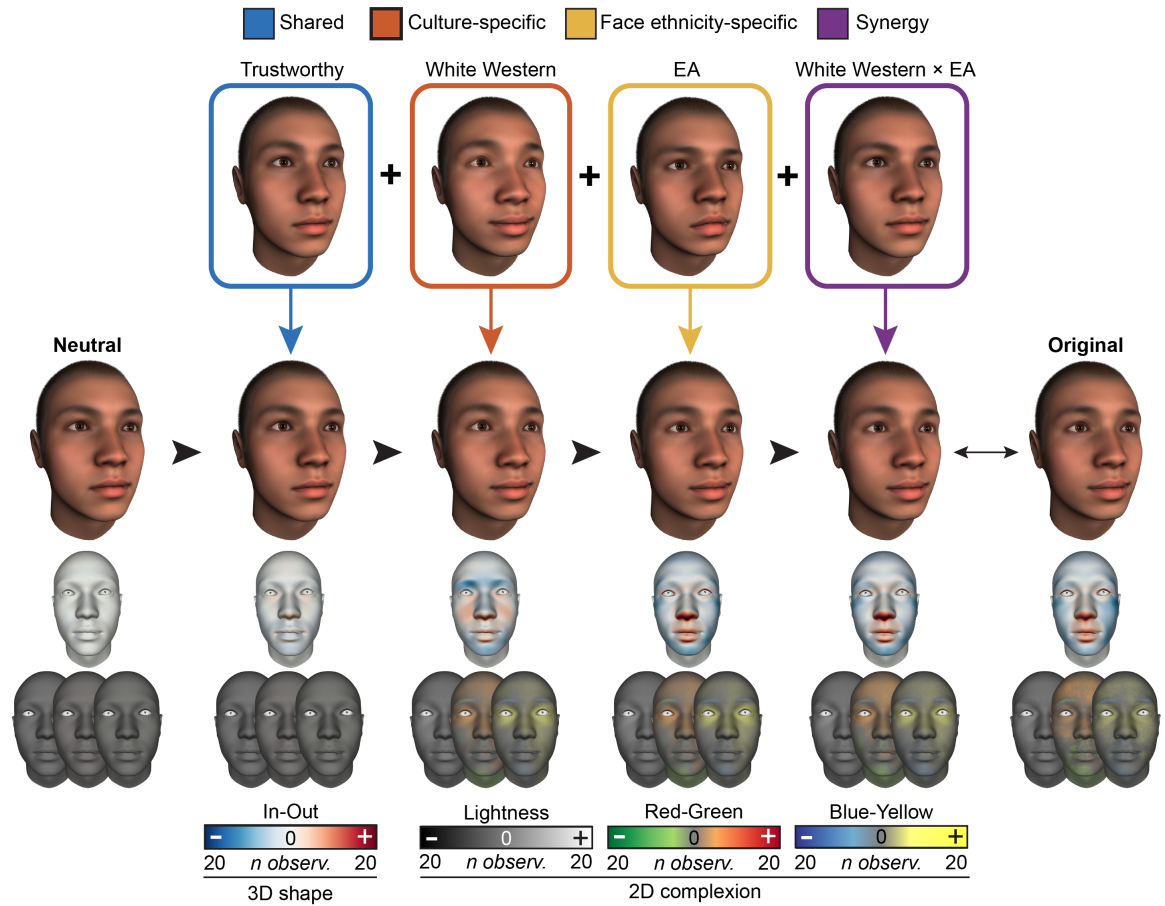
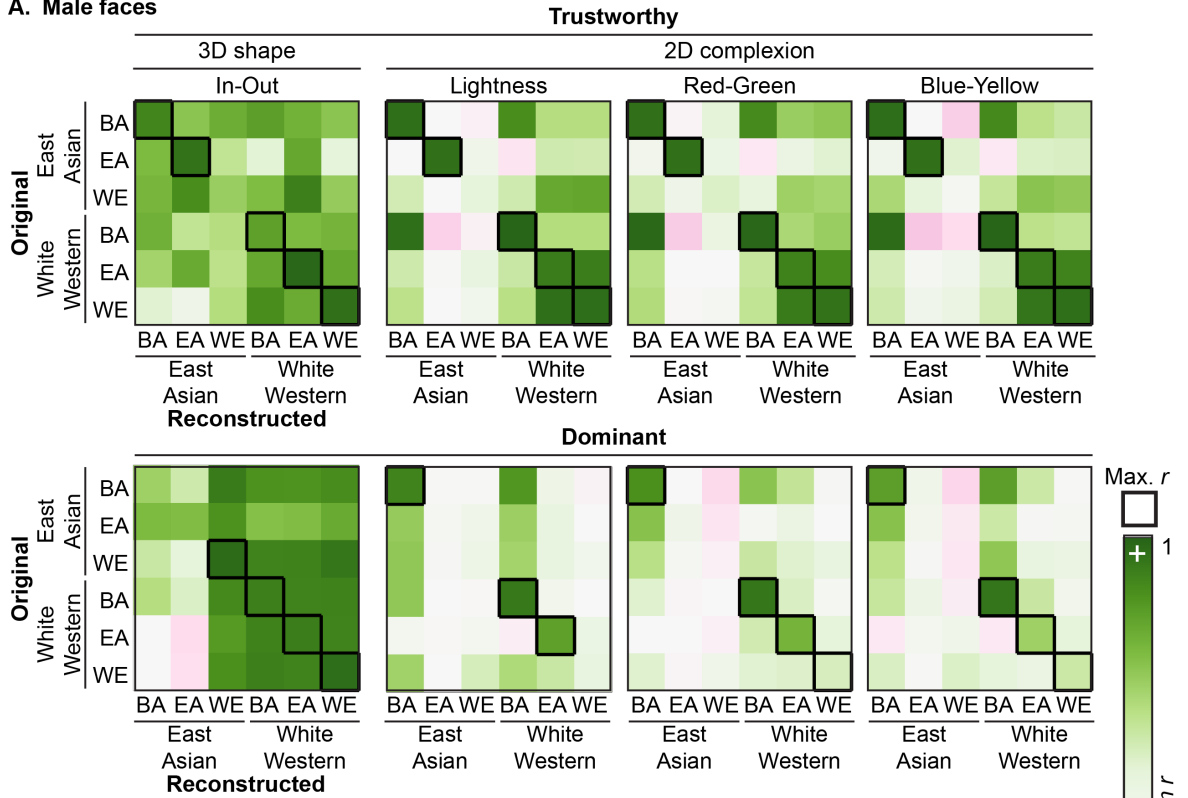


Figure 5.12: Faces at the top show the facial features that are shared (blue colour-coded), culture-specific (orange colour-coded), face ethnicity-specific (yellow colour-coded), and synergistic (purple colour-coded) for EA male faces perceived as trustworthy by White Western observers (see legend at top). Larger faces below show the results of progressively adding each feature set to the ethnically neutral average faces (leftmost face) to reconstruct the original model (rightmost face). Colour-coded faces below show the results for 3D shape and 2D complexion separately, using the same colour-coding format as before (see colorbars below).

Having reconstructed all the models using the approach described above and illustrated in Figure 5.12, I next represented each of them as vertex deviations from the average ethnically neutral face for 3D shape ($N = 14,319$ front-face vertices), and as pixel deviations in each L^*a^*b channel for 2D complexion ($N = 61,218$ downsampled pixels). Similarly, I represented the original models as vertex and pixel representations, aggregated across observers and weighted by the proportion of observers for whom each identity component was significant. Finally, I correlated (Pearson r) the vertex and pixel representations of the reconstructed models to those representing the original models, separately for each social trait and stimulus sex. Figure 5.13 shows the results.

A. Male faces



B. Female faces

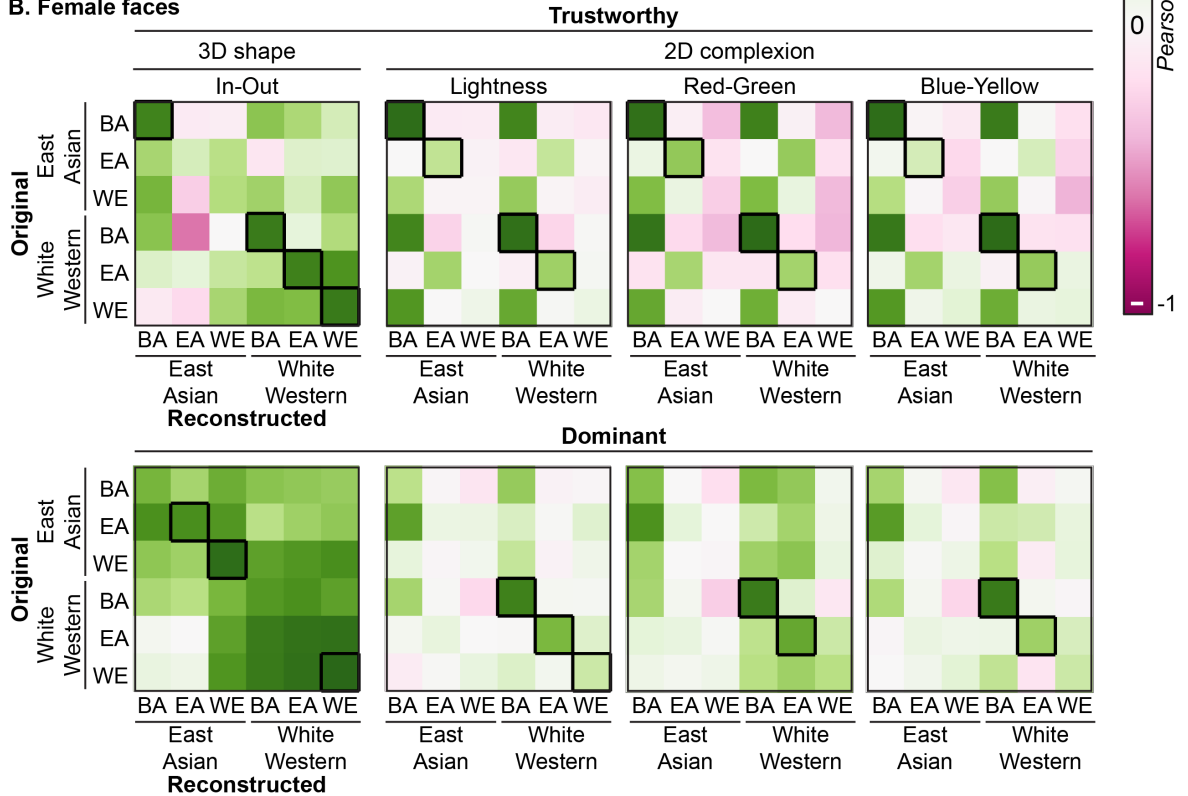


Figure 5.13: In each panel (A: Male faces; B: Female faces), colour-coded show the Pearson correlation between the original (y axis) and reconstructed models (x axis) for trustworthy (top) and dominant perceptions (bottom), separately for 3D shape and each L*a*b channel of 2D complexion (see titles above). Colour-coding indicates the correlation direction and magnitude (see colorbar on right). Thick black outlines highlight statistically significant ($p < 0.05$; Bonferroni-Holm corrected) positive correlations along the diagonal that are larger than the correlation of the original model with any other reconstructed model.

In each panel (A: Male faces; B: Female faces), colour-coded matrices show the Pearson correlation between the original models (y axis) and the reconstructed models (x axis) of trustworthy-looking (top) and dominant-looking faces (bottom), separately for 3D shape (leftmost matrix) and each L*a*b channel of 2D complexion (rightmost matrices; see titles above). Colour-coding indicates the correlation direction and magnitude (see colorbar on right). Thick black outlines highlight statistically significant ($p < 0.05$; Bonferroni-Holm corrected) positive correlations along the diagonal that are larger than the correlation of the original model with any other reconstructed model.

Results show that, for both social traits and stimulus sexes, most of the original models statistically significantly positively correlate with their reconstruction, and do so more strongly than with other reconstructions (see thick black outlines along the diagonal). For example, across 3D shape and 2D complexion, all the original models for trustworthy male faces most strongly positively correlate with their respective reconstruction aside from the models of East Asian observers perceiving WE faces. These effects are however on average weaker for 2D complexion than for 3D shape (see hue variations between matrices). Finally, though overall similar, results for female faces (B) are slightly weaker and less consistent than for male faces (A). For example, relative to male faces, the reconstructed models of trustworthy female faces approximate their corresponding original models less precisely (i.e., correlate less strongly than with other reconstructions). Nevertheless, taken together these results show that, despite comprising a lower proportion of modelled identity components, the features sets identified above (i.e., culture-specific; face ethnicity-specific, synergistic, shared) capture enough facial feature variance to closely approximate the appearance of the original models.

5.4 Discussion

In this final chapter, I aimed to extend current efforts to quantify the relative contributions of face, observer, and face $\times$ observer characteristics to social trait perception (e.g., Hehman et al., 2017; Xie et al., 2019) by examining whether and how the facial features that drive perceptions of trustworthiness and dominance vary only according to the ethnicity of the face, only according to the culture of the observer, and only according to their unique combinations. To do this, I formally compared the ethnically diverse social trait models of East Asian (see Chapter 3) and White Western observers (see Chapter 2) using two complementary approaches. Specifically, I first trained an SVM to show that the modelled facial features are sufficiently distinct to enable correct classification of both the ethnicity of the face and the culture of the observer. I then used a combination of information-theoretic analyses to identify which facial features underpin this distinctiveness—that is, which facial features are culture-specific regardless of face ethnicity, which are face ethnicity-specific regardless of observer culture, which are specific to each combination of face ethnicity and observer culture (i.e., synergy), and which are shared across face ethnicities and observer cultures. Each feature set comprised unique variations in facial appearance, suggesting that

the characteristics of the face (here, face ethnicity), the observer (here, observer culture), and their interaction all uniquely influence the causal facial features used for social trait perception. Subsequent analyses of these feature sets showed that, while they only represent a portion of the modelled facial features, they still capture enough facial feature variance to closely mirror (i.e., positively correlate with) the original models. I discuss the contributions of these results below.

Previous work has shown that variance in social trait ratings is uniquely explained by the characteristics of the face, the observer, and their interaction (e.g., Hehman et al., 2017; Xie et al., 2019). My results extend these findings to the features driving these effects. Specifically, I show that social trait perception is driven by a subtle set of facial features which are shared across face ethnicities and observer cultures plus those that vary only according to the characteristics of the face (i.e., face ethnicity-specific), those that vary only according to the characteristics of the observer (i.e., culture-specific), and those that vary according to their combination (i.e., synergy). In doing so, these findings provide a mechanistic explanation of *how* face, observer, and face $\times$ observer characteristics influence social trait judgements—a key step in advancing understanding of social perception (see Schyns et al., 2022, for further discussion). More broadly, these results support current theoretical perspectives (e.g., Freeman & Ambady, 2011; Over et al., 2020, see also subsections 1.2.5 and 1.2.6) by showing that top-down (i.e., observer culture) and bottom-up (i.e., face ethnicity) sources of information combine and interact to causally influence which facial features are used for social trait perception. Together, these findings therefore complement recent efforts to quantify the relative contributions of face, observer, and face $\times$ observer characteristics and provide empirical support for modern theories of social perception.

Subsequent analyses of the feature sets show that they closely resemble the appearance of the original models, and that this similarity is often strongest for corresponding reconstructed-original model pairs. This suggests not only that the feature sets represent key sources of variance in the causal features used for social trait perception, but also that they sufficiently preserve differences across face ethnicities and observer cultures. Notably, this remains true despite the feature sets account for considerably less variance in the modelled facial features (on average less than half) than what previous work has accounted for in ratings variance (e.g., up to 75% Hehman et al., 2017). One simple explanation for this is that, relative to ratings, variance in facial appearance is high-dimensional and thus more complex to represent (see Jack & Schyns, 2017, for further discussion). Additionally, the analytical approach taken in this chapter is quite conservative. Specifically, I only consider facial features that meet the population prevalence threshold ($n = 4$ observers) for each relevant characteristic (e.g., culture-specific features must be at or above population prevalence in each face ethnicity). While a more liberal or otherwise diversified approach (e.g., culture-specific features in two out of three face ethnicities) would likely capture more variance in the modelled facial features, it would not necessarily contribute to current knowledge in more meaningful ways. Indeed, the current framework still enables conceptual replication of previous ratings-based work (e.g., Hehman et al., 2017; Xie et al., 2019).

Many of the features comprised in each set can also be explained in line with previous findings or otherwise identifiable mechanisms, which further validates the present feature-based approach. Specifically, as discussed in further detail in Chapter 4, the culture-specific features reflect differences in the use of emotion cues (e.g., Adams et al., 2012; Albohn & Adams, 2021b; Said et al., 2009). Some—though not all—of the face ethnicity-specific features also appear to reflect previous work. For example, across social traits and stimulus sexes, the BA-specific features comprise clear complexion effects (e.g., lighter vs darker eye regions) that may serve to alter the appearance of shape features (e.g., bigger vs smaller eyes) in line with shape from shading visual mechanisms (e.g., see Kleffner & Ramachandran, 1992). Similarly, across both stimulus sexes, the WE-specific dominant features comprise anger-resembling cues (e.g., angled eyebrows; e.g., Oosterhof & Todorov, 2008; Said et al., 2009), though these may serve different purposes in each culture. As discussed in Chapter 2, in a Western context these features may reflect the need for a higher perceptual threshold to perceive same-ethnicity faces as dominant (Hugenberg, 2005). In contrast, East Asian observers perceive negative emotions with higher intensity from White vs East Asian faces (Q. Wang et al., 2014) and may thus expect clearer anger cues in WE faces. Finally, the shared features—though subtle—also reflect previous findings, such as the role of softer features in perceptions of trustworthiness (e.g., Oosterhof & Todorov, 2008; Todorov & Oosterhof, 2011) and of facial width in perceptions of dominance (e.g., Albert et al., 2021; V. R. Mileva et al., 2014).

In contrast, the synergistic features are harder to explain. Previous work has defined face $\times$ observer interactions as "personal taste" (Hönekopp, 2006) or more general idiosyncrasies (Hehman et al., 2017). Though this could apply to my results, the synergistic features were derived at the group level (i.e., across observers) and are therefore less likely to capture individual differences. An alternative explanation is to think of these features as serving a similar purpose to facial expression accents (e.g., Elfenbein & Ambady, 2002; Jack et al., 2016; Marsh et al., 2003). That is, rather than fundamentally altering perception, these features may serve to subtly distinguish between combinations of face ethnicity and observer culture. Nevertheless, this explanation remains tentative and future work should seek to further examine and explain these synergistic effects.

Despite adding to growing efforts to understand the relative contributions of face, observer, and face $\times$ observer characteristics, the current findings focus on the influence of face ethnicity and observer culture only. As other face characteristics such sex (e.g., Sutherland et al., 2015) and age (e.g., S. Y. Ng et al., 2016) are known to influence social trait judgements, including interactively (Xie et al., 2019), future work should examine whether and how they affect the causal facial features. Similarly, future work should consider additional characteristics of the observer, such as their ethnicity (e.g., Kawakami et al., 2022), and stereotype (e.g., Fiske, 2017; Xie et al., 2021) or otherwise conceptual knowledge (e.g., Stoler et al., 2018). Finally, as previous work shows that different social messages (e.g., attractiveness) are differentially affected by the face, the observer, and their interaction (Hehman et al., 2017), future work should broaden investigations to additional social domains.

In summary, this chapter demonstrates that social trait perception is driven by distinct facial features that are influenced independently by the characteristics of the face (i.e., face ethnicity), the characteristics of the observer (i.e., culture), and their interaction. These results extend prior work (e.g., Hehman et al., [2017](#); Xie et al., [2019](#)) by shifting focus from social trait ratings to their causal facial features, offering deeper insight into the complexity of social trait perception (Schyns et al., [2022](#)). In doing so, these findings also support modern theoretical accounts (Freeman & Ambady, [2011](#); Over et al., [2020](#)) by providing a mechanistic account of how top-down and bottom-up sources of information interact in social perception.

Chapter 6

General discussion

In this thesis, I aimed to diversify current knowledge of social trait perception by examining whether and how ethnic variance in the face as well as cultural variance in the observer affect the causal facial features used to perceive trustworthiness and dominance. Across four empirical chapters, I show that face ethnicity, observer culture, and their combination systematically influence which facial features are used for social trait perception. Below, I summarize the implications of my findings for the field, before discussing their limitations and suggesting future lines of inquiry.

6.1 Main findings and contributions

In Chapter 2, I aimed to address the White-centric bias of prominent feature-based models of social trait perception (e.g., Oosterhof & Todorov, 2008; Sutherland et al., 2015; Vernon et al., 2014; see Cook & Over, 2021, for discussion) by modelling the causal facial features that drive perceptions of trustworthiness and dominance from three face ethnicities—Black African (BA), East Asian (EA), and White European (WE)—in White Western observers. Comparisons of the models revealed that social trait perception is driven by a core set of features which are shared across face ethnicities and largely mirror current models and related findings (e.g., V. R. Mileva et al., 2014; Oosterhof & Todorov, 2008; Said et al., 2009; Vernon et al., 2014) plus novel face ethnicity-specific variations. The WE-specific features similarly reflect previous work (e.g., Oosterhof & Todorov, 2008), which in turn suggests that current models are not universal but rather specific to White European faces. Further highlighting the White-centric bias of current knowledge, the features specific to Black African and East Asian faces are largely unrepresented in prominent feature-based models. Subsequent analyses of these face ethnicity-specific features revealed that they modulate ethnic phenotypic appearance for trustworthiness and alter signal strength for dominance, in line with previous work on ethnic stereotyping (e.g., Blair et al., 2002; Hugenberg, 2005; Hutchings et al., 2024; Maddox, 2004). Together, these results challenge the generalizability of prominent feature-based models (e.g., Oosterhof & Todorov, 2008), provide a causal explanation for *how* face ethnicity biases social trait perception (e.g., Blair et al., 2002; Eberhardt et al., 2006; Hutchings et al., 2024; Maddox, 2004; Xie et al., 2021), and motivate the need to

consider ethnic diversity when examining fundamental social behaviours (see Cook & Over, 2021; Henrich et al., 2010; Rad et al., 2018, for discussion)

Chapter 2 focused on the perception of White Western observers to enable direct comparisons with previous work (e.g., Oosterhof & Todorov, 2008; Sutherland et al., 2015; Vernon et al., 2014). Observer culture is another key source of variance in social trait perception that is often overlooked (e.g., B. C. Jones et al., 2021; Maeng et al., 2022; see also Cook et al., 2022; Henrich et al., 2010; Kline et al., 2018, for further discussion). To this end, Chapter 3 examined whether face ethnicity causally influences the facial features East Asian observer use to perceive trustworthiness and dominance. Results showed that, like White Western observers, East Asian observers also perceive social traits from a combination of shared plus face ethnicity-specific features. The shared facial features largely aligned with both prominent models (e.g., Oosterhof & Todorov, 2008) and the previous results for White Western observer, suggesting that certain mechanisms of social trait perception generalize across ethnicities and cultures. However, the face ethnicity-specific features were not represented in prominent models, did not resemble those identified for White Western observers, and did not reflect previous findings on ethnic stereotyping (e.g., Blair et al., 2002; Hugenberg, 2005; Hutchings et al., 2024). Thus, while conceptually replicating the results of Chapter 2, these findings suggest that the ways in which face ethnicity alters social trait perception varies across cultural boundaries, possibly due to observer-related differences in how ethnic stereotypes and in-group biases manifest (e.g., Fiske, 2017; Kawakami et al., 2022).

To more directly examine whether and how the culture of the observer causally influence social trait perception, Chapter 4 formally compared the models of White Western (Chapter 2) and East Asian observers (Chapter 3). Specifically, this chapter sought to address two key questions: are the features that drive social trait perception across face ethnicities universal, as implied in key theories (Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008) and supported by previous ratings-based work (e.g., Albright et al., 1997; Walker et al., 2011; A. Wang et al., 2024; Zebrowitz et al., 2012); and do evolutionarily adaptive emotion cues generalize across face ethnicities and observer cultures (e.g., Adams et al., 2012; Albohn & Adams, 2021b; Jaeger & Jones, 2022; Montepare & Dobish, 2003; Said et al., 2009). Results showed that, while *ratings* of trustworthiness and dominance positively correlate between cultures, the causal features driving these ratings are culturally variant. These findings question previous claims of universality in line with emerging evidence (e.g., Mo et al., 2022; Rostovtseva et al., 2024) and underscore the importance of moving beyond coarse measures of similarity to derive causal explanations of social behaviour (see Schyns et al., 2022, for further discussion). Further analyses of the culture-specific features revealed that only those specific to Western culture resemble specific emotion cues (e.g., smiling, frowning) which in turn correlate with valence-congruent emotions (e.g., trustworthy-happy). This suggests that previous work on emotion overgeneralization captures Western-specific, rather than universal, mechanisms (e.g., Albohn & Adams, 2021b; Jaeger & Jones, 2022; Said et al., 2009; Zebrowitz & Montepare, 2008). In sum, Chapter 4 challenges the cultural universality of social trait perception implied by prominent theories (e.g., Oosterhof & Todorov, 2008; Ze-

browitz & Montepare, 2008) and supporting work (e.g., Albright et al., 1997; Said et al., 2009; Walker et al., 2011) by showing that, akin to face ethnicity, observer culture causally influences how social traits are perceived from ethnically diverse faces.

The previous chapters of this thesis showed that social trait perception is causally influenced by face ethnicity within each culture (Chapters 2 and 3), as well as by culture across face ethnicities (Chapter 4). In line with modern theories (e.g., Freeman & Ambady, 2011; Over et al., 2020), emerging evidence now shows that the characteristics of both the face (here, ethnicity) and the observer (here, culture) interact in perception to influence social trait ratings (e.g., Hehman et al., 2017; Xie et al., 2019). Chapter 5 aimed to complement these findings by examining how the ethnicity of the face and the culture of the observer individually and jointly influence the specific facial features that underpin social trait ratings. Results from an SVM indicated that the facial features driving social trait perception from ethnically diverse faces in culturally diverse observers are sufficiently distinct to enable correct classification. Subsequent information-theoretic analyses revealed that this distinctiveness can be captured by four sets of facial features: those that are shared across face ethnicities and observer cultures, those that are face ethnicity-specific, those that are culture-specific, and those that are specific (i.e., synergistic) to each combination. Further examination of these feature sets confirmed that they accurately represent the original models, thus indicating that they capture key sources of variance in social trait perception. These results therefore complement and extend previous work (e.g., Hehman et al., 2017; Xie et al., 2019) by providing a mechanistic understanding of *how* face, observer, and face $\times$ observer characteristics influence social trait judgements. In doing so, these results also provide direct empirical support for current theoretical perspectives (e.g., Freeman & Ambady, 2011; Over et al., 2020) by showing that bottom-up (i.e., face ethnicity) and top-down (i.e., observer culture) sources of information combine and interact to causally influence social trait perception.

Together, the findings reported in this thesis respond to mounting calls to diversify psychological science (e.g., see Cook & Over, 2021; Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018) by directly examining how ethnic and cultural diversity causally influence the perception of two key social traits—trustworthiness and dominance. By revealing systematic differences in how observers from different cultural backgrounds socially perceive ethnically diverse faces, my results challenge previous claims of universality (e.g., Albright et al., 1997; Walker et al., 2011; A. Wang et al., 2024; Zebrowitz et al., 2012) and question the generalizability of prominent feature-based models (e.g., Oosterhof & Todorov, 2008; Sutherland et al., 2015; Vernon et al., 2014) and leading theories (e.g., Zebrowitz & Montepare, 2008). Further, by combining a data-driven approach with a high-fidelity generative model of human faces, my results enable precise identification and formal analysis of the causal features driving social trait perception—a key step in deriving objective explanations of social behaviour (Schyns et al., 2022). In doing so, my results provide a mechanistic account of *how* diversity influences social trait perception, thereby extending previous work (e.g., Hehman et al., 2017; Hutchings et al., 2024; B. C. Jones et al., 2021; Xie et al., 2021) and offering direct empirical support for modern theories (Freeman & Ambady, 2011; Over

et al., 2020). Finally, my work bridges methods and insights from traditionally separate fields of inquiry, such as social perception (e.g., Oosterhof & Todorov, 2008), ethnic stereotyping (e.g., Maddox, 2004), and vision science (e.g., Mangini & Biederman, 2004), thereby highlighting the value of cross-disciplinary approaches in deriving cohesive accounts of how diversity impacts complex social behaviours.

6.2 Limitations

6.2.1 Western-centric social trait dimensions

In this thesis, I focused on perceptions of trustworthiness and dominance to enable direct comparisons with the large majority of extant work (e.g., Hutchings et al., 2024; Jaeger et al., 2020; Oosterhof & Todorov, 2008; Said et al., 2009; Vernon et al., 2014; A. Wang et al., 2024). Although highly influential, growing evidence challenges the universality of these dimensions. Specifically, replications of the dominance-trustworthiness model have only been successful in Western contexts (e.g., Morrison et al., 2017; H. Wang et al., 2016). When cultural diversity is considered, this two-dimensional solution does not readily emerge (B. C. Jones et al., 2021) and is instead supplemented with additional dimensions (e.g., attractiveness/youthfulness, masculinity/femininity; e.g., see Lin et al., 2021; Sutherland et al., 2018). Further, the meaning of these dimensions can vary across cultures. For example, while trustworthiness appears to be relatively cross-cultural (e.g., B. C. Jones et al., 2021), the "opposing" dimension (e.g., dominance) tends to reflect physical strength in a Western context (e.g., Oosterhof & Todorov, 2008), whereas in East Asians it captures intellect (e.g., Sutherland et al., 2018; H. Wang et al., 2019). Finally, while masculinity tends to be a component part of dominance in Western culture (e.g., Albert et al., 2021; V. R. Mileva et al., 2014), it is a dimension in its own right in East Asian culture (Lin et al., 2021). Therefore, while this thesis still adds to growing efforts to diversify knowledge of social trait perception by showing that ethnic and cultural variance do causally affect perception of key social traits (regardless of their fundamental status), it is nevertheless not wholly free of Western-centric biases.

6.2.2 Limited number of ethnic and cultural groups

As discussed at length in this thesis, current knowledge of social trait perception is predominantly based on White Western observers perceiving White faces (see Cook & Over, 2021, for further discussion). Here, I broaden investigations to East Asian observers and Black African and East Asian faces. Although this is a valuable first step in diversifying knowledge, it is also evident that these categories do not comprehensively capture all the ethnic and cultural variance that exists. Relative to ethnicity, studies on social trait perception have rarely—to my knowledge—included further ethnic groups, although some have focused on specific sub-groups such as Buryats (Rostovtseva et al., 2024) and Koreans (Zebrowitz et al., 2010). However, related work on ethnic prototypicality shows that additional

face ethnicities (e.g., Middle Eastern, Latinx) are evaluated differently (e.g., Ma et al., 2018; Norman et al., 2024; Strom et al., 2012). As ethnic prototypicality is known to influence social trait perception (e.g., Blair et al., 2002; Hutchings et al., 2024; Kleider-Offutt et al., 2018; Maddox, 2004, see also Chapter 2), including further ethnic groups in future examinations is a needed next step. Parallel to this, considering the demographic makeup of the country wherein research is conducted may be particularly informative when selecting which face ethnicities to include. This is because two additional known sources of variance in social perception—namely, experience and stereotypes—are likely to vary according to national, ethnic, and cultural context (e.g., Durante et al., 2017; Fiske, 2017; Kawakami et al., 2022). To illustrate, according to the Scottish 2022 census (see Bond, 2025, for details), South Asians (i.e., Pakistani, Indian, Bangladeshi) are the largest minority in Scotland. Including South Asian faces may therefore be more relevant to understanding how ethnic diversity influences social trait perception in a Scottish context.

Relative to culture, most cross-cultural work—including my own—focuses on Western vs East Asian observers (e.g., Maeng et al., 2022; Sutherland et al., 2018; Walker et al., 2011). While this is not inherently negative, there are now mounting calls to move beyond such "East-West" dichotomy (e.g., Kitayama & Salvador, 2024; Vignoles et al., 2016). This is because of several reasons. Most broadly, insofar as East Asian culture is the only (or at least most widely) available comparison to Western culture, cross-cultural knowledge still risks falling into ethnocentric frameworks where the West is compared to "the rest" (e.g., see Kline et al., 2018, for further discussion). This "rest" assumption is in turn challenged by growing evidence of fundamental differences across multiple cultural groups. For example, just as Western and East Asian culture are known to show distinctive cognitive patterns (e.g., individualism vs collectivism), so do Arab, South Asian, and Latin cultures (see Kitayama et al., 2022, for a review). Additionally, the mechanisms underpinning social behaviours vary between these cultural groups—for example, while both Latin and Western people tend to express positive emotions with higher intensity than East Asians, this effect is most pronounced for socially engaging emotions (e.g., comfort) in Latin culture vs socially disengaging emotions (e.g., pride) in Western culture (Salvador et al., 2020). Finally, as discussed previously, the fundamental dimensions underpinning social trait perception vary across cultures (B. C. Jones et al., 2021). Together, this evidence highlights the need to move beyond the common "East-West" dichotomy to better understand how social traits are perceived within and across societies.

6.2.3 Computer-generated vs real faces

The findings presented in this thesis reflect perceptions of computer-generated faces, as is common in the field (e.g., Hutchings et al., 2024; Oosterhof & Todorov, 2008; Peterson et al., 2022; see also Dawel et al., 2022, for a review). Although the GMF has high fidelity and expressivity (see Yu et al., 2012; Zhan, Garrod, et al., 2019, for details, see subsection 7.1.1 for a direct demonstration), a clear concern is whether my own and previous results based on

computer-generated faces would generalize to real faces. That is, computer-generated faces enable precise control over the facial features examined, which in turn allows for precise identification and formal analyses of their effects (Jack & Schyns, 2017). Further, through standardization and parametric control, computer-generated faces minimize noise. However, noise is in many ways an active component of real-life social perception. For example, makeup (e.g., Etcoff et al., 2011), facial body modifications (e.g., Timming & Perrett, 2016), hairstyles (e.g., Takeda et al., 2006), and head tilt (e.g., Zhang et al., 2020)—all sources of "noise" which we nonetheless encounter on a daily basis—influence which social traits are perceived from faces. Notably, removing noise may be beneficial if not necessary for data-driven investigations such as the present, where agnostically sampling all these sources variance would produce an unfeasibly large number of experimental trials. Nevertheless, assuming that effects based on noise-free stimuli would directly replicate in noisy stimuli is premature.

Beyond noise considerations, growing evidence shows that perceptions of computer-generated faces can differ from those of real faces. For example, previous work has shown that known perceptual effects such as ORE (see Kawakami et al., 2022, for a review) are considerably dampened for computer-generated vs real faces (Crookes et al., 2015). Similarly, using real faces produces more varied social trait dimensional spaces (e.g., Sutherland et al., 2013, 2018) and different face- and observer-level contributions to social trait ratings (Hehman et al., 2017). More broadly, computer-generated faces tend to be evaluated less favourably than real faces (e.g., Balas & Pacella, 2017; Di Natale et al., 2024), possibly due to their lower realism (see Kätsyri et al., 2019, for further discussion). Together, this questions the extent to which findings based on computer-generated faces directly map onto perceptions of real faces. This is not to say that previous work, including my own, does not in any way reflect real-life social perception—for example, classification images of the same social trait dimension derived from real (e.g., Sutherland et al., 2013) and computer-generated faces (e.g., Oosterhof & Todorov, 2008) are in many ways similar. Rather, findings based computer-generated faces should be seen as a valuable tool to reduce the sampling space and refine hypotheses before real-life testing. Nevertheless, ongoing technological advances in computer graphics, artificial intelligence, and virtual reality hold promise to improve the realism, and thus ecological validity, of computer-generated faces.

6.2.4 Lack of context and further considerations for ecological validity

Extending from the above, the findings of this thesis—and again, of most extant work in the field (e.g., Hutchings et al., 2024; Oosterhof & Todorov, 2008; Sutherland et al., 2015, 2018; Vernon et al., 2014; Xie et al., 2021)—are context-free. That is, in each experiment observers were asked to rate disembodied faces shown against a black backdrop according to how trustworthy and dominant they looked with no other information beyond the appearance of the face. Obviously, real-life social perception does not occur in such vacuums. However, context is difficult to define and therefore study. If we take context to mean the situation

wherein a face is socially perceived, previous work has shown that scenes suggesting wealth (e.g., Freeman et al., 2011; Keres & Chartier, 2016), threat (e.g., Brambilla et al., 2018; Matavelli et al., 2022), and even simple block-colour backgrounds (e.g., N. Chen et al., 2024) all influence social judgements. Alternatively, if context refers to additional knowledge about an individual, descriptions of previous behaviour can alter initial perceptions (e.g., Mende-Siedlecki et al., 2013), though these changes are seldom lasting (e.g., see Jaeger et al., 2019, 2020). As a result, context-free findings may capture what a trustworthy or dominant face should "on average" look like, rather than more targeted and realistic perceptions.

In the same way that faces naturally exist in context, the knowledge each observer brings to bear during social trait perception is also likely context-dependent. To illustrate, trusting a stranger to look after your bag is fundamentally different from trusting a surgeon to perform open-heart surgery on you. As fundamental dimensions, trustworthiness and dominance are necessarily broad concepts. For example, in Oosterhof and Todorov (2008)'s original study, trustworthiness captures a range of traits (e.g., responsible, caring). Similarly, dominance can be interpreted as physical (e.g., Oosterhof & Todorov, 2008) or intellectual prowess (e.g., H. Wang et al., 2019). This breadth of meaning raises the possibility that different observers may frame the task differently, particularly since individuals are likely to vary in how they structure their conceptual knowledge of social traits (e.g., Stoller et al., 2018, 2020; Xie et al., 2021; see also Over et al., 2020). Arguably, because each observer model was cross-validated against the rest of the sample, the findings of this thesis reflect at least a partly shared conceptual space for each observer group. Nevertheless, better contextualizing perceptual tasks is likely to move the field forward.

Finally, it is important to note that task-irrelevant characteristics of a face—for example, ethnic or gendered appearance when judging trustworthiness—can also be considered context (Hess et al., 2009b). As shown in this thesis and previous work (e.g., Hutchings et al., 2024; Kleider-Offutt et al., 2018; Sutherland et al., 2015), "face context" is a powerful source of variance in social perception that is often hard, if not impossible, to disentangle. This does not negate the need to better understand how other contextual cues—such as scenarios, behavioural evidence, and stricter conceptual definitions—further nuance social trait perception, but does highlight the complexity of doing so.

6.3 Future directions

6.3.1 Examining other facets of diversity and intersectionality

As discussed in subsection 6.2.2, current diversification efforts remain focused on few ethnic and cultural groups. Beyond the need for future work to examine different ethnicities and cultures, another important line of inquiry is understanding how other key factors of human diversity, such as age and gender, combine and interact with ethnicity and culture to influence social trait perception. Specifically, previous work has shown that both gender (e.g., Oh, Dotsch, et al., 2020; Sutherland et al., 2015) and age (e.g., Hess et al., 2023;

Zebrowitz & McDonald, 1991) influence how social traits are perceived. Indeed, masculine vs feminine and baby-like vs mature features are known sources of overgeneralization effects (see Zebrowitz & Montepare, 2008, for a review). As is common, however, this works only attests to how White faces of different genders and ages are perceived. Taking an intersectional approach to understanding social trait perception is not only generally valuable for knowledge diversification but also timely, mirroring growing efforts to understand how intersectional identities are categorized and stereotyped (e.g., see Nicolas et al., 2017; Petsko & Bodenhausen, 2020, for reviews).

Further motivating this, previous work on emotion perception has shown that gender and ethnicity interact to influence the processing speed of positive vs negative emotions in stereotype-consistent ways (Craig, Koch, & Lipp, 2017; Lipp et al., 2015), and that this occurs cross-culturally (Craig, Zhang, & Lipp, 2017). These intersectional effects have also been previously reported for social trait perception (Xie et al., 2021), although these findings focus on ratings rather than causal features. Notably, my own work shows that perceptions of male and female faces of different ethnicities are driven by similar, though not identical, features both in Western and East Asian culture. Although I do not examine these differences directly (i.e., using analyses that formally test these differences), my results nonetheless evidence that multiple social categories interact in perception to causally influence social judgements. As intersectionality is at the core of many modern theories of social perception (e.g., Adams & Kveraga, 2015; Freeman & Ambady, 2011), understanding these effects more deeply holds promise to not only further diversify current knowledge but also inform theoretical perspectives.

6.3.2 Diversity in multimodal social perception

Traditionally, research on social trait perception has focused on judgements of neutral faces (e.g., Hutchings et al., 2024; Oosterhof & Todorov, 2008; Said et al., 2009; Vernon et al., 2014; Zebrowitz & Montepare, 2008). More recently, this focus has broadened to other sources of information, including facial expressions (e.g., Gill et al., 2014; Hensel et al., 2021), body shape (e.g., Hu et al., 2018; Wildman & Ramsey, 2022), and the voice (e.g., McAleer et al., 2014; Wood et al., 2017). However, much like traditional research, these new lines of inquiry are primarily centred on White Western individuals. Beyond more general concerns about the generalizability of White- and Western-centric research (e.g., Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018), evidence suggests that diversity influences social perception across modalities. For example, facial expressions of emotion are interpreted differently depending on the ethnicity (e.g., Hugenberg, 2005), gender (e.g., Hess et al., 2009b), and age (e.g., Freudenberg et al., 2015) of the expresser as well as the culture of the observer (e.g., Jack, 2013), suggesting that facial expressions of social traits may be similarly affected. Similarly, emotional body postures of men vs women are processed differently (Bijlstra et al., 2019), and Black vs White men's bodies tend to be perceived as larger and more threatening by prejudiced White observers (Wilson

et al., 2017). This again highlights the likelihood that social trait judgements inferred from body shape are nuanced by diversity. Finally, local vs foreign accents influence perceptions of competence and warmth in both Western (Hansen et al., 2017) and East Asian culture (M. H. S. Ng & Cheng, 2024). Together, this evidence indicates that diversity likely plays an influential role in how social traits are perceived from different sources of information. As research continues to move toward multimodal investigations, directly examining how diversity influences different modalities is a key next step.

6.3.3 Designing ethnically and culturally sensitive virtual agents

Virtual agents are fast becoming part of mainstream human society, including in homes, schools, hospitals, museums, and shopping centres to perform a variety of tasks (e.g., Dai et al., 2022; Gasteiger et al., 2021; Pereira Santos et al., 2023). Many state-of-the-art agents have impressive human-likeness, which is known to enhance user trust, engagement, and overall experience (e.g., Baylor, 2009; Bickmore & Cassell, 2001; Shamekhi et al., 2018). However, these technical advances disproportionately benefit a narrow demographic—namely, White individuals from WEIRD societies—because they are designed using psychological insights grounded in White- and Western-centric frameworks. For example, digital agents often resemble White phenotypes (e.g., see Cave & Dihal, 2020, for a review) and reflect White and Western-centric expectations and ideals of important social traits, such as trustworthiness (e.g., Song & Luximon, 2024). Virtual agents' biased design therefore poses significant risks to equity in technology access and benefit.

Designing virtual agents using psychological insights that account for human diversity—such as those presented in this thesis—could alleviate these risks. For example, previous work has shown that user perceptions of likeability are influenced by the agent's perceived ethnicity (e.g., Baylor & Kim, 2004; Davis et al., 2023; Zhao et al., 2024). To mitigate this ethnic bias, agent design could integrate facial features associated with culturally and ethnically sensitive perceptions of trustworthiness. This could improve the effectiveness of ethnically diverse agents, in turn addressing representativeness and thus potentially fostering a sense of identification among user from under-represented groups (e.g., Rosenberg-Kima et al., 2008). Embedding ethnic and cultural diversity into agent design in this or similar ways could support more equitable user engagement and reduce disparities in the societal benefits of these technologies.

Additionally, virtual agents hold promise as tools for addressing real-life biases and prejudice because they can evoke similar cognitive and behavioural responses as their human counterparts. For example, White-appearing agents are treated with more empathy in medical simulations than Black-appearing agents (Rossen et al., 2008). These findings suggest that virtual agents could be used to challenge implicit biases when incorporated into targeted intervention programs. To illustrate, modern workplace environments are increasingly diverse. While some evidence suggests that interracial contact alone can help mitigate existing biases (e.g., Darling-Hammond et al., 2021), a considerable body of work underscores the

persistence of implicit racial bias in professional settings (e.g., Holder et al., 2015; Triana et al., 2015). These findings highlight the need for more targeted and effective interventions (Metinyurt et al., 2021). To this end, ethnically diverse agents could be programmed to not only appear trustworthy—based on culturally and ethnically sensitive results such as those reported in this thesis—but also demonstrate these traits through meaningful actions, such as participating constructively to brainstorming sessions (e.g., Nomura et al., 2024). Additionally, interventions could include agents designed to reflect users’ own biased perceptions of ethnically diverse individuals—such as the association between ethnic phenotypes and trustworthiness (e.g., Blair et al., 2002; Hutchings et al., 2024, see also Chapter 2)—to offer a safe and controlled environment in which users can confront and reassess their implicit stereotypes. This strategy aligns with recent successes in reducing facial stereotyping of White faces through similar training efforts (Chua & Freeman, 2021).

Together, designing virtual agents with ethnic and cultural diversity in mind holds promise to not only address access inequity but also tackle real-life bias and prejudice. Future work should seek to realize, test, and expand upon the above suggestions.

6.4 Conclusion

In sum, this thesis addresses the White- and Western-centric bias of current knowledge of social trait perception. Across four empirical chapters, I show that two key factors of human diversity—namely, face ethnicity and observer culture—systematically alter the facial features used to perceive trustworthiness and dominance. These findings challenge the generalizability of prominent feature-based models (e.g., Oosterhof & Todorov, 2008, see Chapters 2 and 3), question previous claims of universality (e.g., A. Wang et al., 2024, see Chapter 4), and extend current efforts to understand how face and observer characteristics interact in perception (e.g., Hehman et al., 2017, see Chapter 5). In doing so, my results support modern theories of social trait perception (e.g., Freeman & Ambady, 2011; Over et al., 2020) and underscore the importance of considering diversity when deriving causal explanations of fundamental social behaviours (Henrich et al., 2010; Kline et al., 2018; Rad et al., 2018).

Chapter 7

Supplementary materials

7.1 Chapter 1

7.1.1 Expressivity of the Generative Model of 3D human Faces

The performance of any generative model hinges on the quality and representativeness of its training database. Given the focus of this thesis on modelling social trait perception from faces of different ethnicities, and the GMF's unbalanced sample of face ethnicities (245 White European, 149 East Asian, 8 Black African), I tested the fidelity and expressivity of the GMF in representing faces of different ethnicities. To do so, I constructed a cross-validation test that evaluates how well the GMF reconstructs left-out faces (i.e., faces not used for training). Generally, on each iteration I rebuilt the GMF using a subset of the total 402 3D faces in the database. I then used the rebuilt GMF to fit (R^2) and reconstruct the left-out faces using the identity components of 3D shape and of the 5 spatial frequency (SF) bands of 2D complexion. I used 5 ranges of identity component coefficients to capture increasing variance, following the standard GMF building procedure (see Yu et al., 2012; Zhan, Ince, et al., 2019, for details). I applied this general procedure using two cross-validation approaches:

1. A k-fold cross-validation approach ($k = 10$) applied to all faces in the database, with a randomly selected subset of 40 faces left out for testing on each fold. I did this to test the overall expressivity of the GMF for faces of all ethnicities.
2. A leave-one-out cross-validation (LOOCV) approach applied only to Black African faces in the database ($n = 8$), with each face left out for testing in turn. I did this to test the expressivity of the GMF for Black African faces specifically.

Figure 7.1 shows the results. Specifically, line plots in Figure 7.1A show the mean fit performance (R^2) for each cross-validation approach (colour-coded, see legend at top), averaged across test faces. Results are shown separately for 3D shape (leftmost plot) and 2D complexion across 5 spatial frequency bands (right plots, ordered from highest to lowest SF), as well as for 5 ranges of identity component coefficients (x axis). Figure 7.1B provides

visual comparisons of the ground truth faces (left) and their reconstructions (right), organized by cross-validation approach (see colour-coded labels on left) and identity component coefficient range (see labels at top).

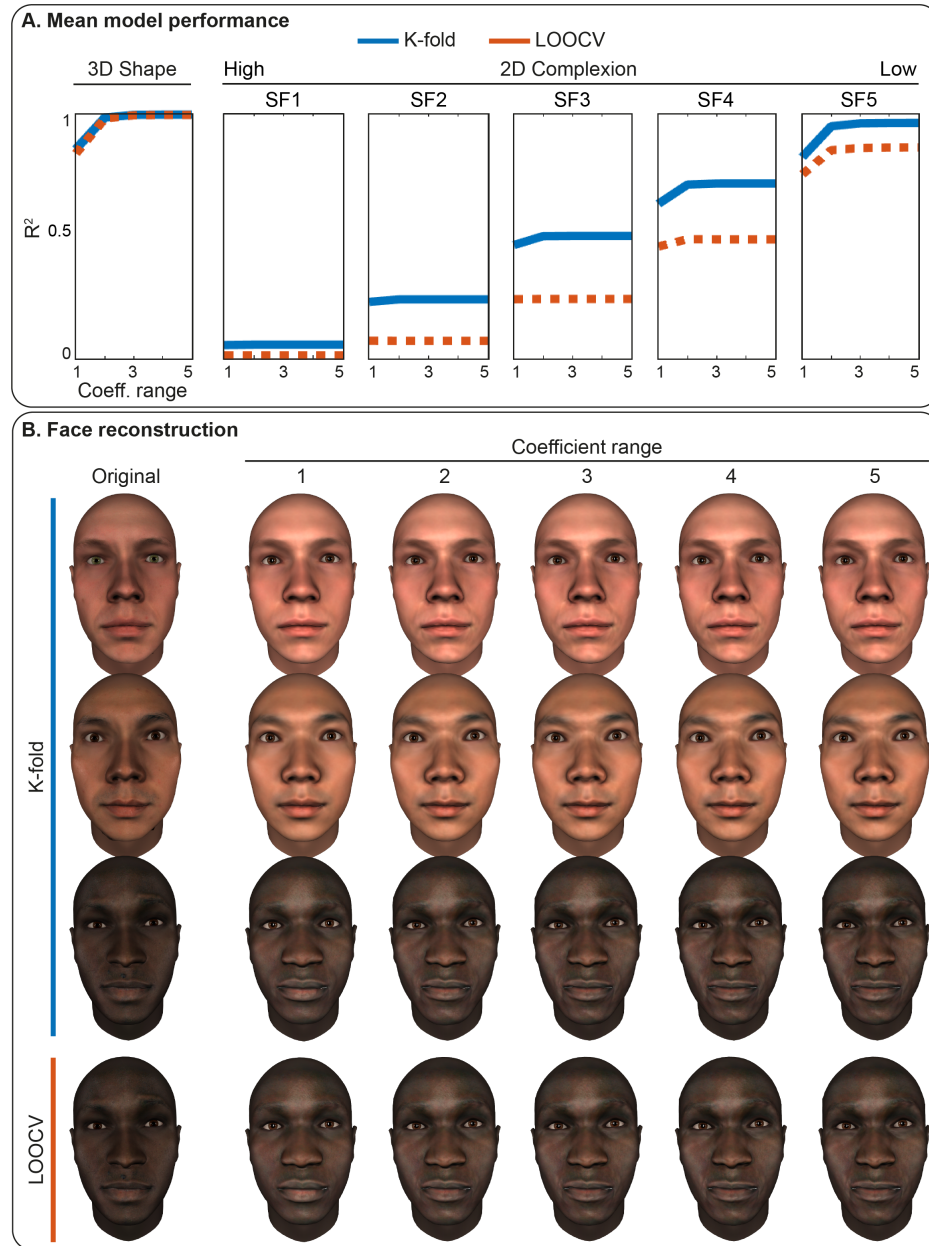


Figure 7.1: **A. Mean model performance.** To assess the expressivity of the GMF (see also Figure 2.1A), I fitted (R^2) ground truth test faces using identity components of 3D shape and of 5 spatial frequency bands of 2D complexion, using two different cross-validation approaches (color-coded, see legend at the top). Fits were performed using 5 ranges (x axis) of identity component coefficients, each capturing increasing variance, with R^2 values averaged across test faces. The line plots present results separately for 3D shape (leftmost plot) and 2D complexion across the 5 spatial frequency bands (right plots, ordered from highest to lowest SF). **B. Face reconstruction.** Rows of faces illustrate the ground truth test face (left) and its reconstructions (right) for each cross-validation approach (color-coded labels) and identity component coefficient range.

For 3D shape, both cross-validation approaches yield similar mean performance across identity component coefficient ranges, indicating a strong fit (ranging from 80% to 99%).

For 2D complexion, performance increases from the highest to the lowest SF. Specifically, R^2 ranges between 1% and 5% at the highest spatial frequency (SF1), and from 75% to 96% at the lowest (SF5), reflecting the type of information each spatial frequency band captures. The highest SF captures fine complexion details (e.g., wrinkles, blemishes), which vary significantly across faces and are more challenging to fit, while the lowest SF captures overall skin tone, which is more stable across faces and particularly important for ethnicity representation (e.g., Farkas et al., 2005; Stepanova & Strube, 2012). Examples in Figure 7.1B demonstrate high fidelity reconstruction (right) of ground truth faces (left) across both cross-validation approaches. Together, these results show that the GMF has high and comparable fidelity and expressivity across face ethnicities. Therefore, throughout this thesis I used the GMF to generate novel face stimuli of all three ethnicities it currently represents—Black African, East Asian, White European.

7.2 Chapter 2

7.2.1 Screening questionnaire for White Western observers

Each White Western observer completed the following screening questionnaire to assess their exposure to and contact with non-Western cultures. I only recruited observers who answered ‘no’ to all questions or who reported minimal exposure (e.g., having visited a non-Western country for a short period of time and not recently).

1. Have you ever lived in a non-Western* country (e.g., on a gap year, summer work, due to parental employment)?
2. Have you every visited a non-Western* country (e.g., on vacation)?
3. Have you ever dated or had a very close friendship with a non-Western* person?
4. Have you ever been involved with any non-Western* culture societies/groups?

*By Western countries/groups/people, we are referring to Europe (Eastern and Western), USA, Canada, United Kingdom, Australia, and New Zealand.

7.2.2 Determining the appropriate number of experimental trials

To determine the number of trials required to obtain stable 3D face models for each social trait, stimulus sex, and face ethnicity, I collected pilot data from 6 White Western observers ($n = 2$ per face ethnicity; 3 female, 3 male, mean age = 23.83, SD = 3.19) using the same experimental procedure described in Chapter 2. I then modelled each pilot observers’ data using ridge regression (I chose this approach over linear regression for time efficiency), separately for each social trait, stimulus sex, and each 3D shape and 2D complexion identity component. I then repeated this modelling procedure using increasing numbers of trials ($n = 10$ to 1,200 in steps of 10), estimating at each iteration the bootstrapped ($n = 100$) 95% upper and lower confidence intervals (CIs) of the identity components’ beta coefficients. Figure 7.2 shows the resulting upper and lower CIs (z-scored; see legend at top) for male (A) and female face models (B) at each iteration (x axis) for each social trait and face ethnicity (see titles at top), separately for 3D shape (top row) and 2D complexion (bottom row). Results are averaged across identity components (402 for 3D shape and 402×5 for 2D complexion) and observers ($n = 2$ per face ethnicity).

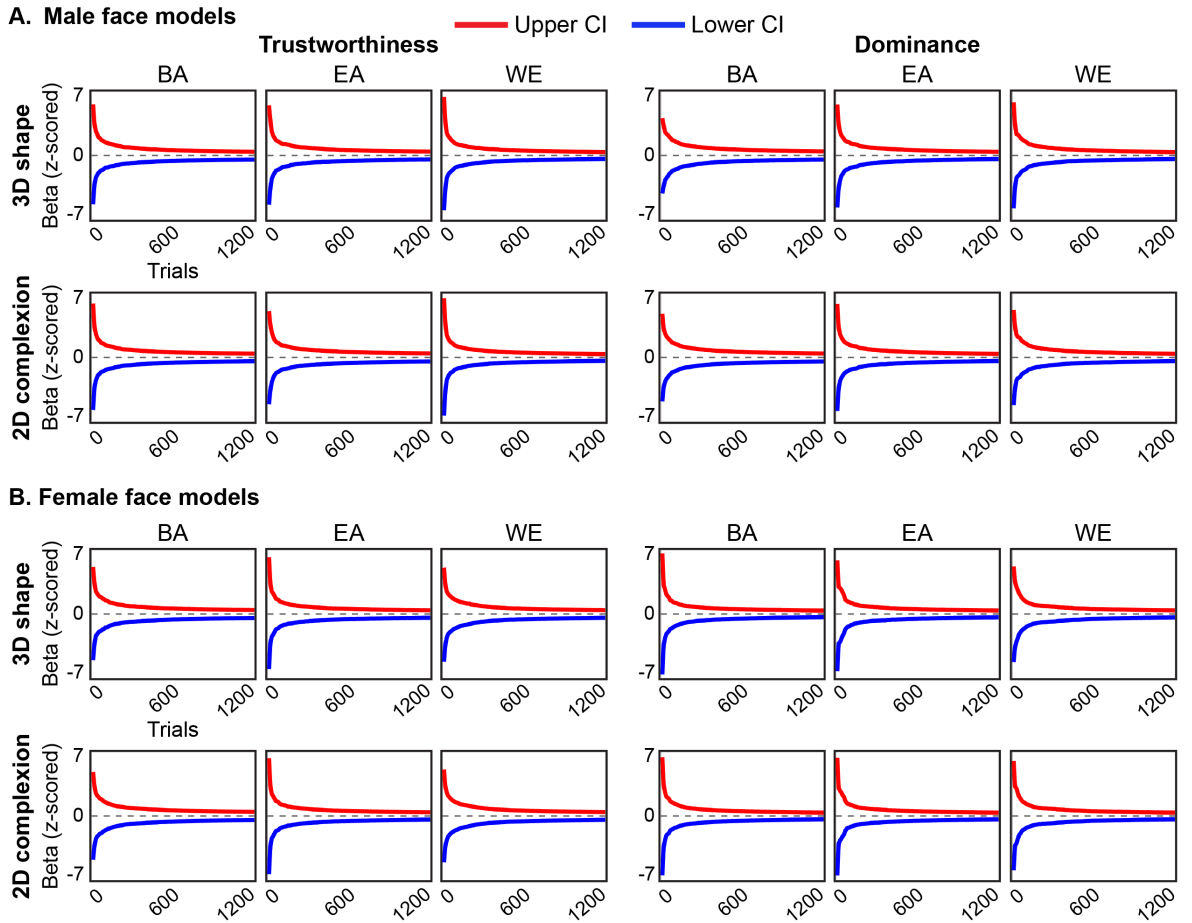


Figure 7.2: In each panel (A: Male face models, B: Female face models), subplots show the bootstrapped ($n = 100$) 95% upper and lower CIs (z-scored; see legend at top) of the models' beta coefficients at increasing trial numbers (x axis), separately for 3D shape (top row) and 2D complexion (bottom row) and for each social trait and face ethnicity (see titles at top). In each subplot, the beta coefficients are averaged across identity components (402 for 3D shape and 402×5 for 2D complexion) and observers ($n = 2$ per face ethnicity)

Results suggest that on average models stabilize around 600 trials (i.e., the CIs do not drastically change beyond this point). However, given the large naturalistic variations of face shape and complexion, particularly across ethnicities (e.g., Farkas et al., 2005; Y. F. Wen et al., 2015), as well as the potentially large individual differences in the causal facial features for social trait perception (e.g., Hehman et al., 2017), and the trial number used by previous similar studies with similar designs (e.g., Bjornsdottir et al., 2024; Hensel et al., 2020; Zhan et al., 2021), I chose to use the full 1,200 trials per social trait and stimulus sex in the final experimental design.

7.2.3 Checking the linearity assumption to model 3D shape and 2D complexion for White Western observers

To ensure that linear regression was an appropriate measure to model the features of 3D shape and 2D complexion that drive social trait perception in White Western observers, I tested the linearity of the relationship between the observers' social trait ratings and the identity component weights of 3D shape and 2D complexion separately. To do so, I proceeded as

follows: for each face ethnicity, stimulus sex, and social trait rating task, I first rebinned each observer's ratings into four rating bins for visualization clarity. I then computed the mean weight of each identity component for each of the four rating bins across observers ($n = 20$ per face ethnicity) and trials ($n = 1,200$ per social trait rating task, stimulus sex). Finally, I represented the resulting mean identity component weights as vertex values in cartesian space ($N = 14,319$ front-face vertices) for 3D shape and pixel values in L^*a^*b space ($N = 61,218$ pixels, downsampled to every third pixel) for 2D complexion, and computed their deviation from the average ethnically neutral face. Figures 7.3 and 7.4 show the results for male and female faces respectively as colour-coded matrices, separately for 3D shape (top row) and each L^*a^*b channel of 2D complexion (bottom rows), for each face ethnicity and social trait rating task (see titles at top). For clarity, each colour-coded matrix is ordered ascendingly according to the values in the first rating bin.

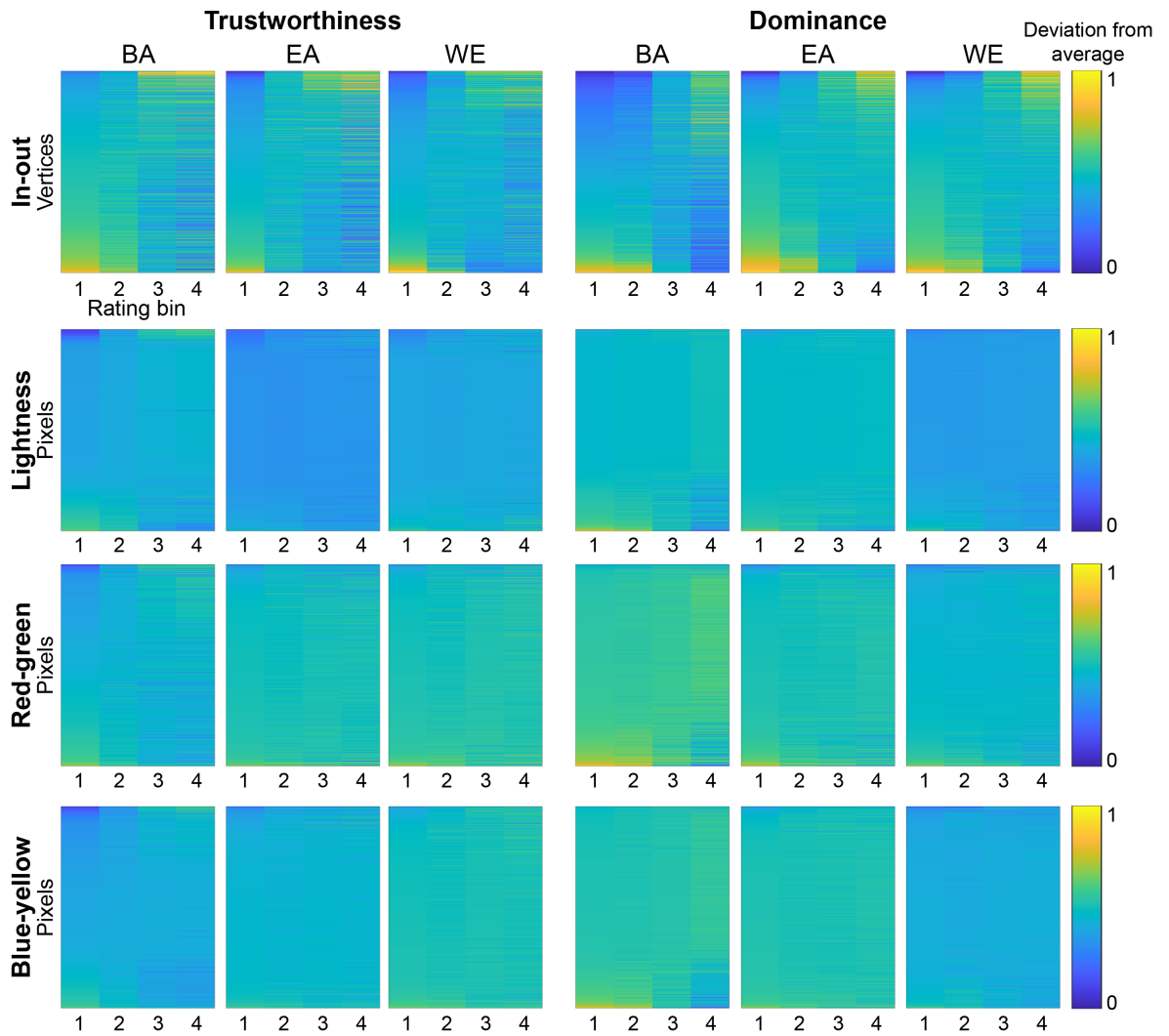


Figure 7.3: For each social trait and face ethnicity (see titles at top), colour-coded matrices show the normalized vertex ($n = 14,319$ front-face vertices; top row) or pixel ($n = 61,218$ downsampled pixels; bottom rows) deviations from the ethnically neutral average face for each of 4 rating bins (x axis) for male faces only, averaged across White Western observers ($n = 20$) and trials ($n = 1,200$). Colour-coding represents the normalized deviation value (see colorbars).

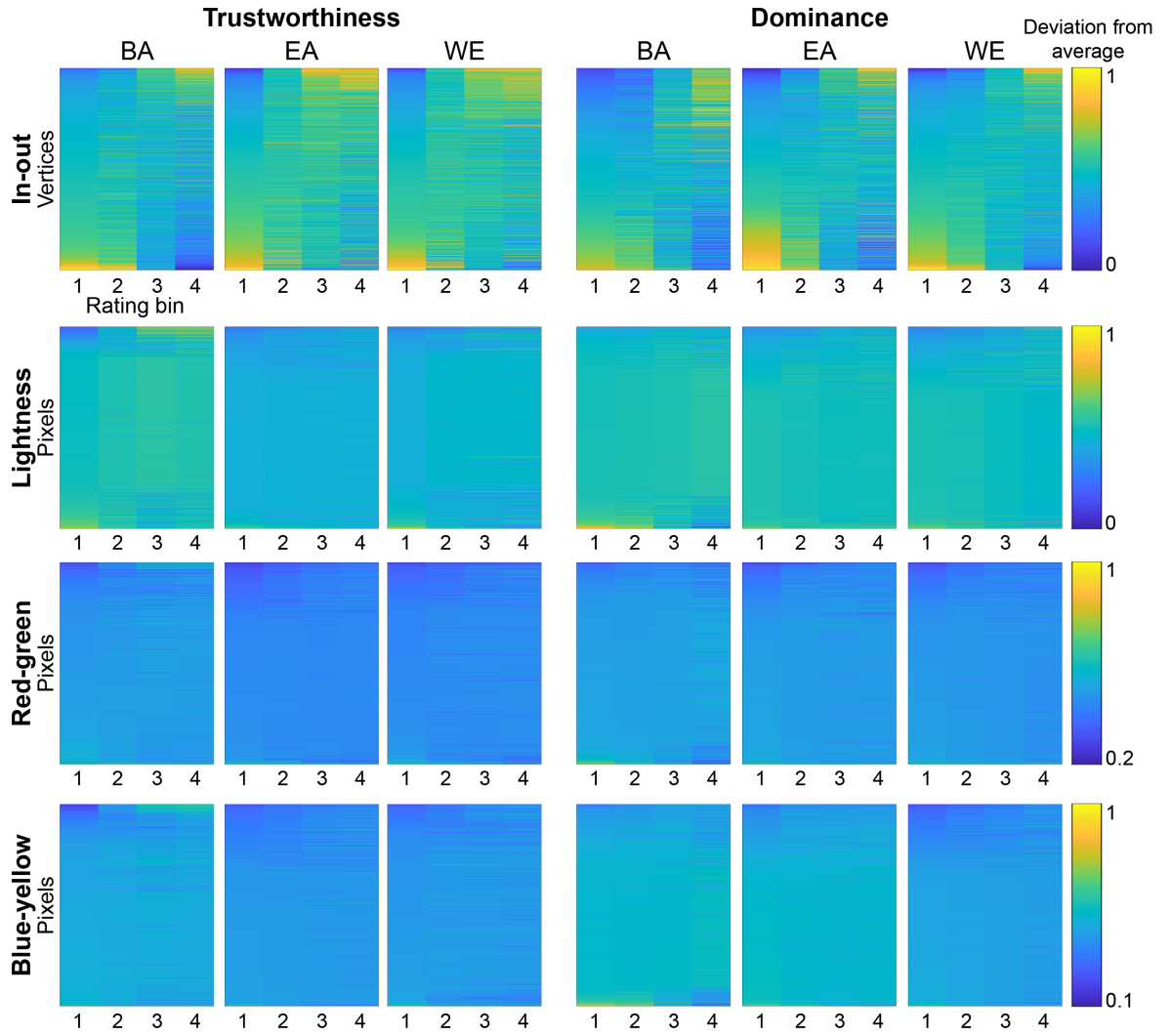


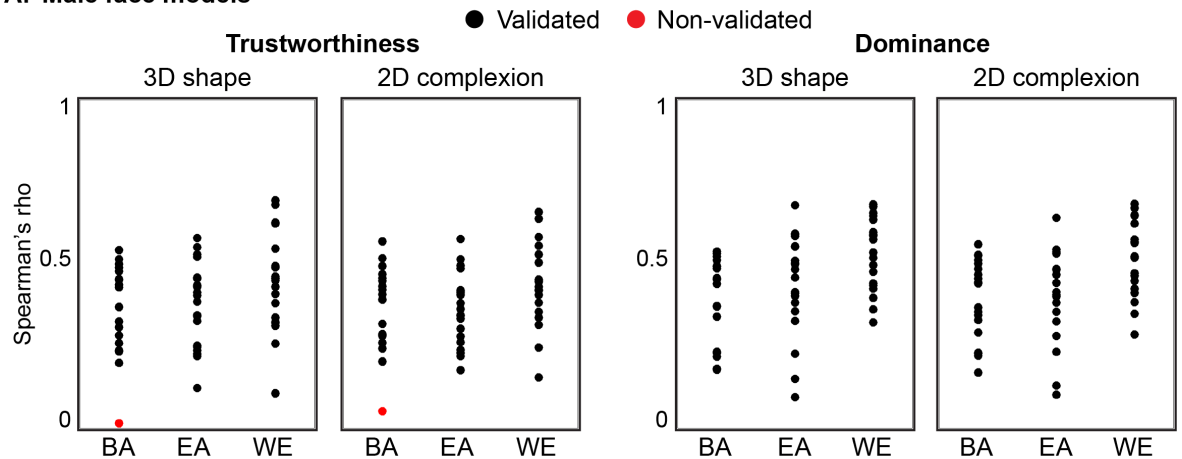
Figure 7.4: Following the same format as Figure 7.3, colour-coded matrices show the normalized deviation values (see colorbars on right) for each of 4 rating bins (x axis), for each social trait and face ethnicity (see titles at top) for female faces only.

In each colour-coded matrix, dark blue denotes a negative deviation from the average and yellow denotes a positive deviation from the average (values normalized; see colorbars on the right). If the relationship between the observers' ratings and the identity component weights were linear, each matrix should show a linear transition from yellow to dark blue (or dark blue to yellow) across the four rating bins (i.e., row-wise). The results for 3D shape shown in the top row of Figures 7.3 and 7.4 align with this pattern, suggesting that the relationship between the observers' ratings and the 3D shape identity component weights is linear. The results for 2D complexion shown in the bottom three rows of Figures 7.3 and 7.4 are overall much weaker, though they do seem to also reflect a linear relationship. Such differences between 3D shape and 2D complexion are not unexpected, given the higher variability in 2D complexion results observed in similar work (e.g., Bjornsdottir et al., 2024; Hensel et al., 2020). Given the above-shown results and these considerations, I deemed linear regression to be an appropriate measure to model the data.

7.2.4 Cross-validating the 3D face models of White Western observers

As described in Chapter 2, I used a leave-one-out cross-validation approach to validate the 3D face models I obtained for each observer, separately for 3D shape and 2D complexion. Figure 7.5 shows the results for male (A) and female face models (B), separately for each social trait rating task, and for 3D shape and 2D complexion (see titles at top).

A. Male face models



B. Female face models

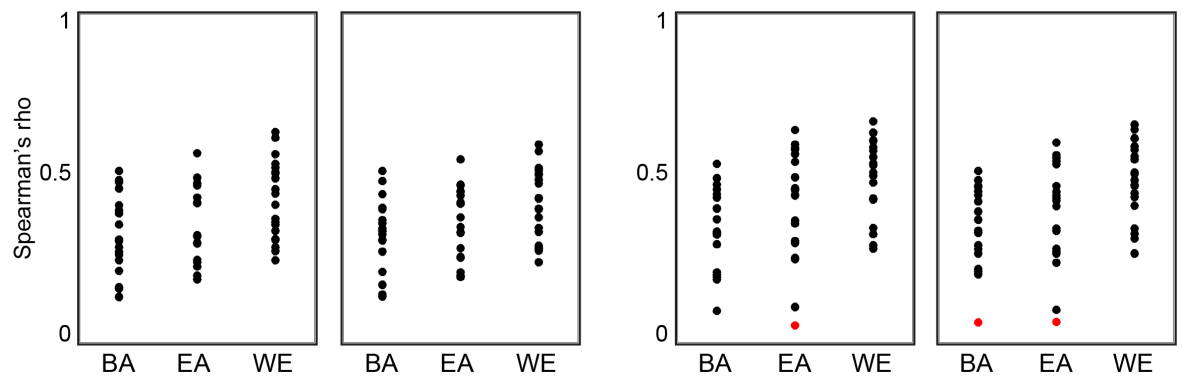


Figure 7.5: In each panel (A: Male face models, B: Female face models), dot plots show the prediction accuracy (measured as Spearman's rho) of each individual observer's social trait models based on a leave-one-out cross-validation analysis, separately for each social trait and for 3D shape and 2D complexion (see titles at top). Black dots denote that the model's prediction accuracy is statistically significantly above chance; red dots indicate it is not (see legend at top). Most of the 240 face models showed cross-observer validity (238 3D shape models; 237 2D complexion models)

In each subplot, colour-coded dots (black = validated, red = non-validated) show the prediction accuracy (Spearman's rho) of each individual observer's model, based on a LOOCV analysis. Results show that most of the models are valid (238 3D shape models; 237 2D complexion models).

7.2.5 Number and distribution of shared vs face ethnicity-specific identity components

Table 7.1 reports the total number of shared vs face ethnicity-specific identity components of 3D shape and each of the 5 spatial frequency (SF) bands of 2D complexion identified using MI, separately for each stimulus sex (A: Male faces; B: Female faces) and social trait (see titles at top).

A. Male faces

	Trustworthy				Dominant			
	Shared	Specific			Shared	Specific		
		BA	EA	WE		BA	EA	WE
3D shape	25	9	10	14	31	8	4	18
2D comp. SF1	2	6	3	2	3	4	6	7
2D comp. SF2	1	5	0	3	2	3	2	3
2D comp. SF3	0	6	2	3	1	4	0	8
2D comp. SF4	1	6	4	8	3	4	5	6
2D comp. SF5	1	10	5	5	3	11	1	7

B. Female faces

	Trustworthy				Dominant			
	Shared	Specific			Shared	Specific		
		BA	EA	WE		BA	EA	WE
3D shape	24	9	15	13	39	9	4	13
2D comp. SF1	1	3	1	2	4	2	2	8
2D comp. SF2	0	4	4	4	3	5	3	5
2D comp. SF3	3	3	5	2	4	6	1	4
2D comp. SF4	1	2	3	5	3	1	1	1
2D comp. SF5	3	10	3	5	5	10	2	8

Table 7.1: For each stimulus sex (A: Male faces; B: Female faces) and social trait (see titles at top), rows report the number of shared vs face ethnicity-specific identity components identified using MI, separately for 3D shape and each spatial frequency (SF) band of 2D complexion.

Additionally, Figure 7.6 shows the number of observers with a statistically significant effect for each shared and face ethnicity-specific identity component of 3D shape and each of the 5 SFs of 2D complexion (ordered from highest to lowest), separately for each social trait and face ethnicity (see titles at top), separately for male (A) and female faces (B). Colour-coding indicates the identity component sign (red = positive, blue = negative), and colour saturation indicates the number of observers with a statistically significant effect (see colorbar on the right). Only identity components at or above population prevalence threshold

for at least one face ethnicity ($n = 4$ observers per face ethnicity) are shown. In each colour-coded matrix, black horizontal lines divide the shared identity components (top section), from the face ethnicity-specific identity components (bottom section).

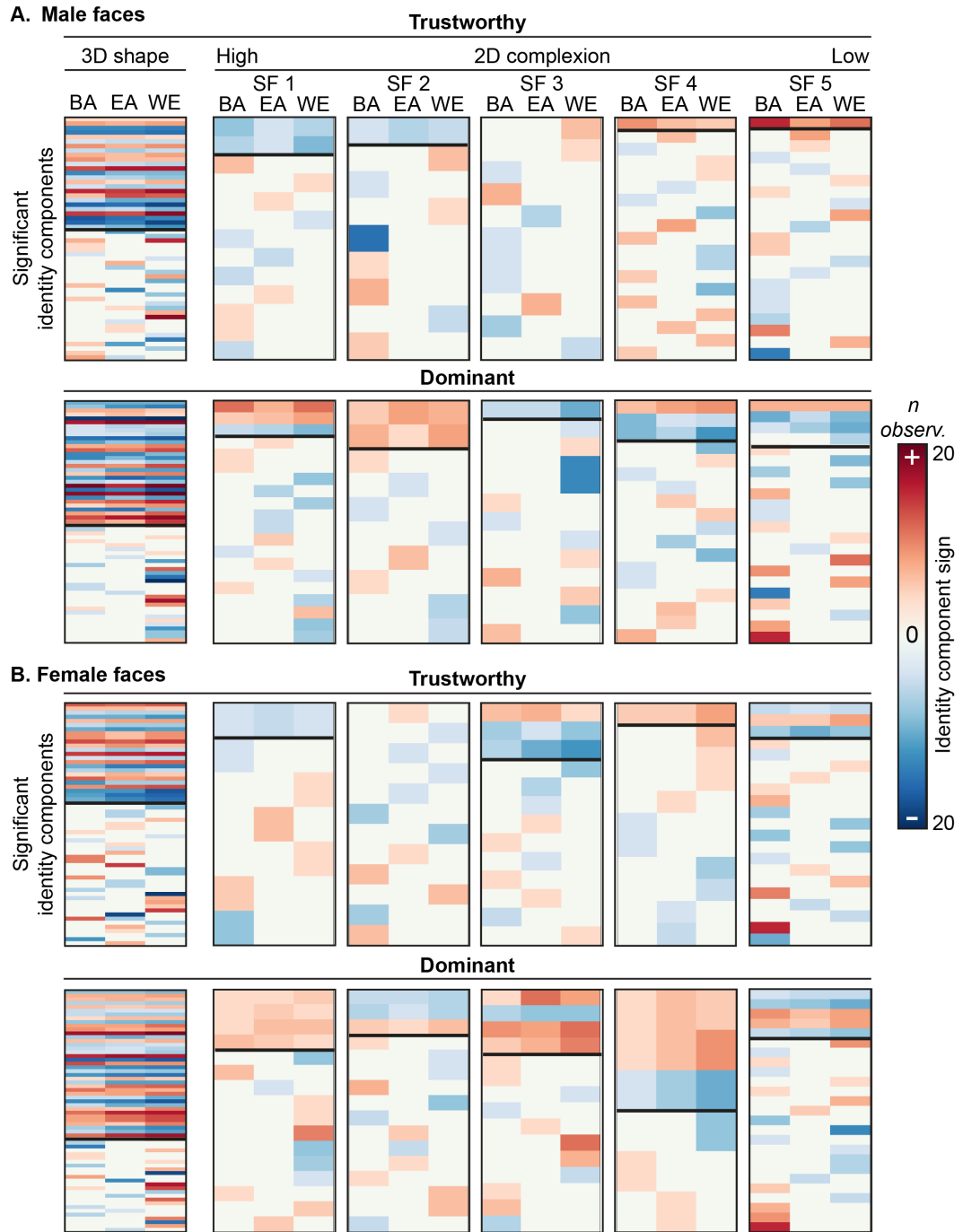


Figure 7.6: In each panel (A: Male faces; B: Female faces), colour-coded matrices show the number of observers with a statistically significant effect for each shared and face ethnicity-specific identity component of 3D shape (leftmost matrix) and of the 5 spatial frequency bands of 2D complexion (right matrices, ordered from highest to lowest SF; see labels at top), separately for trustworthy (top) and dominant judgements (bottom). Colour-coding indicates the identity component sign, where a positive sign (red colour-coding) indicates the addition of the identity component and a negative sign (blue colour-coding) indicates the subtraction of the same identity component in a PCA sense. Colour saturation indicates the number of observers (see colorbar on the right). Only identity components at or above population prevalence threshold for at least one face ethnicity ($n = 4$ observers per face ethnicity) are shown. Black horizontal lines divide the shared identity components (top section) from the face ethnicity-specific identity components (bottom section).

7.2.6 Ethnic phenotypes in the 3D Generative Model of human Faces

To extract the facial features of 3D shape and 2D complexion that represent each ethnic phenotype (i.e., ethnic average in the GMF), I first generated an ethnicity- and sex-neutral average face by setting all categorical factors in the generative model of the human face to 0 and adding no identity components. I then generated a sex-neutral average face for each face ethnicity (Black African, East Asian, White European) by setting each corresponding level of the ethnicity categorical factor to 1 in turn and again adding no sex information nor identity components. Finally, I computed the difference between each of these ethnicity-specific average faces and the ethnicity- and sex-neutral average face, separately for 3D shape and 2D complexion. Figure 7.7 shows the results for each ethnicity.

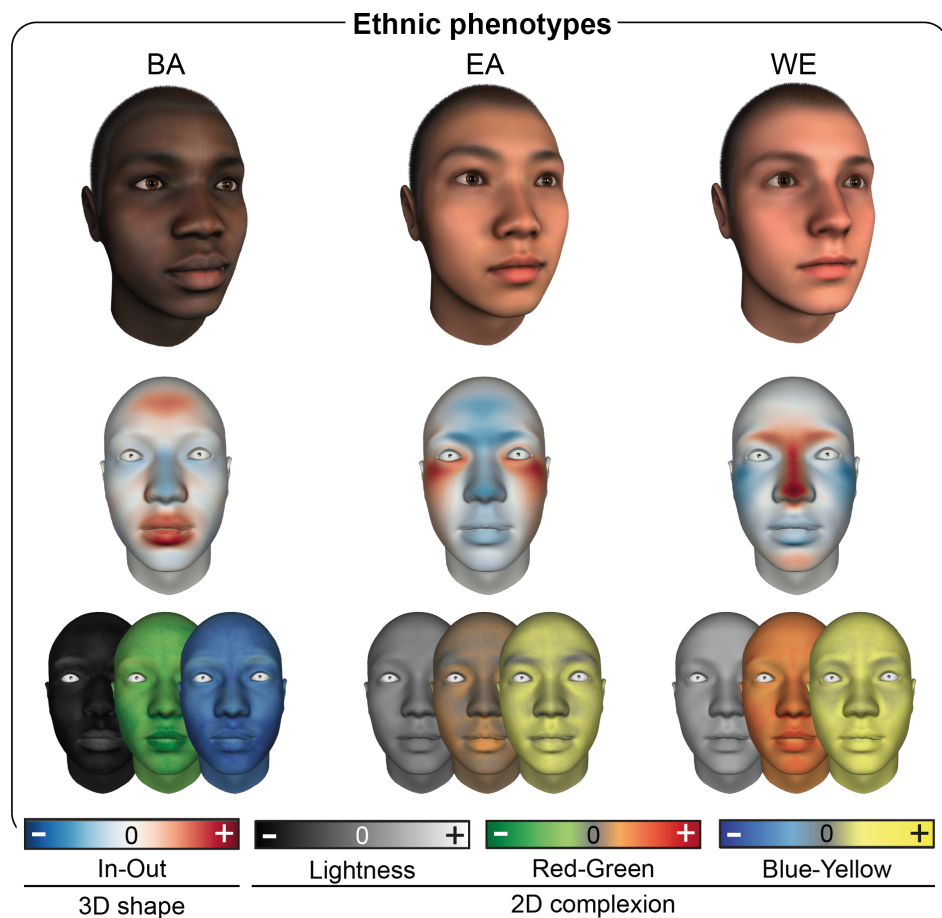


Figure 7.7: Faces at the top show the phenotypic facial features of each ethnicity—Black African (BA), East Asian (EA), and White European (WE)—with shape and complexion combined and displayed on the same sex- and ethnicity-neutral average face. Corresponding colour-coded faces below show the results separately for 3D shape (top) and 2D complexion (bottom), with colour-coding following the same format as throughout Chapter 2 (see colorbars below).

Faces at the top show phenotypic facial features of each ethnicity, with shape and complexion features combined and displayed on the same sex- and ethnicity-neutral average face for comparisons. Colour-coded faces below show the results separately for 3D shape and each L^*a^*b channel of 2D complexion, with colour-coding following the same format as throughout Chapter 2 (see colorbars below). BA-phenotypic features comprise a bigger

head, flatter nose bridge with a wider nose tip, and fuller lips together with a darker, cooler skin tone than the sex- and ethnicity-neutral average face; EA-phenotypic features comprise a smaller head, wider face, higher cheekbones, and a flatter nose and brow ridge together with a warmer skin tone than the sex- and ethnicity-neutral average face; and WE-phenotypic features comprise a narrower face, more prominent brow ridge, higher nose bridge, thinner lips and pointier chin together with a lighter, warmer skin tone than the sex- and ethnicity-neutral average face.

7.2.7 Identifying above-chance face vertex correlations

To identify correlations that were above-chance—that is, correlations which occur due to specific effects rather than natural similarity in face vertex structure—I first generated 500 random ethnically neutral 3D face identities per stimulus sex. For each random face, I then computed their correlation (Pearson r) with each set of ethnic phenotypic features (see Supplementary materials 7.2.6). Finally, I used the 5th and 95th percentile of each distribution as threshold for above-chance correlations. Figure 7.8 shows the distribution for each ethnic phenotypic feature set (see titles at top), separately for each stimulus sex (A: Male faces, B: Female faces). Colour-coded lines and corresponding labels show the 5th and 95th percentile of each distribution (see legend at top).

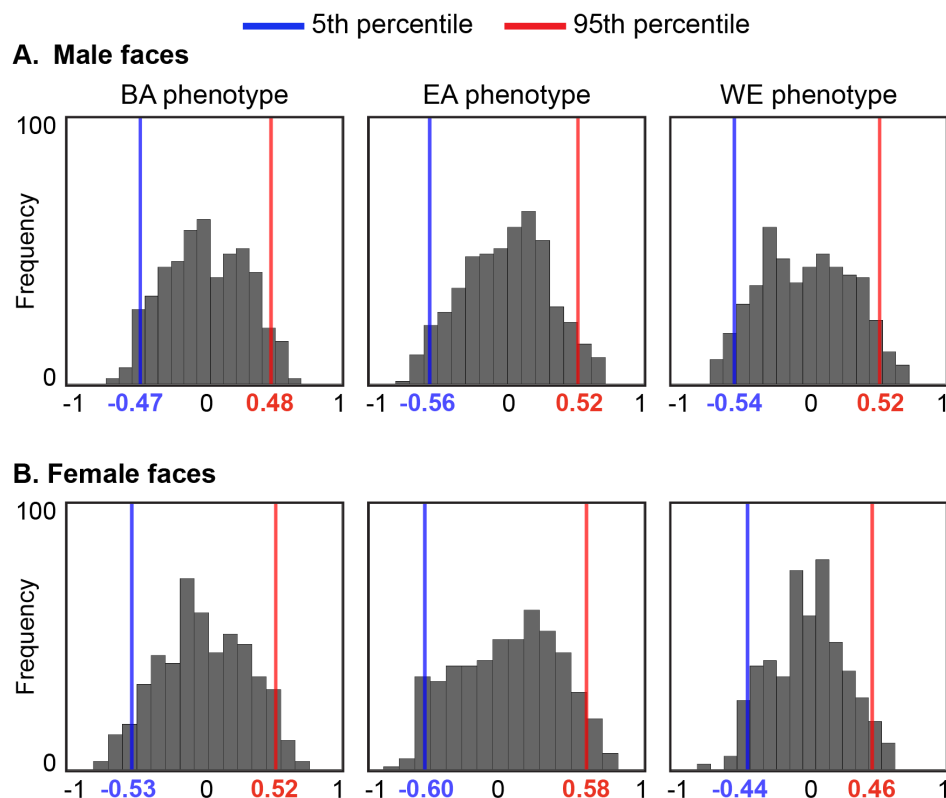


Figure 7.8: In each panel (A: Male faces, B: Female faces), histograms show the distribution of Pearson r correlation values between the 500 randomly generated 3D face identities and each ethnic phenotypic feature set (see titles at top). Colour-coded vertical lines and corresponding colour-coded labels show the 5th and 95th percentile of each distribution (see legend at top).

7.3 Chapter 3

7.3.1 Screening questionnaire for East Asian observers

Each East Asian observer completed the following screening questionnaire to assess their exposure to and contact with non-Western cultures. I only recruited observers who answered ‘no’ to all questions or who reported minimal exposure (e.g., having visited a non-Eastern country for a short period of time and not recently). Additionally, I only recruited East Asian observers who had arrived in the UK no more than 3 months prior to testing, and with a minimum IELTS score of 6 (or equivalent). Native speakers of any main dialect were eligible.

1. At what date did you first enter the UK?
2. What is your IELTS score (or equivalent)?
3. What is your main dialect (Mandarin, Cantonese, Other)?
4. How long have you spent in a non-Eastern* country in total since you were 10 years old?
5. Have you every been in contact with any non-Eastern* person(s) who have been your friend or acquaintance for quite some time?
6. Have you ever been involved with any non-Eastern* culture societies/groups?
7. When do you plan to leave Glasgow?

*By Eastern countries/groups/people, we are referring to China, Japan, Korea, Thailand, and Taiwan

7.3.2 Checking the linearity assumption to model 3D shape and 2D complexion for East Asian observers

Prior to modelling the data for East Asian observers using linear regression, I tested the linearity of the relationship between the observers’ social trait ratings (normalized as described in subsection 3.2.4) and the identity component weights of 3D shape and 2D complexion. To do this, I followed the same procedure I used for White Western observers detailed in Supplementary materials 7.2.3. Figures 7.9 and 7.10 show the results for male and female faces respectively as colour-coded matrices, separately for 3D shape (top row) and each L*a*b channel of 2D complexion (bottom rows), for each face ethnicity and social trait rating task (see titles at top). For clarity, each colour-coded matrix is ordered ascendingly according to the values in the first rating bin.

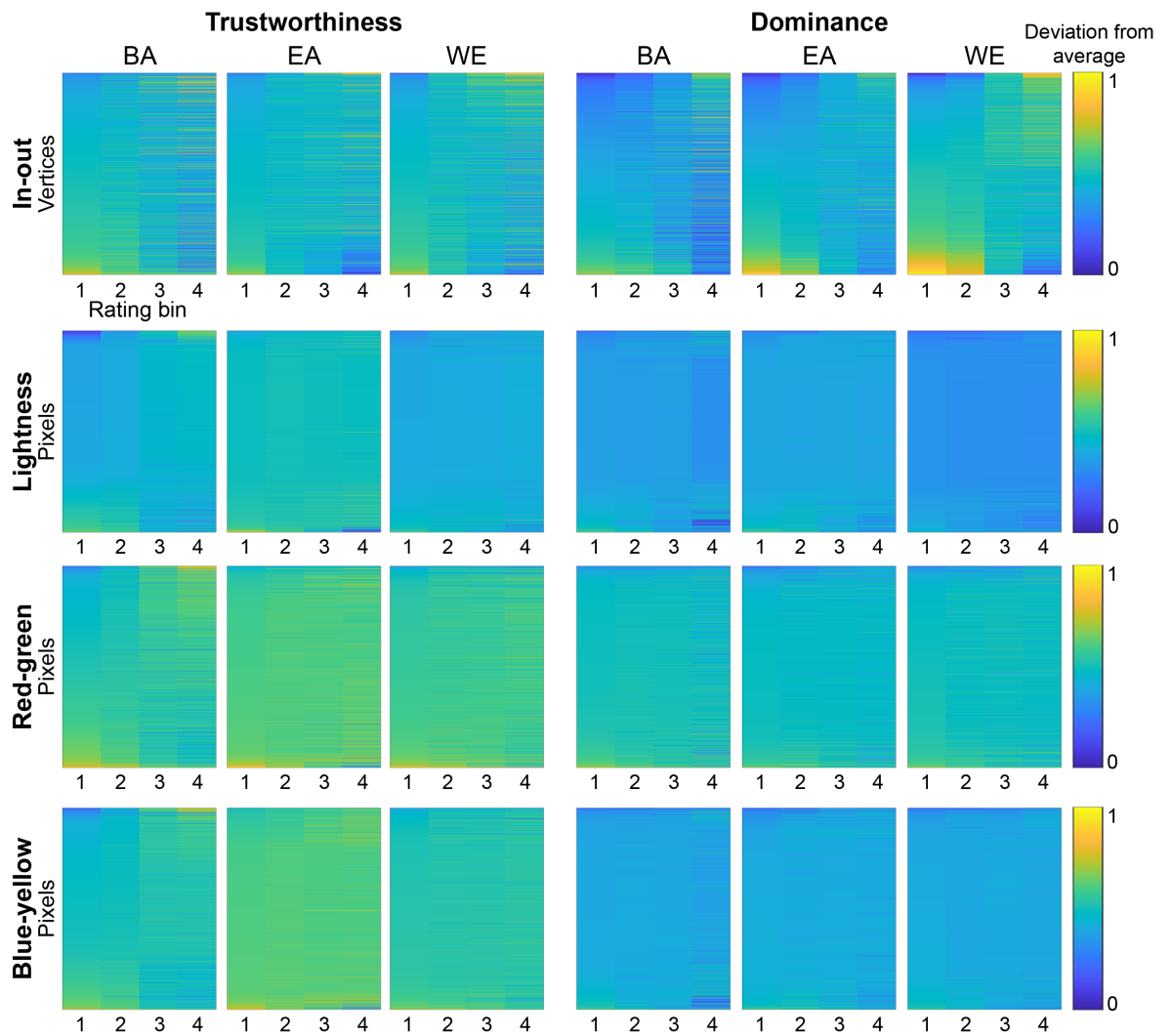


Figure 7.9: For each social trait and face ethnicity (see titles at top), colour-coded matrices show the normalized vertex ($n = 14,319$ front-face vertices; top row) or pixel ($n = 61,218$ downsampled pixels; bottom rows) deviations from the ethnically neutral average face for each of 4 rating bins (x axis) for male faces only, averaged across East Asian observers ($n = 20$) and trials ($n = 1,200$). Colour-coding represents the normalized deviation value (see colorbars).

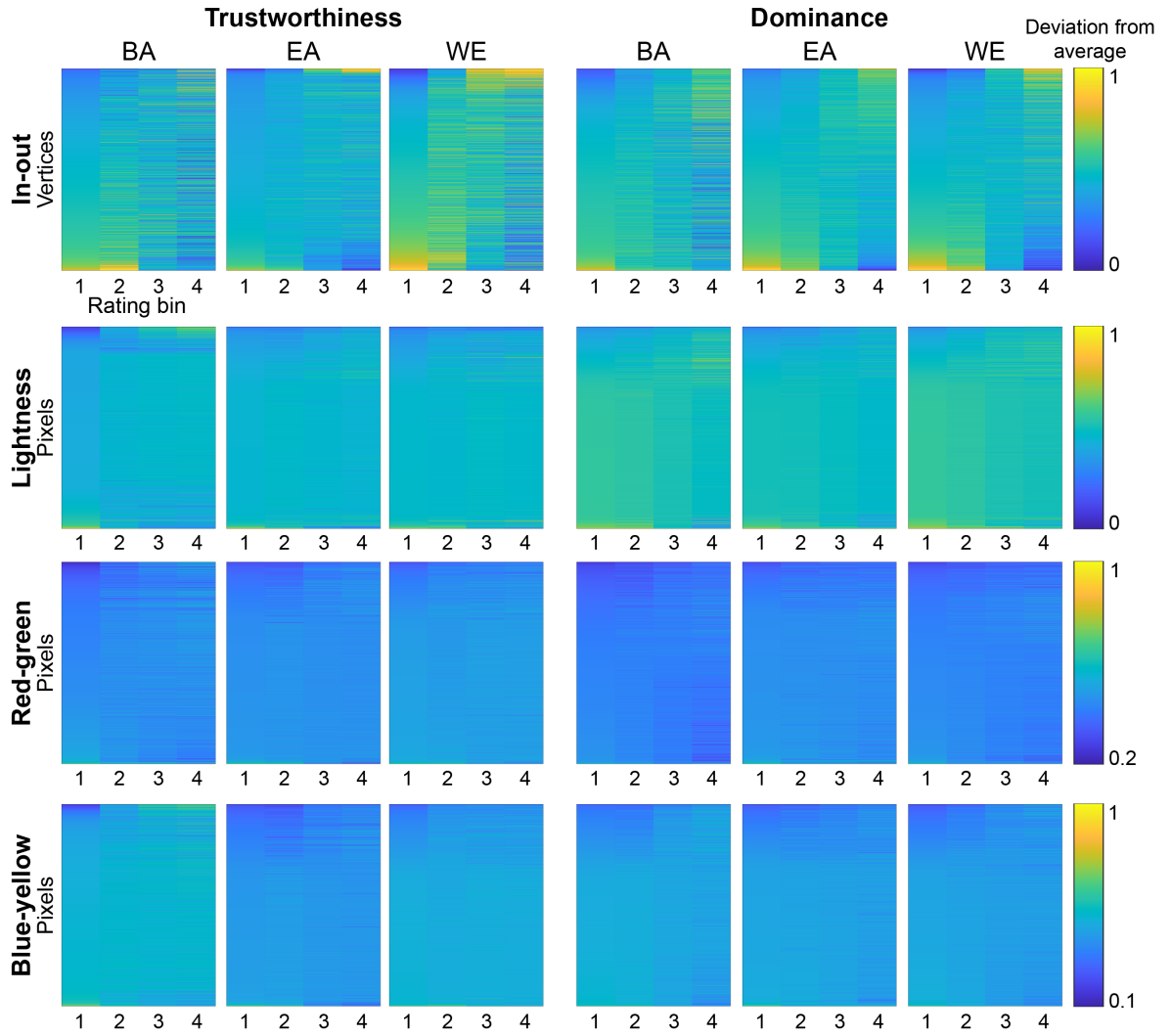


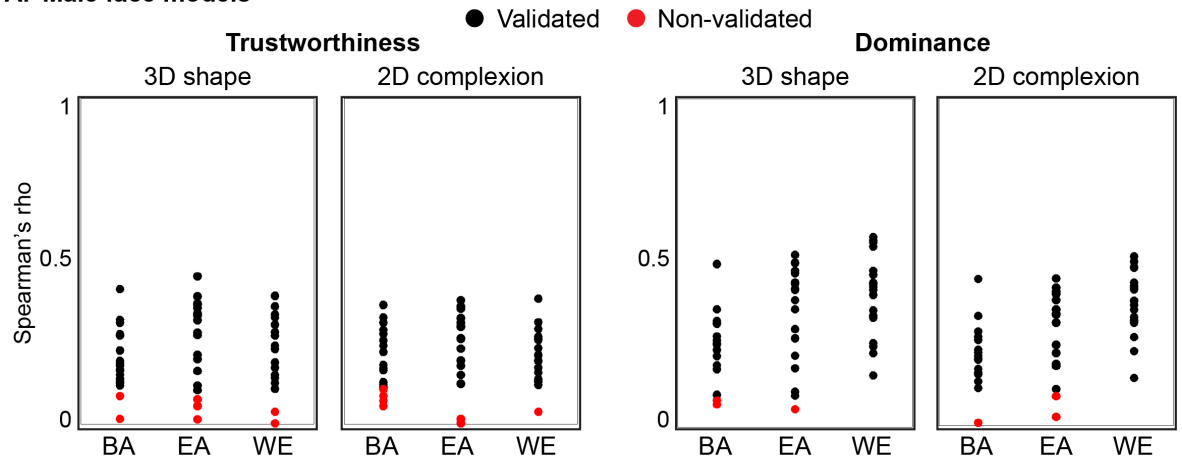
Figure 7.10: Following the same format as Figure 7.9, colour-coded matrices show the normalized deviation values (see colorbars on right) for each of 4 rating bins (x axis), for each social trait and face ethnicity (see titles at top) for female faces only.

In each colour-coded matrix, smooth transitions from yellow to dark blue (or dark blue to yellow) across the four rating bins (i.e., row-wise) would indicate a linear relationship between the observers' ratings and the identity components weights (represented as vertex deviations from the average for 3D shape, and pixel deviations from the average for 2D complexion). 3D shape results (top row) for both male (Figure 7.9) and female faces (Figure 7.10) reflect this smooth transition pattern and thus suggest that, like for White Western observers, the relationship between East Asian observers' ratings and the weights of the 3D shape identity components is linear. 2D complexion results (bottom rows, shown separately for each L^*a^*b channel) for both stimulus sexes also show, on average, a linear row-wise transition—though this pattern is much weaker than for 3D shape. Nevertheless, this mirrors the higher variability of 2D complexion reported in previous work (e.g., Bjornsdottir et al., 2024; Hensel et al., 2020). Given this, and the similarity of these results to those of White Western observers (see Supplementary materials 7.2.3), I deemed linear regression to be an appropriate modelling approach.

7.3.3 Cross-validating the 3D face models of East Asian observers

Using the same approach described in Chapter 2, I used a leave-one-out cross-validation approach to validate the 3D face models I obtained for each East Asian observer, separately for 3D shape and 2D complexion. Figure 7.11 shows the results.

A. Male face models



B. Female face models

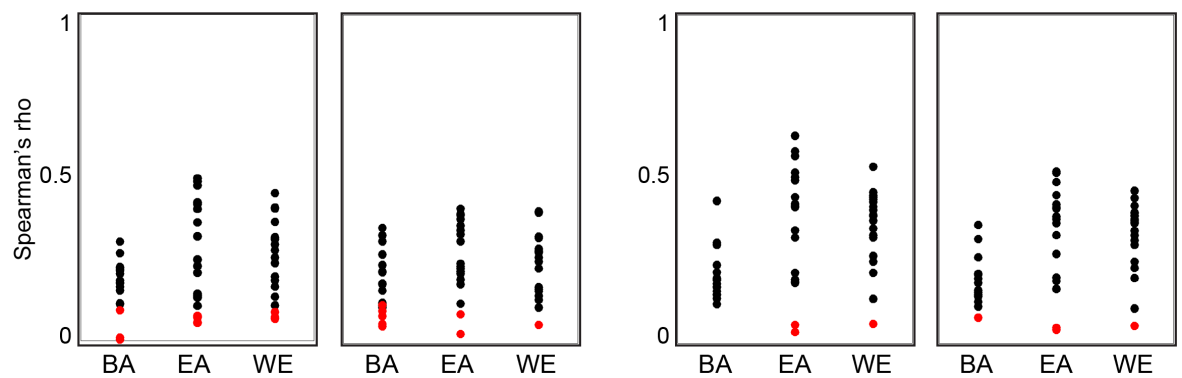


Figure 7.11: Following the same format as Figure 7.5, dot plots in each panel (A: Male face models, B: Female face models) show the prediction accuracy (measured as Spearman's rho) of each individual observer's social trait models based on a leave-one-out cross-validation analysis, separately for each social trait rating task and for 3D shape and 2D complexion (see titles at top). Black dots denote that the model's prediction accuracy is statistically significantly above chance; red dots denote it is not (see legend at top). Most of the 240 face models showed cross-observer validity (213 3D shape models; 210 2D complexion models).

In each subplot, colour-coded dots (black = validated, red = non-validated) show the prediction accuracy (Spearman's rho) of each individual observer's model, based on a LOOCV analysis. Results show that most of the models are valid (213 3D shape models; 210 2D complexion models).

7.3.4 Number and distribution of shared vs face ethnicity-specific identity components

Table 7.2 reports the total number of shared vs face ethnicity-specific identity components of 3D shape and each of the 5 spatial frequency (SF) bands for 2D complexion identi-

fied using MI in Chapter 3, for each stimulus sex and social trait separately (see titles).

A. Male faces

	Trustworthy				Dominant			
	Shared	Specific			Shared	Specific		
		BA	EA	WE		BA	EA	WE
3D shape	14	4	6	1	19	1	4	4
2D comp. SF1	0	0	1	4	2	1	0	5
2D comp. SF2	0	2	1	1	0	1	1	3
2D comp. SF3	0	1	2	0	0	0	4	4
2D comp. SF4	0	4	0	1	1	0	2	2
2D comp. SF5	2	5	2	2	3	4	4	4

B. Female faces

	Trustworthy				Dominant			
	Shared	Specific			Shared	Specific		
		BA	EA	WE		BA	EA	WE
3D shape	10	4	8	3	22	3	3	8
2D comp. SF1	0	2	4	2	0	2	3	1
2D comp. SF2	0	1	2	1	0	1	7	1
2D comp. SF3	0	0	1	2	3	0	2	3
2D comp. SF4	0	0	0	2	0	1	6	0
2D comp. SF5	1	3	3	6	3	0	1	3

Table 7.2: For each stimulus sex (A: Male faces; B: Female faces) and social trait (see titles at top), rows report the number of shared vs face ethnicity-specific identity components identified using MI, separately for 3D shape and each spatial frequency (SF) band of 2D complexion.

Additionally, Figure 7.12 shows the number of observers ($n = 20$ per face ethnicity) with a statistically significant effect for each shared and face ethnicity-specific identity component of 3D shape and each of the 5 SFs of 2D complexion for each social trait and face ethnicity (see titles at top), separately for male (A) and female faces (B). Colour-coding indicates the identity component sign (red = positive, blue = negative), and colour saturation indicates the number of observers with a statistically significant effect (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers per face ethnicity) for at least one face ethnicity are shown. In each colour-coded matrix, black horizontal lines divide the shared identity components (top section), from the face ethnicity-specific identity components (bottom section).

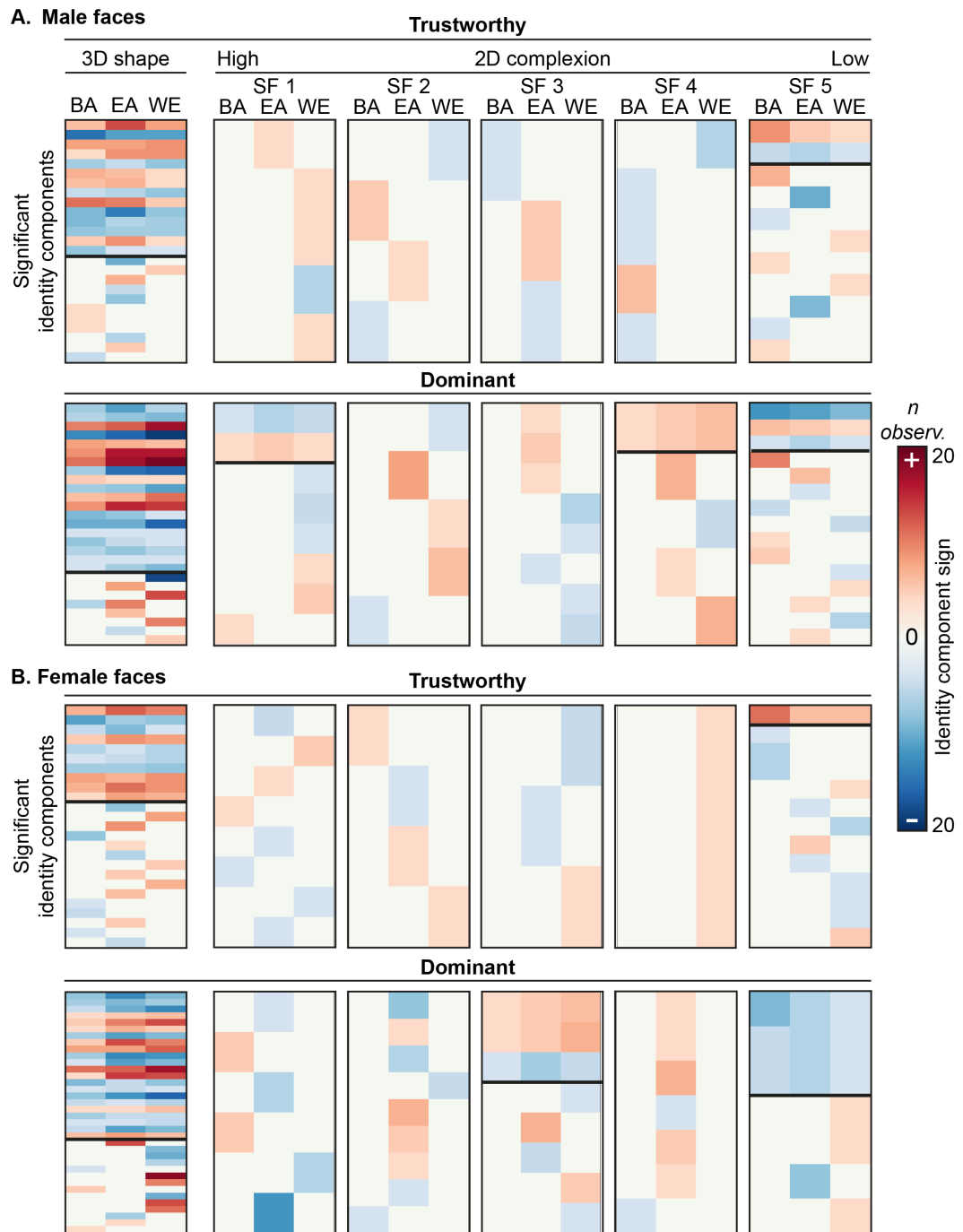


Figure 7.12: Following the same format as Figure 7.6, colour-coded matrices in each panel (A: Male faces; B: Female faces) show the number of observers with a statistically significant effect for each shared and face ethnicity-specific identity component of 3D shape (leftmost matrix) and of the 5 spatial frequency bands of 2D complexion (right matrices, ordered from highest to lowest SF; see labels at top), separately for trustworthy (top) and dominant judgements (bottom). Colour-coding indicates the identity component sign (red = positive; blue = negative), and colour saturation indicates the number of observers (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers per face ethnicity) for at least one face ethnicity are shown. Black horizontal lines divide the shared identity components (top section) from the face ethnicity-specific identity components (bottom section).

7.4 Chapter 4

7.4.1 Number and distribution of culturally shared vs culture-specific identity components used across face ethnicities

Table 7.3 reports the total number of culturally shared vs culture-specific identity components of 3D shape and each of the 5 spatial frequency (SF) bands for 2D complexion identified using MI in Chapter 4, for each stimulus sex (A: Male faces; B: Female faces) and social trait separately (see titles at top).

A. Male faces

	Trustworthy			Dominant		
	Shared	Specific		Shared	Specific	
		East Asian	Western		East Asian	Western
3D shape	2	2	20	3	2	22
2D comp. SF1	0	0	1	0	1	3
2D comp. SF2	0	0	2	0	0	2
2D comp. SF3	0	0	0	0	0	2
2D comp. SF4	0	0	1	1	0	1
2D compl. SF5	0	0	1	1	0	0

B. Female faces

	Trustworthy			Dominant		
	Shared	Specific		Shared	Specific	
		East Asian	Western		East Asian	Western
3D shape	3	1	20	10	4	26
2D comp. SF1	0	0	0	0	0	0
2D comp. SF2	0	0	0	0	0	0
2D comp. SF3	0	0	0	1	0	2
2D comp. SF4	0	0	0	0	0	0
2D comp. SF5	0	0	0	1	0	1

Table 7.3: For each stimulus sex (A: Male faces; B: Female faces) and social trait (see titles at top), rows report the number of culturally shared vs culture-specific (East Asian; Western) identity components, separately for 3D shape and each spatial frequency (SF) band of 2D complexion.

Additionally, Figure 7.13 shows the number of observers ($n = 60$ [20 observers $\times$ 3 face ethnicity]) with a statistically significant effect for each culturally shared and culture-specific identity component of 3D shape and each of the 5 SFs of 2D complexion for each social trait and observer culture (see titles at top), separately for male (A) and female faces (B). Colour-coding indicates the identity component sign (red = positive, blue = negative), and colour saturation indicates the number of observers with a statistically significant effect

(see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) in each face ethnicity condition are shown. In each colour-coded matrix, black horizontal lines divide the culturally shared identity components (top section), from the culture-specific identity components (bottom section).

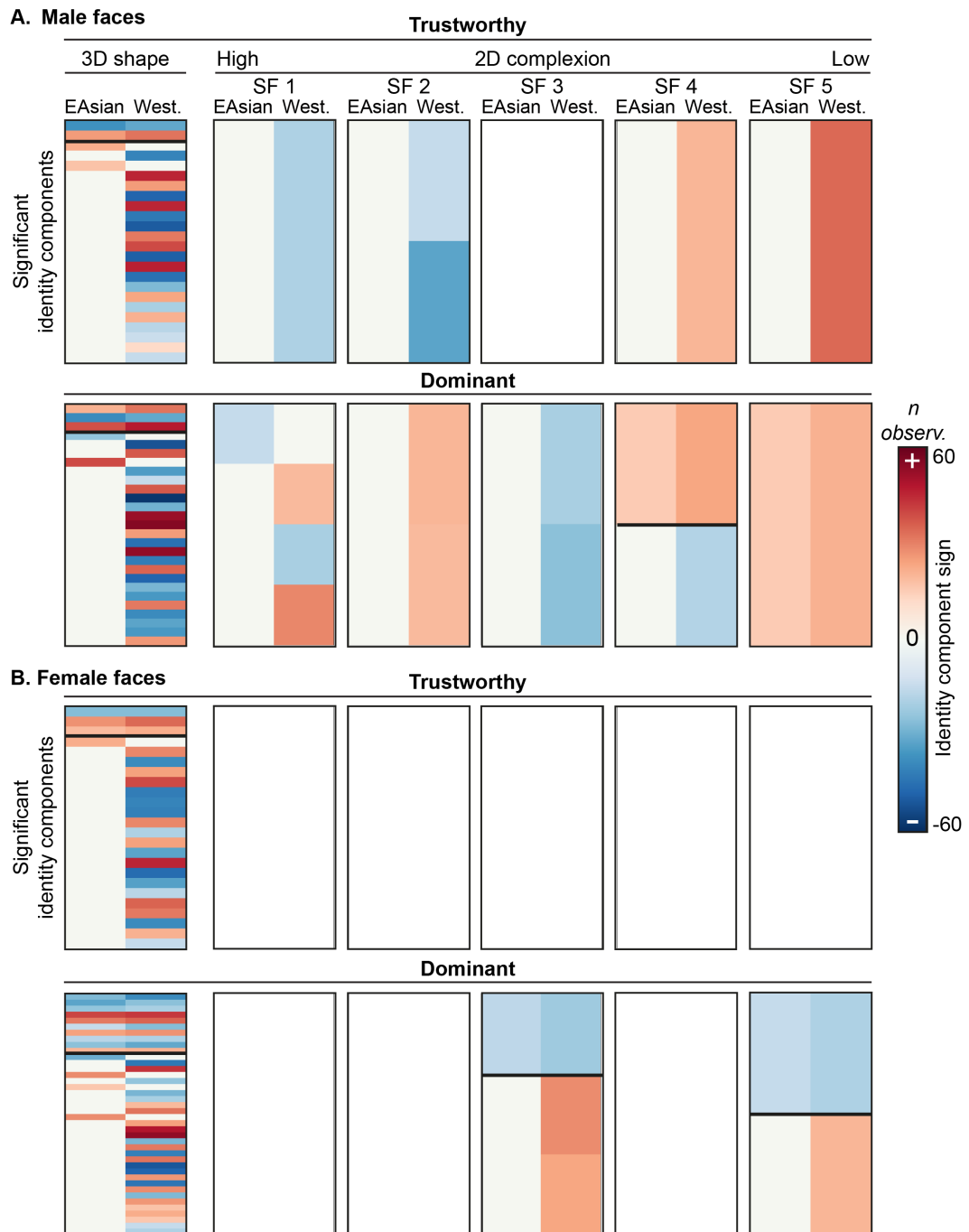


Figure 7.13: In each panel (A: Male faces; B: Female faces), colour-coded matrices show the number of observers with a statistically significant effect for each shared and culture-specific identity component of 3D shape (leftmost matrix) and of the 5 spatial frequency bands of 2D complexion (right matrices, ordered from highest to lowest SF; see labels at top), separately for trustworthy (top) and dominant judgements (bottom). Colour-coding indicates the identity component sign (red = positive; blue = negative), and colour saturation indicates the number of observers (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) in each face ethnicity for at least one observer culture are shown. Black horizontal lines divide the culturally shared identity components (top section) from the culture-specific identity components (bottom section).

7.4.2 Relevant 3D face vertices and cartesian dimension of bilateral Action Units

Red colour-coded faces in Figure 7.14 show the relevant 3D face vertices for each of 28 Action Units (see bolded titles above), identified using the approach described in Chapter 4. Subtitles above each colour-coded face report the cartesian dimension (x, y, z, axis) with the largest absolute effect size. Star icons next to a colour-coded face indicate that the AU has a smaller movement magnitude, and its relevant 3D face vertices were therefore identified using the 10th and 90th percentile of its vertex deviation distribution, rather than the 5th and 95th percentile.

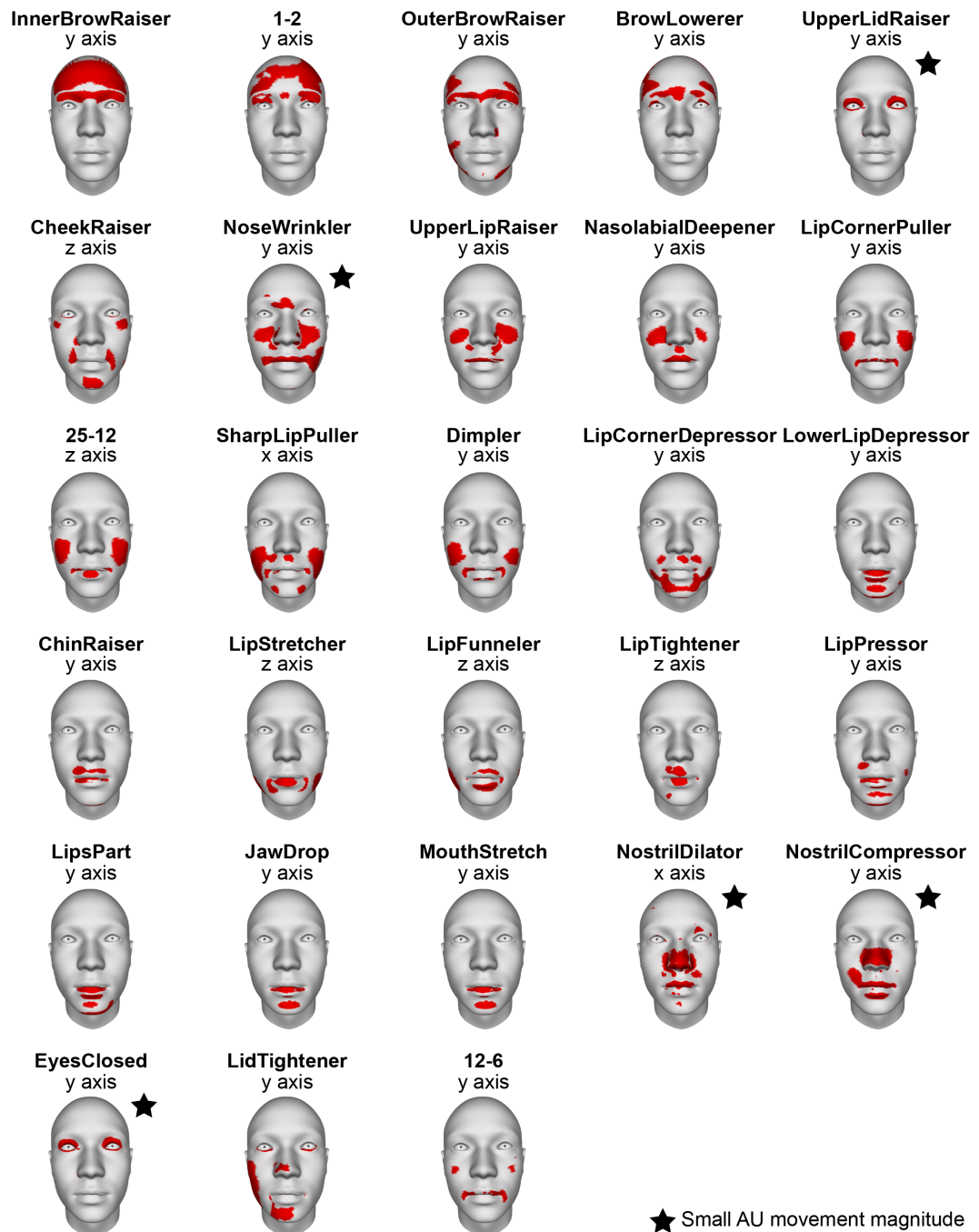


Figure 7.14: For each bilateral Action Unit (AU; see bolded titles above), colour-coded faces shows the 3D face vertices that best capture each facial movement (red = relevant; white = irrelevant). Subtitles above each face indicate the cartesian dimension (x, y, z axis) with the largest absolute effect size; star icons indicate small AU movement magnitude, and thus larger vertex threshold (10th and 90th percentile).

7.4.3 Action Unit patterns representing the culture-specific facial features

Using the approach described in Chapter 4, I represented the culture-specific facial features used for social trait perception by East Asian and White Western observers as Action Units (AUs). Figure 7.15 shows the resulting AU patterns for each social trait (x axis) in each observer culture (see labels above), separately for male (A) and female (B) faces and with

results aggregated across individual observers ($n = 60$ [20 observers $\times$ 3 face ethnicity]). Colour saturation indicates the number of observers with a statistically significant effect (see colorbar below), for all bilateral AUs at or above population prevalence threshold ($n = 4$ observers) in each face ethnicity condition.

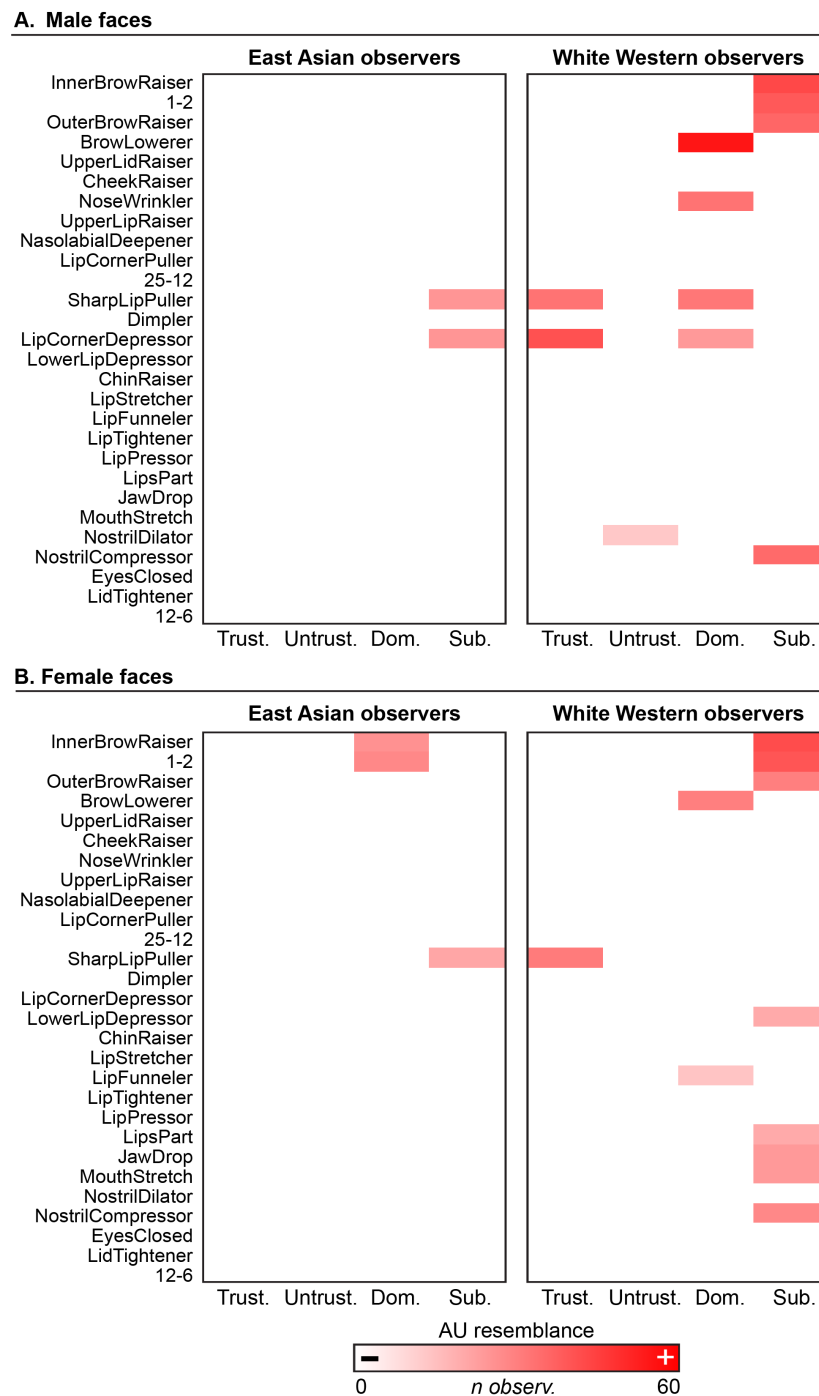


Figure 7.15: In each panel (A: Male faces; B: Female faces), colour-coded matrices show the AU patterns that represent the culture-specific social facial features used by East Asian (left) and White Western observers (right). Colour saturation indicates the number of observers with a statistically significant effect (see colorbar below), for all bilateral AUs at or above population prevalence threshold ($n = 4$ observers) in each face ethnicity condition.

7.5 Chapter 5

7.5.1 Number and distribution of face ethnicity-specific identity components across observer cultures

Table 7.4 reports the total number of face ethnicity-specific identity components of 3D shape and each of the 5 spatial frequency (SF) bands for 2D complexion identified using MI in Chapter 5, for each stimulus sex (A: Male faces; B: Female faces) and social trait separately (see titles at top).

A. Male faces						
	Trustworthy			Dominant		
	BA	EA	WE	BA	EA	WE
3D shape	6	11	2	1	1	24
2D comp. SF1	0	0	2	0	0	3
2D comp. SF2	0	0	0	0	1	5
2D comp. SF3	0	0	0	0	0	3
2D comp. SF4	0	0	0	0	0	2
2D comp. SF5	3	1	0	4	0	3

B. Female faces						
	Trustworthy			Dominant		
	BA	EA	WE	BA	EA	WE
3D shape	5	9	4	1	0	17
2D comp. SF1	0	0	0	1	0	2
2D comp. SF2	0	0	1	0	3	0
2D comp. SF3	0	0	0	0	1	3
2D comp. SF4	0	1	0	0	0	2
2D comp. SF5	5	2	3	2	0	0

Table 7.4: For each stimulus sex (A: Male faces; B: Female faces) and social trait (see titles at top), rows report the number of face ethnicity-specific identity components identity components used across cultures, separately for 3D shape and each spatial frequency (SF) band of 2D complexion.

Additionally, Figure 7.16 shows the number of observers ($n = 40$ [20 observers $\times$ 2 observer cultures]) with a statistically significant effect for face ethnicity-specific identity component of 3D shape and each of the 5 SFs of 2D complexion for each social trait and face ethnicity (see titles at top), separately for male (A) and female faces (B). Colour-coding indicates the identity component sign (red = positive, blue = negative), and colour saturation indicates the number of observers with a statistically significant effect (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) in both observer cultures are shown.

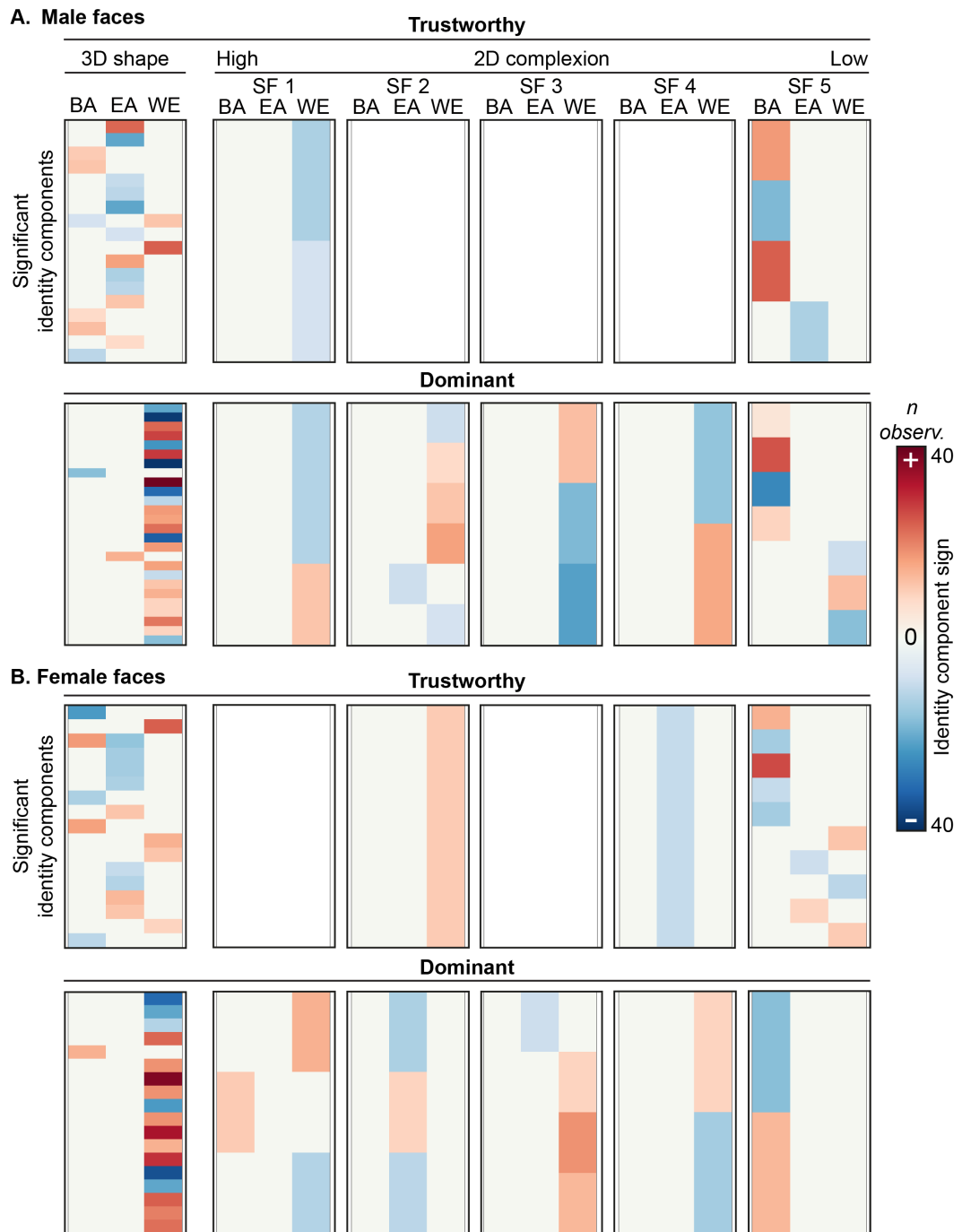


Figure 7.16: In each panel (A: Male faces; B: Female faces), colour-coded matrices show the number of observers with a statistically significant effect for each face ethnicity-specific identity component of 3D shape (leftmost matrix) and of the 5 spatial frequency bands of 2D complexion (right matrices, ordered from highest to lowest SF; see labels at top), separately for trustworthy (top) and dominant judgements (bottom). Colour-coding indicates the identity component sign (red = positive; blue = negative), and colour saturation indicates the number of observers (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) in both observer cultures for at least one face ethnicity are shown.

7.5.2 Number and distribution of synergistic identity components

Tables 7.5 and 7.6 report, for male and female faces respectively, the total number of synergistic identity components of 3D shape and each of the 5 spatial frequency (SF) bands

for 2D complexion identified using Co-I in Chapter 5 for each combination of face ethnicity and observer culture, separately for trustworthy (top) and dominant judgments (bottom).

Trustworthy						
	East Asian culture			White Western culture		
	BA	EA	WE	BA	EA	WE
3D shape	2	1	1	2	4	4
2D comp. SF1	0	0	0	1	1	0
2D comp. SF2	1	0	1	2	0	0
2D comp. SF3	0	2	0	1	4	0
2D comp. SF4	1	0	1	1	0	0
2D comp. SF5	1	2	2	1	1	3

Dominant						
	East Asian culture			White Western culture		
	BA	EA	WE	BA	EA	WE
3D shape	1	2	2	3	0	4
2D comp. SF1	1	0	1	1	3	2
2D comp. SF2	0	0	1	1	0	0
2D comp. SF3	0	2	0	2	0	2
2D comp. SF4	0	1	0	3	2	1
2D comp. SF5	2	2	3	1	2	1

Table 7.5: For trustworthy (top) and dominant judgments (bottom), rows report the number of synergistic identity components of male faces for each combination of face ethnicity and observer culture (see titles above), separately for 3D shape and each spatial frequency (SF) band of 2D complexion.

Trustworthy						
	East Asian culture			White Western culture		
	BA	EA	WE	BA	EA	WE
3D shape	3	5	1	2	5	4
2D comp. SF1	1	1	1	2	2	2
2D comp. SF2	0	0	1	1	1	0
2D comp. SF3	0	0	0	0	2	0
2D comp. SF4	0	0	2	0	0	0
2D comp. SF5	1	1	2	3	1	1

Dominant						
	East Asian culture			White Western culture		
	BA	EA	WE	BA	EA	WE
3D shape	1	7	3	1	5	7
2D comp. SF1	1	0	0	1	1	1
2D comp. SF2	0	4	0	2	0	3
2D comp. SF3	0	1	0	2	1	1
2D comp. SF4	0	4	0	3	1	1
2D comp. SF5	1	0	2	5	1	1

Table 7.6: For trustworthy (top) and dominant judgments (bottom), rows report the number of synergistic identity components of female faces for each combination of face ethnicity and observer culture (see titles above), separately for 3D shape and each spatial frequency (SF) band of 2D complexion.

Additionally, Figure 7.17 shows the number of observers ($n = 20$) with a statistically significant effect for each synergistic identity component of 3D shape and each of the 5 SFs of 2D complexion for trustworthy (top) and dominant judgements (bottom), separately for male (A) and female faces (B). Colour-coding indicates the identity component sign (red = positive, blue = negative), and colour saturation indicates the number of observers with a statistically significant effect (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) for at least one face ethnicity and observer culture are shown.

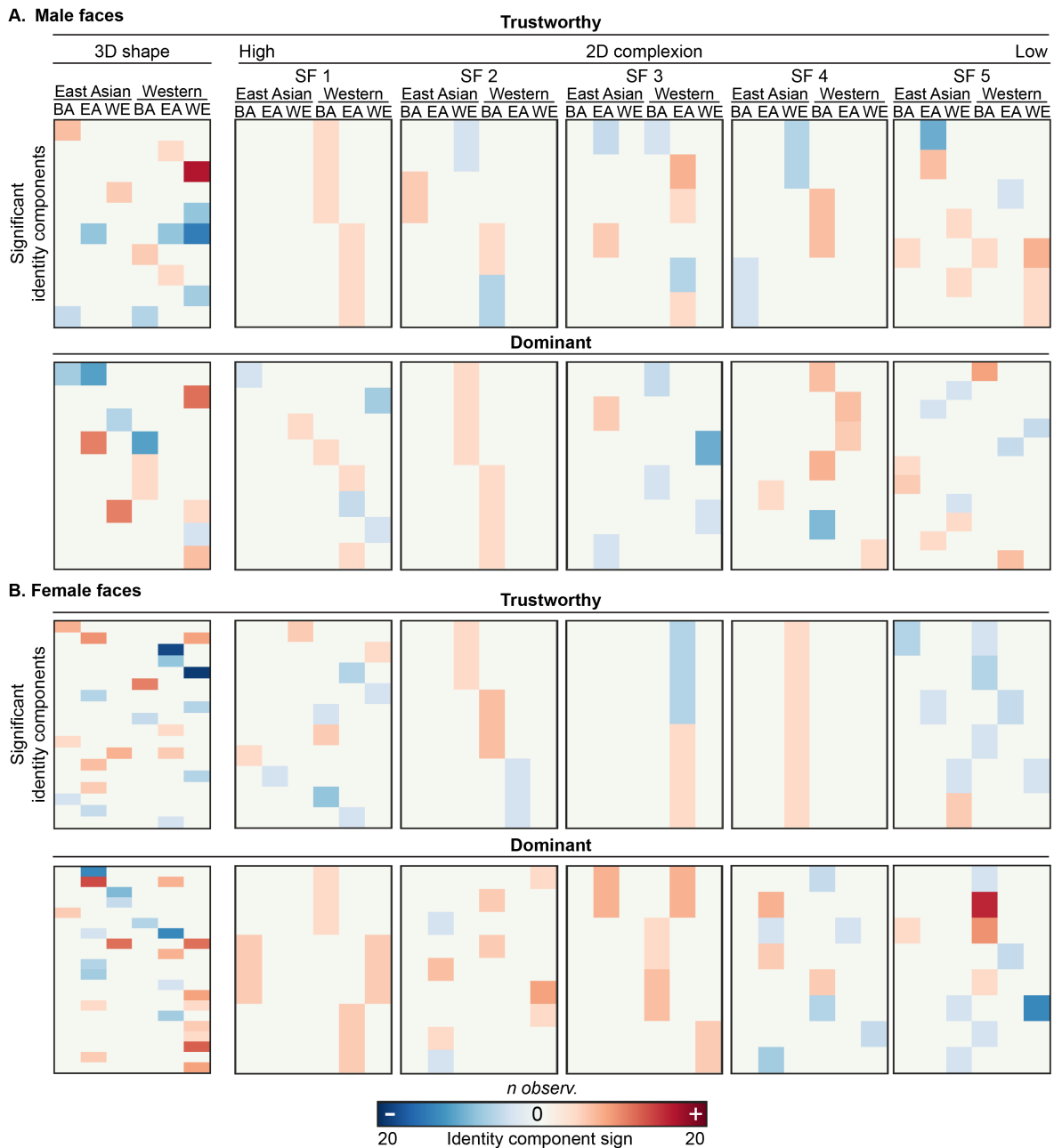


Figure 7.17: In each panel (A: Male faces; B: Female faces), colour-coded matrices show the number of observers ($n = 20$) with a statistically significant effect for each synergistic identity component of 3D shape (leftmost matrix) and of the 5 spatial frequency bands of 2D complexion (right matrices, ordered from highest to lowest SF; see labels at top), separately for trustworthy (left) and dominant judgements (right). Colour-coding indicates the identity component sign (red = positive; blue = negative), and colour saturation indicates the number of observers (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) for at least one face ethnicity and observer culture are shown.

7.5.3 Number and distribution of shared identity components

Table 7.7 reports the total number of shared identity components of 3D shape and each of the 5 spatial frequency (SF) bands for 2D complexion identified in Chapter 5, separately for each stimulus sex and social trait separately (see titles at top).

	Male faces		Female faces	
	Trustworthy	Dominant	Trustworthy	Dominant
3D shape	2	2	6	2
2D comp. SF1	0	0	0	0
2D comp. SF2	0	0	0	0
2D comp. SF3	0	0	0	1
2D comp. SF4	0	1	0	0
2D comp. SF5	0	0	0	1

Table 7.7: For each stimulus sex and social trait (see titles at top), rows report the number of face ethnicity-specific identity components identity components used across cultures, separately for 3D shape and each spatial frequency (SF) band of 2D complexion.

Additionally, Figure 7.18 shows the number of observers ($n = 120$ [20 observers $\times$ 3 face ethnicities $\times$ 2 observer cultures]) with a statistically significant effect for face ethnicity-specific identity component of 3D shape and each of the 5 SFs of 2D complexion for trustworthy (left) and dominant judgements (right), separately for male (A) and female faces (B). Colour-coding indicates the identity component sign (red = positive, blue = negative), and colour saturation indicates the number of observers with a statistically significant effect (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) in all three face ethnicity conditions and both observer cultures are shown.

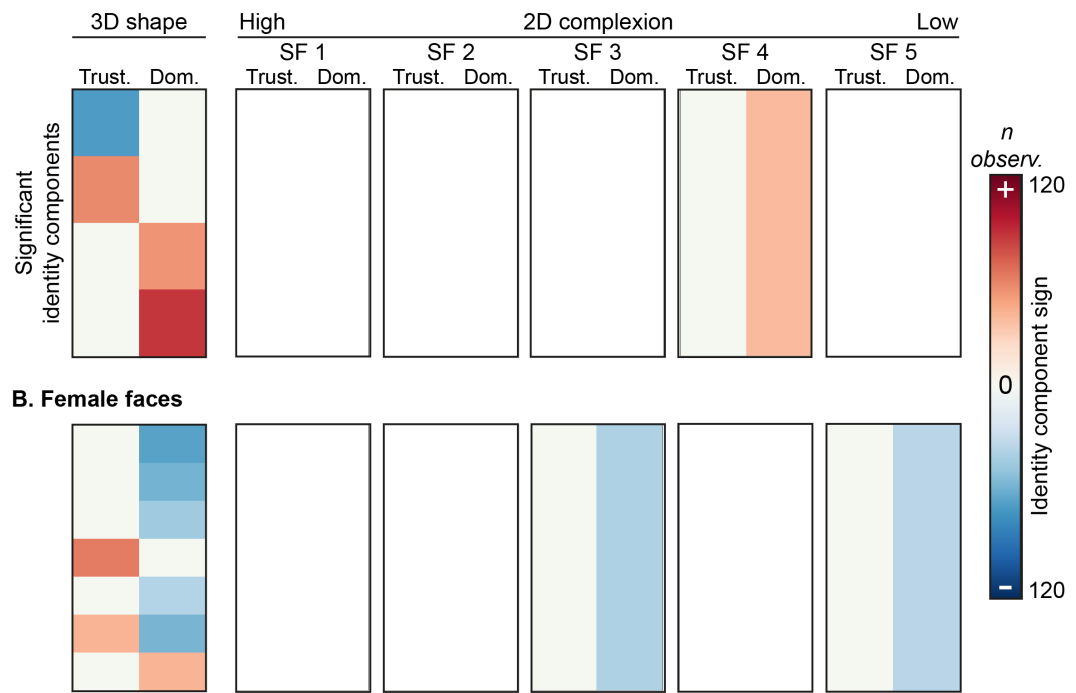
A. Male faces

Figure 7.18: In each panel (A: Male faces; B: Female faces), colour-coded matrices show the number of observers with a statistically significant effect for each shared identity component of 3D shape (leftmost matrix) and of the 5 spatial frequency bands of 2D complexion (right matrices, ordered from highest to lowest SF; see labels at top), separately for trustworthy (left) and dominant judgements (right). Colour-coding indicates the identity component sign (red = positive; blue = negative), and colour saturation indicates the number of observers (see colorbar on the right). Only identity components at or above population prevalence threshold ($n = 4$ observers) in each face ethnicity condition and both observer cultures are shown.

References

- Abbott, J., Middlemiss, M., Bruce, V., Smailes, D., & Dudley, R. (2018). The effect of arousal and eye gaze direction on trust evaluations of stranger's faces: A potential pathway to paranoid thinking. *Journal of Behavior Therapy and Experimental Psychiatry*, 60, 29–36. <https://doi.org/10.1016/j.jbtep.2018.02.007>
- Adams, R. B., Albohn, D. N., Hedgecoth, N., Garrido, C. O., & Adams, K. D. (2022). Angry White Faces: A Contradiction of Racial Stereotypes and Emotion-Resembling Appearance. *Affective Science*, 3(1), 46–61. <https://doi.org/10.1007/s42761-021-00091-5>
- Adams, R. B., Albohn, D. N., & Kveraga, K. (2017). Social Vision: Applying a Social-Functional Approach to Face and Expression Perception [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, 26(3), 243–248. <https://doi.org/10.1177/0963721417706392>
- Adams, R. B., Franklin, R. G., Nelson, A. J., & Stevenson, M. T. (2011, January). Compound Social Cues in Human Face Processing. In *The Science of Social Vision*. Oxford University Press. Retrieved October 31, 2024, from <http://www.scopus.com/inward/record.url?scp=84921882990&partnerID=8YFLogxK>
- Adams, R. B., Hess, U., & Kleck, R. E. (2015). The Intersection of Gender-Related Facial Appearance and Facial Displays of Emotion [Publisher: SAGE Publications]. *Emotion Review*, 7(1), 5–13. <https://doi.org/10.1177/1754073914544407>
- Adams, R. B., & Kleck, R. E. (2003). Perceived Gaze Direction and the Processing of Facial Displays of Emotion [Publisher: SAGE Publications Inc]. *Psychological Science*, 14(6), 644–647. https://doi.org/10.1046/j.0956-7976.2003.psci_1479.x
- Adams, R. B., & Kleck, R. E. (2005). Effects of Direct and Averted Gaze on the Perception of Facially Communicated Emotion. *Emotion*, 5(1), 3–11. <https://doi.org/10.1037/1528-3542.5.1.3>
- Adams, R. B., & Kveraga, K. (2015). Social Vision: Functional Forecasting and the Integration of Compound Social Cues. *Review of Philosophy and Psychology*, 6(4), 591–610. <https://doi.org/10.1007/s13164-015-0256-1>
- Adams, R. B., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition & Emotion*, 26(3), 431–441. <https://doi.org/10.1080/02699931.2012.666502>

- Ahumada, A., & Lovell, J. (1971). Stimulus Features in Signal Detection. *The Journal of the Acoustical Society of America*, 49(6B), 1751–1756. <https://doi.org/10.1121/1.1912577>
- Albert, G., Wells, E., Arnocky, S., Liu, C. H., & Hodges-Simeon, C. R. (2021). Observers use facial masculinity to make physical dominance assessments following 100-ms exposure. *Aggressive Behavior*, 47(2), 226–235. <https://doi.org/10.1002/ab.21941>
- Albohn, D. N., & Adams, R. B. (2021a). The Expressive Triad: Structure, Color, and Texture Similarity of Emotion Expressions Predict Impressions of Neutral Faces [Publisher: Frontiers]. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.612923>
- Albohn, D. N., & Adams, R. B. (2021b). Emotion Residue in Neutral Faces: Implications for Impression Formation [Publisher: SAGE Publications Inc]. *Social Psychological and Personality Science*, 12(4), 479–486. <https://doi.org/10.1177/1948550620923229>
- Albohn, D. N., & Adams, R. B. J. (2022). The Social Face Hypothesis. *Affective Science*, 3(3), 539. <https://doi.org/10.1007/s42761-022-00116-7>
- Albohn, D. N., Brandenburg, J. C., Kveraga, K., & Adams, R. B. (2022). The shared signal hypothesis: Facial and bodily expressions of emotion mutually inform one another. *Attention, Perception, & Psychophysics*, 84(7), 2271–2280. <https://doi.org/10.3758/s13414-022-02548-6>
- Albright, L., Malloy, T. E., Dong, Q., Kenny, D. A., Fang, X., Winkquist, L., & Yu, D. (1997). Cross-cultural consensus in personality judgments. *Journal of Personality and Social Psychology*, 72(3), 558–569. <https://doi.org/10.1037/0022-3514.72.3.558>
- Alrajih, S., & Ward, J. (2014). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. *British Journal of Psychology*, 105(2), 153–161. <https://doi.org/10.1111/bjop.12035>
- Ames, D. R., Kammrath, L. K., Suppes, A., & Bolger, N. (2010). Not So Fast: The (Not-Quite-Complete) Dissociation Between Accuracy and Confidence in Thin-Slice Impressions. *Personality and Social Psychology Bulletin*, 36(2), 264–277. <https://doi.org/10.1177/0146167209354519>
- Arias, P., Belin, P., & Aucouturier, J.-J. (2018). Auditory smiles trigger unconscious facial imitation. *Current Biology*, 28(14), R782–R783. <https://doi.org/10.1016/j.cub.2018.05.084>
- Armstrong, M. M., Lee, A. J., & Feinberg, D. R. (2019). A house of cards: Bias in perception of body size mediates the relationship between voice pitch and perceptions of dominance. *Animal Behaviour*, 147, 43–51. <https://doi.org/10.1016/j.anbehav.2018.11.005>
- Arnocky, S., Carré, J. M., Bird, B. M., Moreau, B. J. P., Vaillancourt, T., Ortiz, T., & Marley, N. (2018). The Facial Width-to-Height Ratio Predicts Sex Drive, Sociosexuality, and Intended Infidelity. *Archives of Sexual Behavior*, 47(5), 1375–1385. <https://doi.org/10.1007/s10508-017-1070-x>

- Balas, B., & Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Computers in Human Behavior*, 77, 240–248. <https://doi.org/10.1016/j.chb.2017.08.045>
- Batres, C., Re, D. E., & Perrett, D. I. (2015). Influence of Perceived Height, Masculinity, and Age on Each Other and on Perceptions of Dominance in Male Faces. *Perception*, 44(11), 1293–1309. <https://doi.org/10.1177/0301006615596898>
- Batres, C., & Shiramizu, V. (2023). Examining the “attractiveness halo effect” across cultures. *Current Psychology*, 42(29), 25515–25519. <https://doi.org/10.1007/s12144-022-03575-0>
- Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: Role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3559–3565. <https://doi.org/10.1098/rstb.2009.0148>
- Baylor, A. L., & Kim, Y. (2004). Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems* (pp. 592–603). Springer. https://doi.org/10.1007/978-3-540-30139-4_56
- Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, 92(2), 179–190. <https://doi.org/10.1037/0022-3514.92.2.179>
- Bell, A. J. (2003). The co-information lattice. *Proceedings of the fifth international workshop on independent component analysis and blind signal separation, 2003*.
- Bente, G., Dratsch, T., Kaspar, K., Häßler, T., Bungard, O., & Al-Issa, A. (2014). Cultures of Trust: Effects of Avatar Faces and Reputation Scores on German and Arab Players in an Online Trust-Game (F. Krueger, Ed.). *PLoS ONE*, 9(6), e98297. <https://doi.org/10.1371/journal.pone.0098297>
- Berry, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology*, 48(2), 312–323. <https://doi.org/10.1037/0022-3514.48.2.312>
- Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, 100(1), 3–18. <https://doi.org/10.1037/0033-2909.100.1.3>
- Bickmore, T., & Cassell, J. (2001). Relational agents: A model and implementation of building user trust. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 396–403. <https://doi.org/10.1145/365024.365304>
- Bijlstra, G., Holland, R. W., Dotsch, R., & Wigboldus, D. H. J. (2019). Stereotypes and prejudice affect the recognition of emotional body postures. *Emotion*, 19(2), 189–199. <https://doi.org/10.1037/emo0000438>
- Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., & Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces’ observed trustworthiness across four samples. *Personality and Individual Differences*, 69, 56–61. <https://doi.org/10.1016/j.paid.2014.05.012>

- Bjornsdottir, R. T., Hensel, L. B., Zhan, J., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2024). Social class perception is driven by stereotype-related facial features. *Journal of Experimental Psychology: General*, 153(3), 742–753. <https://doi.org/10.1037/xge0001519>
- Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of Personality and Social Psychology*, 113(4), 530–546. <https://doi.org/10.1037/pspa0000091>
- Blair, I. V., Chapleau, K. M., & Judd, C. M. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology*, 35(1), 59–68. <https://doi.org/10.1002/ejsp.232>
- Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The Influence of Afrocentric Facial Features in Criminal Sentencing. *Psychological Science*, 15(10), 674–679. <https://doi.org/10.1111/j.0956-7976.2004.00739.x>
- Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, 83(1), 5–25. <https://doi.org/10.1037/0022-3514.83.1.5>
- Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture Shapes How We Look at Faces (A. O. Holcombe, Ed.). *PLoS ONE*, 3(8), e3022. <https://doi.org/10.1371/journal.pone.0003022>
- Bond, R. (2025). Migration and Ethnicity in Scotland's 2022 Census. *Scottish Affairs*, 34(2), 246–261. <https://doi.org/10.3366/scot.2025.0548>
- Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1), 55–59. <https://doi.org/10.1016/j.anbehav.2011.03.024>
- Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, 78, 34–42. <https://doi.org/10.1016/j.jesp.2018.04.011>
- Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, 41(2), 135–143. <https://doi.org/10.1002/ejsp.744>
- Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, 28(1), 333–361. <https://doi.org/10.1080/10463283.2017.1381469>
- Brown, K. A., McKimmie, B. M., & Zarkadi, T. (2018). The Defendant with the Prison Tattoo: The Effect of Tattoos on Mock Jurors' Perceptions. *Psychiatry, Psychology and Law*, 25(3), 386–403. <https://doi.org/10.1080/13218719.2017.1412240>
- Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Univ of California Press.

- Burriss, R. P., Little, A. C., & Nelson, E. C. (2007). 2D:4D and Sexually Dimorphic Facial Characteristics. *Archives of Sexual Behavior*, 36(3), 377–384. <https://doi.org/10.1007/s10508-006-9136-1>
- Cabral, J. C. C., & de Almeida, R. M. M. (2019). Effects of anger on dominance-seeking and aggressive behaviors. *Evolution and Human Behavior*, 40(1), 23–33. <https://doi.org/10.1016/j.evolhumbehav.2018.07.006>
- Caldara, R. (2017). Culture Reveals a Flexible System for Face Processing. *Current Directions in Psychological Science*, 26(3), 249–255. <https://doi.org/10.1177/0963721417710036>
- Caldara, R., & Abdi, H. (2006). Simulating the ‘Other-Race’ Effect with Autoassociative Neural Networks: Further Evidence in Favor of the Face-Space Model. *Perception*, 35(5), 659–670. <https://doi.org/10.1068/p5360>
- Calvo, M. G., Álvarez-Plaza, P., & Fernández-Martín, A. (2017). The contribution of facial regions to judgements of happiness and trustworthiness from dynamic expressions. *Journal of Cognitive Psychology*, 29(5), 618–625. <https://doi.org/10.1080/20445911.2017.1302450>
- Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B: Biological Sciences*, 275(1651), 2651–2656. <https://doi.org/10.1098/rspb.2008.0873>
- Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial Structure Is a Reliable Cue of Aggressive Behavior. *Psychological Science*, 20(10), 1194–1198. <https://doi.org/10.1111/j.1467-9280.2009.02423.x>
- Cartei, V., Bond, R., & Reby, D. (2014). What makes a voice masculine: Physiological and acoustical correlates of women’s ratings of men’s vocal masculinity. *Hormones and Behavior*, 66(4), 569–576. <https://doi.org/10.1016/j.yhbeh.2014.08.006>
- Cave, S., & Dihal, K. (2020). The Whiteness of AI. *Philosophy & Technology*, 33(4), 685–703. <https://doi.org/10.1007/s13347-020-00415-6>
- Charbonneau, I., Robinson, K., Blais, C., & Fiset, D. (2020). Implicit race attitudes modulate visual information extraction for trustworthiness judgments (G. Merlhiot, Ed.). *PLOS ONE*, 15(9), e0239305. <https://doi.org/10.1371/journal.pone.0239305>
- Chen, C., Messinger, D. S., Chen, C., Yan, H., Duan, Y., Ince, R. A. A., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2024). Cultural facial expressions dynamically convey emotion category and intensity information. *Current Biology*, 34(1), 213–223.e5. <https://doi.org/10.1016/j.cub.2023.12.001>
- Chen, C., Lee, S.-y., & Stevenson, H. W. (1995). Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students. *Psychological Science*, 6(3), 170–175. <https://doi.org/10.1111/j.1467-9280.1995.tb00327.x>
- Chen, F. F., Jing, Y., Lee, J. M., & Bai, L. (2016). Culture Matters: The Looks of a Leader Are Not All the Same. *Social Psychological and Personality Science*, 7(6), 570–578. <https://doi.org/10.1177/1948550616644962>

- Chen, N., Yang, Y., Kobayashi, M., Nakamura, K., & Watanabe, K. (2024). Red Backgrounds Enhance Dominance in Human Faces and Shapes. *Evolutionary Psychology*, 22(3), 14747049241284602. <https://doi.org/10.1177/14747049241284602>
- Chua, K.-W., & Freeman, J. B. (2021). Facial Stereotype Bias Is Mitigated by Training. *Social Psychological and Personality Science*, 12(7), 1335–1344. <https://doi.org/10.1177/1948550620972550>
- Chua, K.-W., & Freeman, J. B. (2022). Learning to judge a book by its cover: Rapid acquisition of facial stereotypes. *Journal of Experimental Social Psychology*, 98, 104225. <https://doi.org/10.1016/j.jesp.2021.104225>
- Cogsdill, E. J., & Banaji, M. R. (2015). Face-trait inferences show robust child–adult agreement: Evidence from three types of faces. *Journal of Experimental Social Psychology*, 60, 150–156. <https://doi.org/10.1016/j.jesp.2015.05.007>
- Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring Character From Faces: A Developmental Study. *Psychological Science*, 25(5), 1132–1139. <https://doi.org/10.1177/0956797614523297>
- Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2021). Type I Error Is Inflated in the Two-Phase Reverse Correlation Procedure. *Social Psychological and Personality Science*, 12(5), 760–768. <https://doi.org/10.1177/1948550620938616>
- Cook, R., Eggleston, A., & Over, H. (2022). The cultural learning account of first impressions. *Trends in Cognitive Sciences*, 26(8), 656–668. <https://doi.org/10.1016/j.tics.2022.05.007>
- Cook, R., & Over, H. (2021). Why is the literature on first impressions so focused on White faces? *Royal Society Open Science*, 8(9), 211146. <https://doi.org/10.1098/rsos.211146>
- Cooper, R., & Gerlach, G. L. (2013). Measurement of the blush. In *The Psychological Significance of the Blush* (pp. 39–59). Cambridge University Press.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory* (14. print). Wiley.
- Craig, B. M., Koch, S., & Lipp, O. V. (2017). The influence of social category cues on the happy categorisation advantage depends on expression valence. *Cognition and Emotion*, 31(7), 1493–1501. <https://doi.org/10.1080/02699931.2016.1215293>
- Craig, B. M., Zhang, J., & Lipp, O. V. (2017). Facial race and sex cues have a comparable influence on emotion recognition in Chinese and Australian participants. *Attention, Perception, & Psychophysics*, 79(7), 2212–2223. <https://doi.org/10.3758/s13414-017-1364-z>
- Crookes, K., Ewing, L., Gildenhuis, J.-d., Kloth, N., Hayward, W. G., Oxner, M., Pond, S., & Rhodes, G. (2015). How Well Do Computer-Generated Faces Tap Face Expertise? (A. Key, Ed.). *PLOS ONE*, 10(11), e0141353. <https://doi.org/10.1371/journal.pone.0141353>
- Cuddy, A. J. C., Crotty, S., Chong, J., & Norton, M. I. (2014). Men as cultural ideals: How culture shapes gender stereotypes: (512142015-945). <https://doi.org/10.1037/e512142015-945>

- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92(4), 631–648. <https://doi.org/10.1037/0022-3514.92.4.631>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008, January). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In *Advances in Experimental Social Psychology* (pp. 61–149, Vol. 40). Academic Press. Retrieved September 24, 2024, from <https://www.sciencedirect.com/science/article/pii/S0065260107000020>
- Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleeboos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., ... Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1), 1–33. <https://doi.org/10.1348/014466608X314935>
- Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. *Computers & Education*, 190, 104607. <https://doi.org/10.1016/j.compedu.2022.104607>
- Darling-Hammond, S., Lee, R. T., & Mendoza-Denton, R. (2021). Interracial contact at work: Does workplace diversity reduce bias? *Group Processes & Intergroup Relations*, 24(7), 1114–1131. <https://doi.org/10.1177/1368430220932636>
- Darwin, C. (1872). *The expression of the emotions in man and animals*. University of Chicago Press.
- Davis, N., Olsen, N., Perry, V. G., Stewart, M. M., & White, T. B. (2023). I'm Only Human? The Role of Racial Stereotypes, Humanness, and Satisfaction in Transactions with Anthropomorphic Sales Bots. *Journal of the Association for Consumer Research*, 8(1), 47–58. <https://doi.org/10.1086/722703>
- Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2022). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods*, 54(4), 1889–1901. <https://doi.org/10.3758/s13428-021-01705-3>
- De Leersnyder, J., Mesquita, B., & Kim, H. S. (2011). Where Do My Emotions Belong? A Study of Immigrants' Emotional Acculturation. *Personality and Social Psychology Bulletin*, 37(4), 451–463. <https://doi.org/10.1177/0146167211399103>
- De Neys, W., Hopfensitz, A., & Bonnefon, J.-F. (2017). Split-Second Trustworthiness Detection From Faces in an Economic Game. *Experimental Psychology*, 64(4), 231–239. <https://doi.org/10.1027/1618-3169/a000367>
- Deska, J. C., Kunstman, J. W., Bernstein, M. J., Ogunbadero, T., & Hugenberg, K. (2020). Black racial phenotypicality shapes social pain and support judgments. *Journal of Experimental Social Psychology*, 90, 103998. <https://doi.org/10.1016/j.jesp.2020.103998>

- Devine, P. G., & Elliot, A. J. (1995). Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited. *Personality and Social Psychology Bulletin*, 21(11), 1139–1150. <https://doi.org/10.1177/01461672952111002>
- Di Natale, A. F., La Rocca, S., Simonetti, M. E., & Bricolo, E. (2024). Using computer-generated faces in experimental psychology: The role of realism and exposure. *Computers in Human Behavior Reports*, 14, 100397. <https://doi.org/10.1016/j.chbr.2024.100397>
- Donhauser, P. W., Florin, E., & Baillet, S. (2018). Imaging of neural oscillations with embedded inferential and group prevalence statistics (D. Marinazzo, Ed.). *PLOS Computational Biology*, 14(2), e1005990. <https://doi.org/10.1371/journal.pcbi.1005990>
- Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour*, 1(1), 0001. <https://doi.org/10.1038/s41562-016-0001>
- Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, 3(5), 562–571. <https://doi.org/10.1177/194855061143027>
- Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, 100(6), 999–1014. <https://doi.org/10.1037/a0023026>
- Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic Out-Group Faces Are Biased in the Prejudiced Mind. *Psychological Science*, 19(10), 978–980. <https://doi.org/10.1111/j.1467-9280.2008.02186.x>
- Drummond, P. D., & Quah, S. H. (2001). The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians. *Psychophysiology*, 38(2), 190–196. <https://doi.org/10.1111/1469-8986.3820190>
- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, 6(5), 178–190. [https://doi.org/10.1002/\(SICI\)1520-6505\(1998\)6:5<178::AID-EVAN5>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8)
- Dupuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *Journal of Vision*, 9(2), 10. <https://doi.org/10.1167/9.2.10>
- Durante, F., Fiske, S. T., Gelfand, M. J., Crippa, F., Suttora, C., Stillwell, A., Asbrock, F., Aycan, Z., Bye, H. H., Carlsson, R., Björklund, F., Dagher, M., Geller, A., Larsen, C. A., Latif, A.-H. A., Mähönen, T. A., Jasinskaja-Lahti, I., & Teymouri, A. (2017). Ambivalent stereotypes link to peace, conflict, and inequality across 38 nations. *Proceedings of the National Academy of Sciences*, 114(4), 669–674. <https://doi.org/10.1073/pnas.1611874114>
- Durkee, P. K., & Ayers, J. D. (2021). Is facial width-to-height ratio reliably associated with social inferences? *Evolution and Human Behavior*, 42(6), 583–592. <https://doi.org/10.1016/j.evolhumbehav.2021.06.003>
- Dzhelyova, M., Perrett, D. I., & Jentsch, I. (2012). Temporal dynamics of trustworthiness perception. *Brain Research*, 1435, 81–90. <https://doi.org/10.1016/j.brainres.2011.11.043>

- Eagly, A., & Steffen, V. (1984). Gender Stereotypes Stem from the Distribution of Women and Men in Social Roles. *Journal of Personality and Social Psychology*, 46, 735–754. <https://doi.org/10.1037/0022-3514.46.4.735>
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes. *Psychological Science*, 17(5), 383–386. <https://doi.org/10.1111/j.1467-9280.2006.01716.x>
- Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, Crime, and Visual Processing. *Journal of Personality and Social Psychology*, 87(6), 876–893. <https://doi.org/10.1037/0022-3514.87.6.876>
- Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System. *Environmental Psychology and Nonverbal Behavior*. <https://doi.org/10.1037/t27734-000>
- Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, 10(2), 159–168. <https://doi.org/10.1007/BF00992253>
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2), 203–235. <https://doi.org/10.1037/0033-2909.128.2.203>
- England, D. E., Descartes, L., & Collier-Meek, M. A. (2011). Gender Role Portrayal and the Disney Princesses. *Sex Roles*, 64(7), 555–567. <https://doi.org/10.1007/s11199-011-9930-7>
- Etcoff, N. L., Stock, S., Haley, L. E., Vickery, S. A., & House, D. M. (2011). Cosmetics as a Feature of the Extended Human Phenotype: Modulation of the Perception of Biologically Important Facial Signals. *PLOS ONE*, 6(10), e25656. <https://doi.org/10.1371/journal.pone.0025656>
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, 105(3), 482–498. <https://doi.org/10.1037/0033-295X.105.3.482>
- Farkas, L. G., Katic, M. J., & Forrest, C. R. (2005). International Anthropometric Study of Facial Morphology in Various Ethnic Groups/Races. *Journal of Craniofacial Surgery*, 16(4), 615. <https://doi.org/10.1097/01.scs.0000171847.58031.9e>
- FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences*, 115(7). <https://doi.org/10.1073/pnas.1715227115>
- Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and How Implicit First Impressions Can Be Updated. *Current Directions in Psychological Science*, 28(4), 331–336. <https://doi.org/10.1177/0963721419835206>
- Fernández-Dols, J.-M. (2013). Advances in the Study of Facial Expression: An Introduction to the Special Section. *Emotion Review*, 5(1), 3–7. <https://doi.org/10.1177/1754073912457209>
- Fetscherin, M., Tantleff-Dunn, S., & Klumb, A. (2020). Effects of facial features and styling elements on perceptions of competence, warmth, and hireability of male profession-

- als. *The Journal of Social Psychology*, 160(3), 332–345. <https://doi.org/10.1080/00224545.2019.1671300>
- Fink, B., & Matts, P. (2008). The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology*, 22(4), 493–498. <https://doi.org/10.1111/j.1468-3083.2007.02512.x>
- Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior*, 27(6), 433–442. <https://doi.org/10.1016/j.evolhumbehav.2006.08.007>
- Fink, B., Grammer, K., Mitteroecker, P., Gunz, P., Schaefer, K., Bookstein, F. L., & Manning, J. T. (2005). Second to fourth digit ratio and face shape. *Proceedings of the Royal Society B: Biological Sciences*, 272(1576), 1995–2001. <https://doi.org/10.1098/rspb.2005.3179>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 115(27). <https://doi.org/10.1073/pnas.1711978115>
- Fiske, S. T. (2017). Prejudices in Cultural Contexts: Shared Stereotypes (Gender, Age) Versus Variable Stereotypes (Race, Ethnicity, Religion). *Perspectives on Psychological Science*, 12(5), 791–799. <https://doi.org/10.1177/1745691617708204>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2022). Accuracy in Facial Trustworthiness Impressions: Kernel of Truth or Modern Physiognomy? A Meta-Analysis. *Personality and Social Psychology Bulletin*, 48(11), 1580–1596. <https://doi.org/10.1177/01461672211048110>
- Freeman, J. B., & Ambady, N. (2009). Motions of the Hand Expose the Partial and Parallel Activation of Stereotypes. *Psychological Science*, 20(10), 1183–1188. <https://doi.org/10.1111/j.1467-9280.2009.02422.x>
- Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, 118(2), 247–279. <https://doi.org/10.1037/a0022327>
- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, 137(4), 673–690. <https://doi.org/10.1037/a0013875>

- Freeman, J. B., & Johnson, K. L. (2016). More Than Meets the Eye: Split-Second Social Perception. *Trends in Cognitive Sciences*, 20(5), 362–374. <https://doi.org/10.1016/j.tics.2016.03.003>
- Freeman, J. B., Johnson, K. L., Ambady, N., & Rule, N. O. (2010). Sexual Orientation Perception Involves Gendered Facial Cues. *Personality and Social Psychology Bulletin*, 36(10), 1318–1331. <https://doi.org/10.1177/0146167210378755>
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the Part: Social Status Cues Shape Race Perception (S. Gilbert, Ed.). *PLoS ONE*, 6(9), e25107. <https://doi.org/10.1371/journal.pone.0025107>
- Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020, January). Chapter Five - Dynamic interactive theory as a domain-general account of social perception. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (pp. 237–287, Vol. 61). Academic Press. Retrieved September 26, 2024, from <https://www.sciencedirect.com/science/article/pii/S0065260119300346>
- Freeman, J. B., Stolier, R. M., Brooks, J. A., & Stillerman, B. S. (2018). The neural representational geometry of social perception. *Current Opinion in Psychology*, 24, 83–91. <https://doi.org/10.1016/j.copsyc.2018.10.003>
- Freeman, J. B., Stolier, R. M., Ingbreetsen, Z. A., & Hehman, E. A. (2014). Amygdala Responsivity to High-Level Social Information from Unseen Faces. *Journal of Neuroscience*, 34(32), 10573–10581. <https://doi.org/10.1523/JNEUROSCI.5063-13.2014>
- Freundenberg, M., Adams, R. B., Kleck, R. E., & Hess, U. (2015). Through a glass darkly: Facial wrinkles affect our processing of emotion in the elderly. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01476>
- Friedman, H., & Zebrowitz, L. A. (1992). The Contribution of Typical Sex Differences in Facial Maturity to Sex Role Stereotypes. *Personality and Social Psychology Bulletin*, 18(4), 430–438. <https://doi.org/10.1177/0146167292184006>
- Funk, F., & Todorov, A. (2013). Criminal Stereotypes in the Courtroom: Facial Tattoos Affect Guilt and Punishment Differently. *Psychology, Public Policy, and Law*, 19(4), 466–478. Retrieved October 7, 2024, from <https://heinonline.org/HOL/P?h=hein.journals/psypbclw19&i=476>
- Ganel, T., & Goodale, M. A. (2021). The effect of smiling on the perceived age of male and female faces across the lifespan. *Scientific Reports*, 11(1), 23020. <https://doi.org/10.1038/s41598-021-02380-2>
- Gasteiger, N., Hellou, M., & Ahn, H. S. (2021). Deploying social robots in museum settings: A quasi-systematic review exploring purpose and acceptability. *International Journal of Advanced Robotic Systems*, 18(6), 17298814211066740. <https://doi.org/10.1177/17298814211066740>
- Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from Meta-Analyses of the Facial Width-to-Height Ratio as an Evolved Cue of Threat (I. D. Stephen, Ed.). *PLOS ONE*, 10(7), e0132726. <https://doi.org/10.1371/journal.pone.0132726>

- Gibson, J. J. (1986). *The Ecological Approach to Visual Perception* (1st ed.). Psychology Press.
- Gifford, R. (1994). A Lens-Mapping Framework for Understanding the Encoding and Decoding of Interpersonal Dispositions in Nonverbal Behavior. *Journal of Personality and Social Psychology*, 66, 398–412. <https://doi.org/10.1037//0022-3514.66.2.398>
- Gill, D., Garrod, O. G. B., Jack, R. E., & Schyns, P. G. (2014). Facial Movements Strategically Camouflage Involuntary Social Signals of Face Morphology. *Psychological Science*, 25(5), 1079–1086. <https://doi.org/10.1177/0956797614522274>
- Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Gur, R. C., & Sachser, N. (2009). Baby Schema in Infant Faces Induces Cuteness Perception and Motivation for Caretaking in Adults. *Ethology*, 115(3), 257–263. <https://doi.org/10.1111/j.1439-0310.2008.01603.x>
- Graham, J. R., Harvey, C. R., & Puri, M. (2017). A Corporate Beauty Contest. *Management Science*, 63(9), 3044–3056. <https://doi.org/10.1287/mnsc.2016.2484>
- Grigoryev, D., Fiske, S. T., & Batkhina, A. (2019). Mapping Ethnic Stereotypes and Their Antecedents in Russia: The Stereotype Content Model. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01643>
- Guan, Y., Deng, H., & Bond, M. H. (2010). Examining Stereotype Content Model in a Chinese context: Inter-group structural relations and Mainland Chinese's stereotypes towards Hong Kong Chinese. *International Journal of Intercultural Relations*, 34(4), 393–399. <https://doi.org/10.1016/j.ijintrel.2010.04.003>
- Guimond, S., Crisp, R., Oliveira, P., Kamiejski, R., Kteily, N., Küpper, B., Lalonde, R., Levin, S., Pratto, F., Tougas, F., Sidanius, J., & Zick, A. (2013). Diversity Policy, Social Dominance, and Intergroup Relations: Predicting Prejudice in Changing Social and Political Contexts. *Journal of personality and social psychology*, 104. <https://doi.org/10.1037/a0032069>
- Hagiwara, N., Kashy, D. A., & Cesario, J. (2012). The independent effects of skin tone and facial features on Whites' affective reactions to Blacks. *Journal of Experimental Social Psychology*, 48(4), 892–898. <https://doi.org/10.1016/j.jesp.2012.02.001>
- Hall, E. T. (1966). *The hidden dimension* (1st edition). Doubleday.
- Han, C., Wang, H., Hahn, A. C., Fisher, C. I., Kandrik, M., Fasolt, V., Morrison, D. K., Lee, A. J., Holzleitner, I. J., DeBruine, L. M., & Jones, B. C. (2018). Cultural differences in preferences for facial coloration. *Evolution and Human Behavior*, 39(2), 154–159. <https://doi.org/10.1016/j.evolhumbehav.2017.11.005>
- Hanel, P. H. P., Maio, G. R., Soares, A. K. S., Vione, K. C., de Holanda Coelho, G. L., Gouveia, V. V., Patil, A. C., Kamble, S. V., & Manstead, A. S. R. (2018). Cross-Cultural Differences and Similarities in Human Value Instantiation. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00849>
- Hansen, K., Rakić, T., & Steffens, M. C. (2017). Competent and Warm?: How Mismatching Appearance and Accent Influence First Impressions. *Experimental Psychology*, 64(1), 27–36. <https://doi.org/10.1027/1618-3169/a000348>

- Hareli, S., Shomrat, N., & Hess, U. (2009). Emotional versus neutral expressions and perceptions of social dominance and submissiveness. *Emotion*, 9(3), 378–384. <https://doi.org/10.1037/a0015958>
- Hartung, F., Jamrozik, A., Rosen, M. E., Aguirre, G., Sarwer, D. B., & Chatterjee, A. (2019). Behavioural and Neural Responses to Facial Disfigurement. *Scientific Reports*, 9(1), 8021. <https://doi.org/10.1038/s41598-019-44408-8>
- Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's Facial Width-to-Height Ratio Predicts Aggression: A Meta-Analysis. *PLOS ONE*, 10(4), e0122637. <https://doi.org/10.1371/journal.pone.0122637>
- Haselhuhn, M. P., & Wong, E. M. (2011). Bad to the bone: Facial structure predicts unethical behaviour. *Proceedings of the Royal Society B: Biological Sciences*, 279(1728), 571–576. <https://doi.org/10.1098/rspb.2011.1193>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- Helman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, 49(4), 747–754. <https://doi.org/10.1016/j.jesp.2013.02.015>
- Helman, E., Stoller, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, 13(2), e12431. <https://doi.org/10.1111/spc3.12431>
- Helman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, 113(4), 513–529. <https://doi.org/10.1037/pspa0000090>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hensel, L. B., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2021). Facial expressions of social traits comprise core affective signals in Top Ranked Abstracts from the 2021 Annual Meeting of the Society for Affective Science. *2021 Annual Meeting of the Society for Affective Science*, 8, 264–272. Retrieved November 27, 2024, from https://scholar.google.com/citations?view_op=view_citation&hl=en&user=1R8GM7sAAAAJ&citation_for_view=1R8GM7sAAAAJ:IjCSPb-OGc4C
- Hensel, L. B., Zhan, J., Bjornsdottir, R. T., Garrod, O. G., Schyns, P. G., & Jack, R. E. (2020). Social trait perception is structured by a latent composition of 3D face features. *Journal of Vision*, 20(11), 1365. <https://doi.org/10.1167/jov.20.11.1365>
- Hess, U., Adams, R. B., & Kleck, R. E. (2009a). The Categorical Perception of Emotions and Traits. *Social Cognition*, 27(2), 320–326. <https://doi.org/10.1521/soco.2009.27.2.320>
- Hess, U., Adams, R. B., & Kleck, R. E. (2009b). The face is not an empty canvas: How facial expressions interact with facial appearance. *Philosophical Transactions of the Royal*

- Society B: Biological Sciences*, 364(1535), 3497–3504. <https://doi.org/10.1098/rstb.2009.0165>
- Hess, U., Huppertz, D., Mauersberger, H., & Kastendieck, T. (2023). Wrinkles are neither beautiful nor nice: The effect of facial wrinkles on person perception and interpersonal closeness. *Acta Psychologica*, 241, 104077. <https://doi.org/10.1016/j.actpsy.2023.104077>
- Hodges-Simeon, C. R., Hanson Sobraske, K. N., Samore, T., Gurven, M., & Gaulin, S. J. C. (2016). Facial Width-To-Height Ratio (fWHR) Is Not Associated with Adolescent Testosterone Levels (J. Mpodozis, Ed.). *PLOS ONE*, 11(4), e0153083. <https://doi.org/10.1371/journal.pone.0153083>
- Holder, A. M. B., Jackson, M. A., & Ponterotto, J. G. (2015). Racial microaggression experiences and coping strategies of Black women in corporate leadership. *Qualitative Psychology*, 2(2), 164–180. <https://doi.org/10.1037/qup0000024>
- Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 199–209. <https://doi.org/10.1037/0096-1523.32.2.199>
- Hong, Y., Chua, K.-W., & Freeman, J. B. (2024). Reducing Facial Stereotype Bias in Consequential Social Judgments: Intervention Success With White Male Faces. *Psychological Science*, 35(1), 21–33. <https://doi.org/10.1177/09567976231215238>
- Hu, Y., & O’Toole, A. J. (2023). First impressions: Integrating faces and bodies in personality trait perception. *Cognition*, 231, 105309. <https://doi.org/10.1016/j.cognition.2022.105309>
- Hu, Y., Parde, C. J., Hill, M. Q., Mahmood, N., & O’Toole, A. J. (2018). First Impressions of Personality Traits From Body Shapes. *Psychological Science*, 29(12), 1969–1983. <https://doi.org/10.1177/0956797618799300>
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160(1), 106–154.2. Retrieved July 16, 2024, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1359523/>
- Hugenberg, K. (2005). Social Categorization and the Perception of Facial Affect: Target Race Moderates the Response Latency Advantage for Happy Faces. *Emotion (Washington, D.C.)*, 5, 267–76. <https://doi.org/10.1037/1528-3542.5.3.267>
- Hugenberg, K., & Sacco, D. F. (2008). Social Categorization and Stereotyping: How Social Categorization Biases Person Perception and Face Memory. *Social and Personality Psychology Compass*, 2(2), 1052–1072. <https://doi.org/10.1111/j.1751-9004.2008.00090.x>
- Hutchings, R. J., Freiburger, E., Sim, M., & Hugenberg, K. (2024). Racial Prejudice Affects Representations of Facial Trustworthiness. *Psychological Science*, 35(3), 263–276. <https://doi.org/10.1177/09567976231225094>

- Ibrahimagić-Šeper, L., Čelebić, A., Petričević, N., & Selimović, E. (2006). Anthropometric differences between males and females in face dimensions and dimensions of central maxillary incisors. *Medicinski glasnik*, 3(2).
- Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00386>
- Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, 38(3), 1541–1573. <https://doi.org/10.1002/hbm.23471>
- Ince, R. A. A., Kay, J. W., & Schyns, P. G. (2022). Within-participant statistics for cognitive science. *Trends in Cognitive Sciences*, 26(8), 626–630. <https://doi.org/10.1016/j.tics.2022.05.008>
- Ince, R. A. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of population prevalence. *eLife*, 10, e62461. <https://doi.org/10.7554/eLife.62461>
- Ito, T. A., & Tomelleri, S. (2017). Seeing is not stereotyping: The functional independence of categorization and stereotype activation. *Social Cognitive and Affective Neuroscience*, 12(5), 758–764. <https://doi.org/10.1093/scan/nsx009>
- Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocorical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85(4), 616–626. <https://doi.org/10.1037/0022-3514.85.4.616>
- Jack, R. E. (2013). Culture and facial expressions of emotion. *Visual Cognition*, 21(9-10), 1248–1286. <https://doi.org/10.1080/13506285.2013.835367>
- 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(18), 1543–1548. <https://doi.org/10.1016/j.cub.2009.07.051>
- Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: General*, 141(1), 19–25. <https://doi.org/10.1037/a0023463>
- Jack, R. E., Crivelli, C., & Wheatley, T. (2018). Data-Driven Methods to Diversify Knowledge of Human Psychology. *Trends in Cognitive Sciences*, 22(1), 1–5. <https://doi.org/10.1016/j.tics.2017.10.002>
- Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19), 7241–7244. <https://doi.org/10.1073/pnas.1200155109>
- Jack, R. E., & Schyns, P. G. (2015). The Human Face as a Dynamic Tool for Social Communication. *Current Biology*, 25(14), R621–R634. <https://doi.org/10.1016/j.cub.2015.05.052>

- Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, 68(1), 269–297. <https://doi.org/10.1146/annurev-psych-010416-044242>
- Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General*, 145(6), 708–730. <https://doi.org/10.1037/xge0000162>
- Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General*, 148(6), 1008–1021. <https://doi.org/10.1037/xge0000591>
- Jaeger, B., & Jones, A. L. (2022). Which Facial Features Are Central in Impression Formation? *Social Psychological and Personality Science*, 13(2), 553–561. <https://doi.org/10.1177/19485506211034979>
- Jaeger, B., Todorov, A. T., Evans, A. M., & van Beest, I. (2020). Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. *Journal of Experimental Social Psychology*, 90, 104004. <https://doi.org/10.1016/j.jesp.2020.104004>
- James, E. A., Jenkins, S., & Watkins, C. D. (2018). Negative Effects of Makeup Use on Perceptions of Leadership Ability Across Two Ethnicities. *Perception*, 47(5), 540–549. <https://doi.org/10.1177/0301006618763263>
- Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, 102(1), 116–131. <https://doi.org/10.1037/a0025335>
- Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, 50, 169–179. <https://doi.org/10.1016/j.neubiorev.2014.10.009>
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The Relation Between Culture and Response Styles: Evidence From 19 Countries. *Journal of Cross-Cultural Psychology*, 36(2), 264–277. <https://doi.org/10.1177/0022022104272905>
- Johnston, R., Milne, A., Williams, C., & Hosie, J. (1997). Do Distinctive Faces Come from Outer Space? An Investigation of the Status of a Multidimensional Face-Space. *Visual Cognition*, 4(1), 59–67. <https://doi.org/10.1080/713756748>
- Jones, A. L., Porcheron, A., Sweda, J. R., Morizot, F., & Russell, R. (2016). Coloration in different areas of facial skin is a cue to health: The role of cheek redness and periorbital luminance in health perception. *Body Image*, 17, 57–66. <https://doi.org/10.1016/j.bodyim.2016.02.001>
- Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., ... Coles, N. A. (2021). To which world regions does the

- valence–dominance model of social perception apply? *Nature Human Behaviour*, 5(1), 159–169. <https://doi.org/10.1038/s41562-020-01007-2>
- Joo, J., Steen, F. F., & Zhu, S.-C. (2015). Automated Facial Trait Judgment and Election Outcome Prediction: Social Dimensions of Face. *2015 IEEE International Conference on Computer Vision (ICCV)*, 3712–3720. <https://doi.org/10.1109/ICCV.2015.423>
- Kaisler, R. E., & Leder, H. (2016). Trusting the Looks of Others: Gaze Effects of Faces in Social Settings. *Perception*, 45(8), 875–892. <https://doi.org/10.1177/0301006616643678>
- Kaltwasser, L., Moore, K., Weinreich, A., & Sommer, W. (2017). The influence of emotion type, social value orientation and processing focus on approach-avoidance tendencies to negative dynamic facial expressions. *Motivation and Emotion*, 41(4), 532–544. <https://doi.org/10.1007/s11031-017-9624-8>
- Kang, S. K., & Bodenhausen, G. V. (2015). Multiple Identities in Social Perception and Interaction: Challenges and Opportunities. *Annual Review of Psychology*, 66(1), 547–574. <https://doi.org/10.1146/annurev-psych-010814-015025>
- Kastanakis, M. N., & Voyer, B. G. (2014). The effect of culture on perception and cognition: A conceptual framework. *Journal of Business Research*, 67(4), 425–433. <https://doi.org/10.1016/j.jbusres.2013.03.028>
- Kätsyri, J., de Gelder, B., & Takala, T. (2019). Virtual Faces Evoke Only a Weak Uncanny Valley Effect: An Empirical Investigation With Controlled Virtual Face Images. *Perception*, 48(10), 968–991. <https://doi.org/10.1177/0301006619869134>
- Kawakami, K., Friesen, J. P., & Fang, X. (2022). Perceiving ingroup and outgroup faces within and across nations. *British Journal of Psychology*, 113(3), 551–574. <https://doi.org/10.1111/bjop.12563>
- Keating, C. F., Mazur, A., & Segall, M. H. (1977). Facial Gestures Which Influence the Perception of Status. *Sociometry*, 40(4), 374–378. <https://doi.org/10.2307/3033487>
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98(2), 155–163. <https://doi.org/10.1037/0033-295X.98.2.155>
- Keres, A., & Chartier, C. R. (2016). The Biasing Effects of Visual Background on Perceived Facial Trustworthiness.
- Kitayama, S., & Salvador, C. E. (2024). Cultural Psychology: Beyond East and West. *Annual Review of Psychology*, 75(1), 495–526. <https://doi.org/10.1146/annurev-psych-021723-063333>
- Kitayama, S., Salvador, C. E., Nanakdewa, K., Rossmaier, A., San Martin, A., & Savani, K. (2022). Varieties of interdependence and the emergence of the Modern West: Toward the globalizing of psychology. *American Psychologist*, 77(9), 991–1006. <https://doi.org/10.1037/amp0001073>
- Klapper, A., Dotsch, R., Van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology*, 111(5), 655–664. <https://doi.org/10.1037/pspa0000062>

- Klatt, J., Eimler, S. C., & Krämer, N. C. (2016). Makeup your mind: The impact of styling on perceived competence and warmth of female leaders. *The Journal of Social Psychology, 156*(5), 483–497. <https://doi.org/10.1080/00224545.2015.1129303>
- Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics, 52*(1), 18–36. <https://doi.org/10.3758/BF03206757>
- Kleider-Offutt, H. M., Bond, A. D., Williams, S. E., & Bohil, C. J. (2018). When a face type is perceived as threatening: Using general recognition theory to understand biased categorization of Afrocentric faces. *Memory & Cognition, 46*(5), 716–728. <https://doi.org/10.3758/s13421-018-0801-0>
- Kleisner, K., Tureček, P., Roberts, S. C., Havlíček, J., Valentova, J. V., Akoko, R. M., Leongómez, J. D., Apostol, S., Varella, M. A. C., & Saribay, S. A. (2021). How and why patterns of sexual dimorphism in human faces vary across the world. *Scientific Reports, 11*(1), 5978. <https://doi.org/10.1038/s41598-021-85402-3>
- Kline, M. A., Shamsudheen, R., & Broesch, T. (2018). Variation is the universal: Making cultural evolution work in developmental psychology. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1743), 20170059. <https://doi.org/10.1098/rstb.2017.0059>
- Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20*(3), 165–182. <https://doi.org/10.1007/BF02281954>
- Komori, M., Kawamura, S., & Ishihara, S. (2011). Multiple mechanisms in the perception of face gender: Effect of sex-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 626–633. <https://doi.org/10.1037/a0020369>
- Koppensteiner, M. (2013). Motion cues that make an impression: Predicting perceived personality by minimal motion information. *Journal of Experimental Social Psychology, 49*(6), 1137–1143. <https://doi.org/10.1016/j.jesp.2013.08.002>
- Koppensteiner, M., Stephan, P., & Jäschke, J. P. M. (2016). Moving speeches: Dominance, trustworthiness and competence in body motion. *Personality and Individual Differences, 94*, 101–106. <https://doi.org/10.1016/j.paid.2016.01.013>
- Kramer, R. S. S. (2017). Sexual dimorphism of facial width-to-height ratio in human skulls and faces: A meta-analytical approach [Place: Netherlands Publisher: Elsevier Science]. *Evolution and Human Behavior, 38*(3), 414–420. <https://doi.org/10.1016/j.evolhumbehav.2016.12.002>
- Kramer, R. S. S., Mileva, M., & Ritchie, K. L. (2018). Inter-rater agreement in trait judgments from faces (P. J. Hills, Ed.). *PLOS ONE, 13*(8), e0202655. <https://doi.org/10.1371/journal.pone.0202655>
- Kreibig, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology, 44*(5), 787–806. <https://doi.org/10.1111/j.1469-8986.2007.00550.x>
- Krys, K., Hansen, K., Xing, C., Szarota, P., & Yang, M.-m. (2014). Do Only Fools Smile at Strangers? Cultural Differences in Social Perception of Intelligence of Smiling

- Individuals. *Journal of Cross-Cultural Psychology*, 45(2), 314–321. <https://doi.org/10.1177/0022022113513922>
- Kubota, J. T., & Ito, T. (2017). Rapid race perception despite individuation and accuracy goals. *Social Neuroscience*, 12(4), 468–478. <https://doi.org/10.1080/17470919.2016.1182585>
- Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The Price of Racial Bias: Intergroup Negotiations in the Ultimatum Game. *Psychological Science*, 24(12), 2498–2504. <https://doi.org/10.1177/0956797613496435>
- Kyle, D. J., & Mahler, H. I. M. (1996). The Effects Of Hair Color And Cosmetic Use On Perceptions Of A Female's Ability. *Psychology of Women Quarterly*, 20(3), 447–455. <https://doi.org/10.1111/j.1471-6402.1996.tb00311.x>
- Lee, H., Shimizu, Y., & Uleman, J. S. (2015). Cultural Differences in the Automaticity of Elemental Impression Formation. *Social Cognition*, 33(1), 1–19. <https://doi.org/10.1521/soco.2015.33.1.1>
- Lee, R., Flavell, J. C., Tipper, S. P., Cook, R., & Over, H. (2021). Spontaneous first impressions emerge from brief training. *Scientific Reports*, 11(1), 15024. <https://doi.org/10.1038/s41598-021-94670-y>
- Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: Facial width is associated with testosterone levels in men. *Evolution and Human Behavior*, 34(4), 273–279. <https://doi.org/10.1016/j.evolhumbehav.2013.03.005>
- Lieberson, S. (2000, January). *A Matter of Taste: How Names, Fashions, and Culture Change*. Yale University Press.
- Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize comprehensive trait judgments of faces. *Nature Communications*, 12(1), 5168. Retrieved July 30, 2024, from <https://authors.library.caltech.edu/100792/2/10.31234%EF%80%A2osf.io%EF%80%A287nex.pdf>
- Lipp, O. V., Craig, B. M., & Dat, M. C. (2015). A Happy Face Advantage With Male Caucasian Faces: It Depends on the Company You Keep. *Social Psychological and Personality Science*, 6(1), 109–115. <https://doi.org/10.1177/1948550614546047>
- Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). The many faces of research on face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571), 1634–1637. <https://doi.org/10.1098/rstb.2010.0386>
- Little, A. C., Jones, B. C., Waite, C., Tiddeman, B. P., Feinberg, D. R., Perrett, D. I., Apicella, C. L., & Marlowe, F. W. (2008). Symmetry Is Related to Sexual Dimorphism in Faces: Data Across Culture and Species (T. Reimchen, Ed.). *PLoS ONE*, 3(5), e2106. <https://doi.org/10.1371/journal.pone.0002106>
- Little, A. C., & Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology*, 98(1), 111–126. <https://doi.org/10.1348/000712606X109648>
- Little, A. C., Roberts, S. C., Jones, B. C., & DeBruine, L. M. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions

- differently in wartime and peacetime scenarios. *Quarterly Journal of Experimental Psychology*, 65(10), 2018–2032. <https://doi.org/10.1080/17470218.2012.677048>
- Luo, X., Song, J., Guan, J., Wang, X., & Chen, L. (2024). Influence of facial dimorphism on interpersonal trust: Weighing warmth and competence traits in different trust situations. *Current Psychology*, 43(3), 2158–2172. <https://doi.org/10.1007/s12144-023-04472-w>
- Ma, D. S., Koltai, K., McManus, R. M., Bernhardt, A., Correll, J., & Wittenbrink, B. (2018). Race Signaling Features: Identifying Markers of Racial Prototypicality among Asians, Blacks, Latinos, and Whites. *Social Cognition*, 36(6), 603–625. <https://doi.org/10.1521/soco.2018.36.6.603>
- MacLin, O. H., & Malpass, R. S. (2001). Racial Categorization of Faces: The Ambiguous Race Face Effect Special Theme: The Other-Race Effect and Contemporary Criminal Justice: Eyewitness Identification and Jury Decision Making: Eyewitness Identification. *Psychology, Public Policy, and Law*, 7(1), 98–118. Retrieved October 9, 2024, from <https://heinonline.org/HOL/P?h=hein.journals/psypbclw7&i=100>
- Maddox, K. B. (2004). Perspectives on Racial Phenotypicality Bias. *Personality and Social Psychology Review*, 8(4), 383–401. https://doi.org/10.1207/s15327957pspr0804_4
- Maddox, K. B., Perry, J. M., & Pagan, J. (2022). Cues and categories: Revisiting paths to racial phenotypicality bias. *Social and Personality Psychology Compass*, 16(8), e12699. <https://doi.org/10.1111/spc3.12699>
- Maeng, A., Lee, H.-S., & Miyamoto, Y. (2022). Culture and trait inferences from facial cues. *Culture and Brain*, 10(1), 24–37. <https://doi.org/10.1007/s40167-022-00114-3>
- Mafra, A. L., de Moraes, Y. L., Varella, M. A. C., & Valentova, J. V. (2024). Makeup Usage in Women Is Positively Associated to Narcissism and Extraversion but Negatively to Psychopathy. *Archives of Sexual Behavior*. <https://doi.org/10.1007/s10508-024-02974-7>
- Main, J. C., Jones, B. C., DeBruine, L. M., & Little, A. C. (2009). Integrating Gaze Direction and Sexual Dimorphism of Face Shape When Perceiving the Dominance of Others. *Perception*, 38(9), 1275–1283. <https://doi.org/10.1068/p6347>
- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, 28(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4
- Manssuer, L. R., Pawling, R., Hayes, A. E., & Tipper, S. P. (2016). The role of emotion in learning trustworthiness from eye-gaze: Evidence from facial electromyography. *Cognitive Neuroscience*, 7(1-4), 82–102. <https://doi.org/10.1080/17588928.2015.1085374>
- Marsh, A. A., Adams, R. B., & Kleck, R. E. (2005). Why Do Fear and Anger Look the Way They Do? Form and Social Function in Facial Expressions. *Personality and Social Psychology Bulletin*, 31(1), 73–86. <https://doi.org/10.1177/0146167204271306>

- Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The Effects of Fear and Anger Facial Expressions on Approach- and Avoidance-Related Behaviors. *Emotion*, 5(1), 119–124. <https://doi.org/10.1037/1528-3542.5.1.119>
- Marsh, A. A., Elfenbein, H. A., & Ambady, N. (2003). Nonverbal “Accents”: Cultural Differences in Facial Expressions of Emotion. *Psychological Science*, 14(4), 373–376. <https://doi.org/10.1111/1467-9280.24461>
- Marshall, P., Bartolacci, A., & Burke, D. (2020). Human Face Tilt Is a Dynamic Social Signal That Affects Perceptions of Dimorphism, Attractiveness, and Dominance. *Evolutionary Psychology*, 18(1), 1474704920910403. <https://doi.org/10.1177/1474704920910403>
- Matsumoto, D. (1993). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion*, 17(2), 107–123. <https://doi.org/10.1007/BF00995188>
- Mattavelli, S., Masi, M., & Brambilla, M. (2022). Untrusted under threat: On the superior bond between trustworthiness and threat in face-context integration. *Cognition and Emotion*, 36(7), 1273–1286. <https://doi.org/10.1080/02699931.2022.2103100>
- Matts, P. J., Fink, B., Grammer, K., & Burquest, M. (2007). Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *Journal of the American Academy of Dermatology*, 57(6), 977–984. <https://doi.org/10.1016/j.jaad.2007.07.040>
- McAleer, P., Todorov, A., & Belin, P. (2014). How Do You Say ‘Hello’? Personality Impressions from Brief Novel Voices (C. R. Larson, Ed.). *PLoS ONE*, 9(3), e90779. <https://doi.org/10.1371/journal.pone.0090779>
- McArthur, L. Z., & Apatow, K. (1984). Impressions of Baby-Faced Adults. *Social Cognition*, 2(4), 315–342. <https://doi.org/10.1521/soco.1984.2.4.315>
- McArthur, L., & Baron, R. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90, 215–238. <https://doi.org/10.1037/0033-295X.90.3.215>
- McCormick, C. M., Mondloch, C. J., Carré, J. M., & Short, L. (2010). The Facial Width-to-Height Ratio as a Basis for Estimating Aggression from Emotionally Neutral Faces. *Journal of Vision*, 10(7), 599. <https://doi.org/10.1167/10.7.599>
- McKone, E., Dawel, A., Robbins, R. A., Shou, Y., Chen, N., & Crookes, K. (2023). Why the other-race effect matters: Poor recognition of other-race faces impacts everyday social interactions. *British Journal of Psychology*, 114(S1), 230–252. <https://doi.org/10.1111/bjop.12508>
- McManus, R. M., Young, L., & Sweetman, J. (2023). Psychology Is a Property of Persons, Not Averages or Distributions: Confronting the Group-to-Person Generalizability Problem in Experimental Psychology. *Advances in Methods and Practices in Psychological Science*, 6(3), 25152459231186615. <https://doi.org/10.1177/25152459231186615>
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100(2), 254–278. <https://doi.org/10.1037/0033-295X.100.2.254>

- Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, 8(6), 623–631. <https://doi.org/10.1093/scan/nss040>
- Menegatti, M., Pireddu, S., Crocetti, E., Moscatelli, S., & Rubini, M. (2021). The Ginevra de' Benci Effect: Competence, Morality, and Attractiveness Inferred From Faces Predict Hiring Decisions for Women. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.658424>
- Metinyurt, T., Haynes-Baratz, M. C., & Bond, M. A. (2021). A systematic review of interventions to address workplace bias: What we know, what we don't, and lessons learned. *New Ideas in Psychology*, 63, 100879. <https://doi.org/10.1016/j.newideapsych.2021.100879>
- Mezentseva, A. A., Rostovtseva, V. V., & Butovskaya, M. L. (2024). Facial cues to physical strength are not always associated with facial masculinity: Comparative study of Europeans and Southern Siberians. *American Journal of Human Biology*, 36(1), e23974. <https://doi.org/10.1002/ajhb.23974>
- Mignault, A., & Chaudhuri, A. (2003). The Many Faces of a Neutral Face: Head Tilt and Perception of Dominance and Emotion. *Journal of Nonverbal Behavior*, 27(2), 111–132. <https://doi.org/10.1023/A:1023914509763>
- Mileva, M., Kramer, R. S. S., & Burton, A. (2019). Social Evaluation of Faces Across Gender and Familiarity. *Perception*, 48(6), 471–486. <https://doi.org/10.1177/0301006619848996>
- Mileva, M., Young, A. W., Kramer, R. S. S., & Burton, A. M. (2019). Understanding facial impressions between and within identities. *Cognition*, 190, 184–198. <https://doi.org/10.1016/j.cognition.2019.04.027>
- Mileva, V. R., Cowan, M. L., Cobey, K. D., Knowles, K. K., & Little, A. C. (2014). In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men. *Personality and Individual Differences*, 69, 115–118. <https://doi.org/10.1016/j.paid.2014.05.019>
- Mileva, V. R., Jones, A. L., Russell, R., & Little, A. C. (2016). Sex Differences in the Perceived Dominance and Prestige of Women With and Without Cosmetics. *Perception*, 45(10), 1166–1183. <https://doi.org/10.1177/0301006616652053>
- Mo, C., Cristofori, I., Lio, G., Gomez, A., Duhamel, J.-R., Qu, C., & Sirigu, A. (2022). Culture-free perceptual invariant for trustworthiness. *PLOS ONE*, 17(2), e0263348. <https://doi.org/10.1371/journal.pone.0263348>
- Montepare, J. M., & Dobish, H. (2003). The Contribution of Emotion Perceptions and Their Overgeneralizations to Trait Impressions. *Journal of Nonverbal Behavior*, 27(4), 237–254. <https://doi.org/10.1023/A:1027332800296>
- Morrison, D., Wang, H., Hahn, A. C., Jones, B. C., & DeBruine, L. M. (2017). Predicting the reward value of faces and bodies from social perception (A. Avenanti, Ed.). *PLOS ONE*, 12(9), e0185093. <https://doi.org/10.1371/journal.pone.0185093>

- Mulhern, R., Fieldman, G., Hussey, T., L  v  que, J.-L., & Pineau, P. (2003). Do cosmetics enhance female Caucasian facial attractiveness? *International Journal of Cosmetic Science*, 25(4), 199–205. <https://doi.org/10.1046/j.1467-2494.2003.00188.x>
- Na, J., Kim, S., Oh, H., Choi, I., & O’Toole, A. (2015). Competence Judgments Based on Facial Appearance Are Better Predictors of American Elections Than of Korean Elections. *Psychological Science*, 26(7), 1107–1113. <https://doi.org/10.1177/0956797615576489>
- Na, J., & Kitayama, S. (2011). Spontaneous Trait Inference Is Culture-Specific: Behavioral and Neural Evidence. *Psychological Science*, 22(8), 1025–1032. <https://doi.org/10.1177/0956797611414727>
- Nestler, S., & Back, M. D. (2013). Applications and Extensions of the Lens Model to Understand Interpersonal Judgments at Zero Acquaintance. *Current Directions in Psychological Science*, 22(5), 374–379. <https://doi.org/10.1177/0963721413486148>
- Nestor, A., Plaut, D. C., & Behrmann, M. (2016). Feature-based face representations and image reconstruction from behavioral and neural data. *Proceedings of the National Academy of Sciences*, 113(2), 416–421. <https://doi.org/10.1073/pnas.1514551112>
- Nestor, A., & Tarr, M. J. (2008). Gender Recognition of Human Faces Using Color. *Psychological Science*, 19(12), 1242–1246. <https://doi.org/10.1111/j.1467-9280.2008.02232.x>
- Newell, A. (1998). You Can’t Play 20 Questions with Nature and Win: Projective Comments on the Papers of This Symposium. In *Machine Intelligence*. Routledge.
- Ng, M. H. S., & Cheng, C.-Y. (2024). The social evaluation of accents and perceived social influence in Singapore: A comparison of American and Singaporean English accents. *Asian Journal of Social Psychology*, 27(3), 376–390. <https://doi.org/10.1111/ajsp.12600>
- Ng, S. Y., Zebrowitz, L. A., & Franklin, R. G. (2016). Age Differences in the Differentiation of Trait Impressions From Faces. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 71(2), 220–229. <https://doi.org/10.1093/geronb/gbu113>
- Nicolas, G., de la Fuente, M., & Fiske, S. T. (2017). Mind the overlap in multiple categorization: A review of crossed categorization, intersectionality, and multiracial perception. *Group Processes & Intergroup Relations*, 20(5), 621–631. <https://doi.org/10.1177/1368430217708862>
- Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, 33(6), 417–433. <https://doi.org/10.1017/S0140525X10000865>
- Nomura, M., Ito, T., & Ding, S. (2024). Towards Collaborative Brain-storming among Humans and AI Agents: An Implementation of the IBIS-based Brainstorming Support System with Multiple AI Agents. *Proceedings of the ACM Collective Intelligence Conference*, 1–9. <https://doi.org/10.1145/3643562.3672609>

- Norman, J. B., Castro Lingl, D., Hehman, E., & Chen, J. M. (2024). Race in the eye of the beholder: Decomposing perceiver- and target-level variation in perceived racial prototypicality. *Journal of Experimental Social Psychology*, 115, 104667. <https://doi.org/10.1016/j.jesp.2024.104667>
- Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing Hidden Gender Biases in Competence Impressions of Faces. *Psychological Science*, 30(1), 65–79. <https://doi.org/10.1177/0956797618813092>
- Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, 149(2), 323–342. <https://doi.org/10.1037/xge0000638>
- Oh, D., Dotsch, R., & Todorov, A. (2019). Contributions of shape and reflectance information to social judgments from faces. *Vision Research*, 165, 131–142. <https://doi.org/10.1016/j.visres.2019.10.010>
- Oh, D., Shafir, E., & Todorov, A. (2020). Economic status cues from clothes affect perceived competence from faces. *Nature Human Behaviour*, 4(3), 287–293. <https://doi.org/10.1038/s41562-019-0782-4>
- Okubo, M., Ishikawa, K., Kobayashi, A., & Suzuki, H. (2017). Can I Trust You? Laterality of Facial Trustworthiness in an Economic Game. *Journal of Nonverbal Behavior*, 41(1), 21–34. <https://doi.org/10.1007/s10919-016-0242-z>
- Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, 49(5), 888–902. <https://doi.org/10.1002/ejsp.2569>
- Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2020). Good to Bad or Bad to Bad? What is the relationship between valence and the trait content of the Big Two? *European Journal of Social Psychology*, 50(2), 463–483. <https://doi.org/10.1002/ejsp.2618>
- Olivola, C. Y., Eubanks, D. L., & Lovelace, J. B. (2014). The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *The Leadership Quarterly*, 25(5), 817–834. <https://doi.org/10.1016/j.leaqua.2014.06.002>
- Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans Prefer Republican-Looking Leaders: Political Facial Stereotypes Predict Candidate Electoral Success Among Right-Leaning Voters. *Social Psychological and Personality Science*, 3(5), 605–613. <https://doi.org/10.1177/1948550611432770>
- Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, 46(2), 315–324. <https://doi.org/10.1016/j.jesp.2009.12.002>
- Olszanowski, M., Parzuchowski, M., & Szymków, A. (2019). When the smile is not enough: The interactive role of smiling and facial characteristics in forming judgments about trustworthiness and dominance. *Roczniki Psychologiczne*, 22, 35–52. <https://doi.org/10.18290/rpsych.2019.22.1-3>

- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition*, 170, 190–200. <https://doi.org/10.1016/j.cognition.2017.10.002>
- Over, H., Eggleston, A., & Cook, R. (2020). Ritual and the origins of first impressions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1805), 20190435. <https://doi.org/10.1098/rstb.2019.0435>
- Özener, B. (2012). Facial width-to-height ratio in a Turkish population is not sexually dimorphic and is unrelated to aggressive behavior. *Evolution and Human Behavior*, 33(3), 169–173. <https://doi.org/10.1016/j.evolhumbehav.2011.08.001>
- Park, B., Tsai, J. L., Chim, L., Blevins, E., & Knutson, B. (2016). Neural evidence for cultural differences in the valuation of positive facial expressions. *Social Cognitive and Affective Neuroscience*, 11(2), 243–252. <https://doi.org/10.1093/scan/nsv113>
- Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality Judgments from Natural and Composite Facial Images: More Evidence For A “Kernel Of Truth” In Social Perception. *Social Cognition*, 24(5), 607–640. <https://doi.org/10.1521/soco.2006.24.5.607>
- Pereira Santos, C., Relouw, J., Hutchinson-Lhuissier, K., Buggenum, A. V., Boudry, A., Fransen, A., Ven, M. V. D., & Mayer, I. (2023). Embodied Agents for Obstetric Simulation Training. *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 515–527. <https://doi.org/10.1145/3581641.3584100>
- Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences*, 119(17), e2115228119. <https://doi.org/10.1073/pnas.2115228119>
- Petsko, C. D., & Bodenhausen, G. V. (2020). Multifarious person perception: How social perceivers manage the complexity of intersectional targets. *Social and Personality Psychology Compass*, 14(2), e12518. <https://doi.org/10.1111/spc3.12518>
- Ponsi, G., Panasiti, M. S., Scandola, M., & Aglioti, S. M. (2016). Influence of warmth and competence on the promotion of safe in-group selection: Stereotype content model and social categorization of faces. *Quarterly Journal of Experimental Psychology*, 69(8), 1464–1479. <https://doi.org/10.1080/17470218.2015.1084339>
- Porcheron, A., Latreille, J., Jdid, R., Tschachler, E., & Morizot, F. (2014). Influence of skin ageing features on Chinese women’s perception of facial age and attractiveness. *International Journal of Cosmetic Science*, 36(4), 312–320. <https://doi.org/10.1111/ics.12128>
- Porcheron, A., Mauger, E., & Russell, R. (2013). Aspects of Facial Contrast Decrease with Age and Are Cues for Age Perception (H. P. O. De Beeck, Ed.). *PLoS ONE*, 8(3), e57985. <https://doi.org/10.1371/journal.pone.0057985>

- Puccetti, G., Nguyen, T., & Stroever, C. (2011). Skin colorimetric parameters involved in skin age perception. *Skin Research and Technology*, 17(2), 129–134. <https://doi.org/10.1111/j.1600-0846.2010.00480.x>
- Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, 28(5), 340–344. <https://doi.org/10.1016/j.evolhumbehav.2007.05.002>
- Qian, M. K., Heyman, G. D., Quinn, P. C., Messi, F. A., Fu, G., & Lee, K. (2016). Implicit Racial Biases in Preschool Children and Adults From Asia and Africa. *Child Development*, 87(1), 285–296. <https://doi.org/10.1111/cdev.12442>
- Quist, M. C., Watkins, C. D., Smith, F. G., DeBruine, L. M., & Jones, B. C. (2011). Facial masculinity is a cue to women's dominance. *Personality and Individual Differences*, 50(7), 1089–1093. <https://doi.org/10.1016/j.paid.2011.01.032>
- Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of *Homo sapiens* : Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, 115(45), 11401–11405. <https://doi.org/10.1073/pnas.1721165115>
- Reeves, B., Hancock, J., & Liu, X. (2020). Social robots are like real people: First impressions, attributes, and stereotyping of social robots. *Technology, Mind, and Behavior*, 1(1). <https://doi.org/10.1037/tmb0000018>
- Resenhoeft, A., Villa, J., & Wiseman, D. (2008). Tattoos Can Harm Perceptions: A Study and Suggestions. *Journal of American College Health*, 56(5), 593–596. <https://doi.org/10.3200/JACH.56.5.593-596>
- Rhodes, G. (2006). The Evolutionary Psychology of Facial Beauty. *Annual Review of Psychology*, 57(1), 199–226. <https://doi.org/10.1146/annurev.psych.57.102904.190208>
- Rhodes, G., Locke, V., Ewing, L., & Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception*, 38(2), 232–241. <https://doi.org/10.1068/p6110>
- Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior*, 24(6), 2741–2756. <https://doi.org/10.1016/j.chb.2008.03.017>
- Rossen, B., Johnsen, K., Deladisma, A., Lind, S., & Lok, B. (2008). Virtual Humans Elicit Skin-Tone Bias Consistent with Real-World Skin-Tone Biases. In H. Prendinger, J. Lester, & M. Ishizuka (Eds.), *Intelligent Virtual Agents* (pp. 237–244). Springer. https://doi.org/10.1007/978-3-540-85483-8_24
- Rostovtseva, V. V., Butovskaya, M. L., Mezentseva, A. A., Dashieva, N. B., Korotkova, A. A., Kavina, A., & Singh, M. (2024). Cross-Cultural Differences in Perception of Facial Trustworthiness Based on Geometric Morphometric Morphs. *Journal of Cross-Cultural Psychology*, 55(2), 216–235. <https://doi.org/10.1177/00220221231220013>

- Rostovtseva, V. V., Mezentseva, A. A., Windhager, S., & Butovskaya, M. L. (2020). Second-to-fourth digit ratio and facial shape in Buryats of Southern Siberia. *Early Human Development*, 149, 105138. <https://doi.org/10.1016/j.earlhumdev.2020.105138>
- Rule, N. O., Adams, R. B., Ambady, N., & Freeman, J. B. (2012). Perceptions of Dominance following Glimpses of Faces and Bodies. *Perception*, 41(6), 687–706. <https://doi.org/10.1068/p7023>
- Rule, N. O., & Ambady, N. (2008). The Face of Success: Inferences From Chief Executive Officers' Appearance Predict Company Profits. *Psychological Science*, 19(2), 109–111. <https://doi.org/10.1111/j.1467-9280.2008.02054.x>
- Rule, N. O., & Ambady, N. (2009). She's Got the Look: Inferences from Female Chief Executive Officers' Faces Predict their Success. *Sex Roles*, 61(9), 644–652. <https://doi.org/10.1007/s11199-009-9658-9>
- Rule, N. O., Ambady, N., Adams, R. B., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. *Journal of Personality and Social Psychology*, 98(1), 1–15. <https://doi.org/10.1037/a0017673>
- Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., & Niedenthal, P. M. (2017). Functional Smiles: Tools for Love, Sympathy, and War. *Psychological Science*, 28(9), 1259–1270. <https://doi.org/10.1177/0956797617706082>
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, 9(2), 260–264. <https://doi.org/10.1037/a0014681>
- Salatiello, A., Hovaidi-Ardestani, M., & Giese, M. A. (2021). A Dynamical Generative Model of Social Interactions. *Frontiers in Neurorobotics*, 15. <https://doi.org/10.3389/fnbot.2021.648527>
- Salvador, C., Carlier, S. I., Ishii, K., Castillo, C. T., Nanakdewa, K., Alvaro, S. M., Savani, K., & Kitayama, S. (2020, September). Emotionally Expressive Interdependence in Latin America: Triangulating Through a Comparison of Three Cultural Regions. <https://doi.org/10.31234/osf.io/pw4yk>
- Samson, N., Fink, B., & Matts, P. (2011). Does a Woman's Skin Color Indicate Her Fertility Level?: Preliminary Findings. *Swiss Journal of Psychology*, 70(4), 199–202. <https://doi.org/10.1024/1421-0185/a000057>
- Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, 8(4), 467–487. <https://doi.org/10.1002/ejsp.2420080405>
- Schneider, Z., & Morón, M. (2023). Facial makeup and perceived likelihood of influence tactics use among women: A role of attractiveness attributed to faces with and without makeup. *Current Psychology*, 42(26), 22564–22575. <https://doi.org/10.1007/s12144-022-03373-8>
- Schwartz, S. H. (2014, January). Chapter 20 - National Culture as Value Orientations: Consequences of Value Differences and Cultural Distance. In V. A. Ginsburgh & D.

- Throsby (Eds.), *Handbook of the Economics of Art and Culture* (pp. 547–586, Vol. 2). Elsevier. Retrieved October 30, 2024, from <https://www.sciencedirect.com/science/article/pii/B9780444537768000209>
- Schyns, P. G., Snoek, L., & Daube, C. (2022). Degrees of algorithmic equivalence between the brain and its DNN models. *Trends in Cognitive Sciences*, 26(12), 1090–1102. <https://doi.org/10.1016/j.tics.2022.09.003>
- Scott-Phillips, T. C. (2008). Defining biological communication. *Journal of Evolutionary Biology*, 21(2), 387–395. <https://doi.org/10.1111/j.1420-9101.2007.01497.x>
- Senese, V. P., De Falco, S., Bornstein, M. H., Caria, A., Buffolino, S., & Venuti, P. (2013). Human Infant Faces Provoke Implicit Positive Affective Responses in Parents and Non-Parents Alike (M. Costantini, Ed.). *PLoS ONE*, 8(11), e80379. <https://doi.org/10.1371/journal.pone.0080379>
- Shamekhi, A., Liao, Q. V., Wang, D., Bellamy, R. K. E., & Erickson, T. (2018). Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3173574.3173965>
- Sharmin, T., & Sattar, S. (2018). Gender Politics in the Projection of “Disney” Villains. *Journal of Literature and Art Studies*, 8(1). <https://doi.org/10.17265/2159-5836/2018.01.006>
- Shimizu, Y. (2012). Spontaneous trait inferences among Japanese children and adults: A developmental approach. *Asian Journal of Social Psychology*, 15(2), 112–121. <https://doi.org/10.1111/j.1467-839X.2012.01370.x>
- Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, 69, 79–85. <https://doi.org/10.1016/j.jesp.2016.08.003>
- Silvestri, V., Arioli, M., Baccolo, E., & Cassia, V. M. (2022). Sensitivity to trustworthiness cues in own- and other-race faces: The role of spatial frequency information. *PLOS ONE*, 17(9), e0272256. <https://doi.org/10.1371/journal.pone.0272256>
- Slepian, M. L., & Ames, D. R. (2016). Internalized Impressions: The Link Between Apparent Facial Trustworthiness and Deceptive Behavior Is Mediated by Targets’ Expectations of How They Will Be Judged. *Psychological Science*, 27(2), 282–288. <https://doi.org/10.1177/0956797615594897>
- Sobieszek, A., Siemiątkowski, M., & Imbir, K. K. (2024). Generative neural networks for experimental manipulation: Examining dominance-trustworthiness face impressions with data-efficient models. *British Journal of Psychology*, n/a(n/a). <https://doi.org/10.1111/bjop.12732>
- Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H. J., & Todorov, A. (2017). For Your Local Eyes Only: Culture-Specific Face Typicality Influences Perceptions of Trustworthiness. *Perception*, 46(8), 914–928. <https://doi.org/10.1177/0301006617691786>

- Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What Is Typical Is Good: The Influence of Face Typicality on Perceived Trustworthiness. *Psychological Science*, 26(1), 39–47. <https://doi.org/10.1177/0956797614554955>
- Song, Y., & Luximon, Y. (2024). When Trustworthiness Meets Face: Facial Design for Social Robots. *Sensors (Basel, Switzerland)*, 24(13), 4215. <https://doi.org/10.3390/s24134215>
- South Palomares, J. K., & Young, A. W. (2018). Facial First Impressions of Partner Preference Traits: Trustworthiness, Status, and Attractiveness. *Social Psychological and Personality Science*, 9(8), 990–1000. <https://doi.org/10.1177/1948550617732388>
- Speelman, C. P., & McGann, M. (2013). How Mean is the Mean? *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00451>
- Speelman, C. P., & McGann, M. (2020). Statements About the Pervasiveness of Behavior Require Data About the Pervasiveness of Behavior. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.594675>
- Stanciu, A. (2015). Four Sub-Dimensions of Stereotype Content: Explanatory Evidence from Romania. *International Psychology Bulletin*, 19, 14–20.
- Stanciu, A., Cohrs, J. C., Hanke, K., & Gavreliuc, A. (2017). Within-culture variation in the content of stereotypes: Application and development of the stereotype content model in an Eastern European culture. *The Journal of Social Psychology*, 157(5), 611–628. <https://doi.org/10.1080/00224545.2016.1262812>
- Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, 108(19), 7710–7715. <https://doi.org/10.1073/pnas.1014345108>
- Stepanova, E. V., & Strube, M. J. (2012). The role of skin color and facial physiognomy in racial categorization: Moderation by implicit racial attitudes. *Journal of Experimental Social Psychology*, 48(4), 867–878. <https://doi.org/10.1016/j.jesp.2012.02.019>
- Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General*, 141(4), 715–727. <https://doi.org/10.1037/a0027950>
- Stins, J. F., Roelofs, K., Villan, J., Kooijman, K., Hagenaars, M. A., & Beek, P. J. (2011). Walk to me when I smile, step back when I'm angry: Emotional faces modulate whole-body approach–avoidance behaviors. *Experimental Brain Research*, 212(4), 603–611. <https://doi.org/10.1007/s00221-011-2767-z>
- Stirrat, M., & Perrett, D. (2010). Valid Facial Cues to Cooperation and Trust: Male Facial Width and Trustworthiness. *Psychological Science*, 21(3), 349–354. <https://doi.org/10.1177/0956797610362647>
- Stolier, R. M., Hehman, E., & Freeman, J. B. (2020, January). Trait knowledge forms a common structure across social cognition | Nature Human Behaviour. Retrieved September 26, 2024, from <https://www.nature.com/articles/s41562-019-0800-6>

- Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, 115(37), 9210–9215. <https://doi.org/10.1073/pnas.1807222115>
- Strom, M. A., Zebrowitz, L. A., Zhang, S., Bronstad, P. M., & Lee, H. K. (2012). Skin and Bones: The Contribution of Skin Tone and Facial Structure to Racial Prototypicality Ratings (G. Yovel, Ed.). *PLoS ONE*, 7(7), e41193. <https://doi.org/10.1371/journal.pone.0041193>
- Sutherland, C. A. M., Burton, N. S., Wilmer, J. B., Blokland, G. A. M., Germine, L., Palermo, R., Collova, J. R., & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences*, 117(19), 10218–10224. <https://doi.org/10.1073/pnas.1920131117>
- Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial First Impressions Across Culture: Data-Driven Modeling of Chinese and British Perceivers' Unconstrained Facial Impressions. *Personality and Social Psychology Bulletin*, 44(4), 521–537. <https://doi.org/10.1177/0146167217744194>
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118. <https://doi.org/10.1016/j.cognition.2012.12.001>
- Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, 157, 257–267. <https://doi.org/10.1016/j.cognition.2016.09.006>
- Sutherland, C. A. M., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology*, 113(4), 1056–1078. <https://doi.org/10.1111/bjop.12583>
- Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, 106(2), 186–208. <https://doi.org/10.1111/bjop.12085>
- Sutherland, C. A. M., Young, A. W., & Rhodes, G. (2017). Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *British Journal of Psychology*, 108(2), 397–415. <https://doi.org/10.1111/bjop.12206>
- Swe, D. C., Palermo, R., Gwinn, O. S., Bell, J., Nakanishi, A., Collova, J., & Sutherland, C. A. M. (2022). Trustworthiness perception is mandatory: Task instructions do not modulate fast periodic visual stimulation trustworthiness responses. *Journal of Vision*, 22(11), 17. <https://doi.org/10.1167/jov.22.11.17>
- Swe, D. C., Palermo, R., Gwinn, O. S., Rhodes, G., Neumann, M., Payart, S., & Sutherland, C. A. M. (2020). An objective and reliable electrophysiological marker for implicit trustworthiness perception. *Social Cognitive and Affective Neuroscience*, 15(3), 337–346. <https://doi.org/10.1093/scan/nsaa043>

- Takeda, M. B., Helms, M. M., & Romanova, N. (2006). Hair Color Stereotyping and CEO Selection in the United Kingdom. *Journal of Human Behavior in the Social Environment*, 13(3), 85–99. https://doi.org/10.1300/J137v13n03_06
- Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, 93(1), B1–B9. <https://doi.org/10.1016/j.cognition.2003.09.011>
- Tang, D., & Schmeichel, B. J. (2015). Look Me in the Eye: Manipulated Eye Gaze Affects Dominance Mindsets. *Journal of Nonverbal Behavior*, 39(2), 181–194. <https://doi.org/10.1007/s10919-015-0206-8>
- Thompson, W., Owen, J., De St. Germain, H., Stark, S., & Henderson, T. (1999). Feature-based reverse engineering of mechanical parts. *IEEE Transactions on Robotics and Automation*, 15(1), 57–66. <https://doi.org/10.1109/70.744602>
- Thorstenson, C. A. (2018). The Social Psychophysics of Human Face Color: Review and Recommendations. *Social Cognition*, 36(2), 247–273. <https://doi.org/10.1521/soco.2018.36.2.247>
- Thorstenson, C. A., Elliot, A. J., Pazda, A. D., Perrett, D. I., & Xiao, D. (2018). Emotion-color associations in the context of the face. *Emotion*, 18(7), 1032–1042. <https://doi.org/10.1037/emo0000358>
- Thorstenson, C. A., & Pazda, A. D. (2021). Facial coloration influences social approach-avoidance through social perception. *Cognition and Emotion*, 35(5), 970–985. <https://doi.org/10.1080/02699931.2021.1914554>
- Thorstenson, C. A., Pazda, A. D., Young, S. G., & Elliot, A. J. (2019). Face color facilitates the disambiguation of confusing emotion expressions: Toward a social functional account of face color in emotion communication. *Emotion*, 19(5), 799–807. <https://doi.org/10.1037/emo0000485>
- Timme, N., Alford, W., Flecker, B., & Beggs, J. M. (2014). Synergy, redundancy, and multivariate information measures: An experimentalist's perspective. *Journal of Computational Neuroscience*, 36(2), 119–140. <https://doi.org/10.1007/s10827-013-0458-4>
- Timming, A. R., Nickson, D., Re, D., & Perrett, D. (2017). What Do You Think of My Ink? Assessing the Effects of Body Art on Employment Chances. *Human Resource Management*, 56(1), 133–149. <https://doi.org/10.1002/hrm.21770>
- Timming, A. R., & Perrett, D. (2016). Trust and mixed signals: A study of religion, tattoos and cognitive dissonance. *Personality and Individual Differences*, 97, 234–238. <https://doi.org/10.1016/j.paid.2016.03.067>
- Tinbergen, N. (1948). Social Releasers and the Experimental Method Required for Their Study. *The Wilson Bulletin*, 60(1), 6–51. Retrieved July 16, 2024, from <https://www.jstor.org/stable/4157642>
- Titze, I. R., & Martin, D. W. (1998). Principles of Voice Production. *The Journal of the Acoustical Society of America*, 104(3), 1148. <https://doi.org/10.1121/1.424266>

- Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven Methods for Modeling Social Perception. *Social and Personality Psychology Compass*, 5(10), 775–791. <https://doi.org/10.1111/j.1751-9004.2011.00389.x>
- Todorov, A., Olivola, C., Dotsch, R., & Mende-Siedlecki, P. (2014). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual review of psychology*, 66. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, 66(Volume 66, 2015), 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., & Oosterhof, N. (2011). Modeling Social Perception of Faces [Social Sciences]. *IEEE Signal Processing Magazine*, 28(2), 117–122. <https://doi.org/10.1109/MSP.2010.940006>
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating Faces on Trustworthiness After Minimal Time Exposure. *Social Cognition*, 27(6), 813–833. <https://doi.org/10.1521/soco.2009.27.6.813>
- Torrance, J. S., Holzleitner, I. J., Lee, A. J., DeBruine, L. M., & Jones, B. C. (2020). Evidence Head Tilt Has Dissociable Effects on Dominance and Trustworthiness Judgments, But Does Not Have Category-Contingent Effects on Hypothetical Leadership Judgments. *Perception*, 49(2), 199–209. <https://doi.org/10.1177/0301006619898589>
- Toscano, H., Schubert, T. W., Dotsch, R., Falvello, V., & Todorov, A. (2016). Physical Strength as a Cue to Dominance: A Data-Driven Approach. *Personality and Social Psychology Bulletin*, 42(12), 1603–1616. <https://doi.org/10.1177/0146167216666266>
- Toscano, H., Schubert, T. W., & Giessner, S. R. (2018). Eye Gaze and Head Posture Jointly Influence Judgments of Dominance, Physical Strength, and Anger. *Journal of Non-verbal Behavior*, 42(3), 285–309. <https://doi.org/10.1007/s10919-018-0276-5>
- Triana, M. d. C., Jayasinghe, M., & Pieper, J. R. (2015). Perceived workplace racial discrimination and its correlates: A meta-analysis. *Journal of Organizational Behavior*, 36(4), 491–513. <https://doi.org/10.1002/job.1988>
- Trope, Y., & Higgins, E. T. (1993). The What, when, and How of Dispositional Inference: New Answers and New Questions. *Personality and Social Psychology Bulletin*, 19(5), 493–500. <https://doi.org/10.1177/0146167293195002>
- Tsai, J. L. (2017). Ideal affect in daily life: Implications for affective experience, health, and social behavior. *Current Opinion in Psychology*, 17, 118–128. <https://doi.org/10.1016/j.copsyc.2017.07.004>
- Tsai, J. L., Ang, J. Y. Z., Blevins, E., Goernandt, J., Fung, H. H., Jiang, D., Elliott, J., Kölzer, A., Uchida, Y., Lee, Y.-C., Lin, Y., Zhang, X., Govindama, Y., & Haddouk, L. (2016). Leaders' smiles reflect cultural differences in ideal affect. *Emotion*, 16(2), 183–195. <https://doi.org/10.1037/emo0000133>
- Tsai, J. L., Blevins, E., Bencharit, L. Z., Chim, L., Fung, H. H., & Yeung, D. Y. (2019). Cultural variation in social judgments of smiles: The role of ideal affect. *Journal*

- of Personality and Social Psychology*, 116(6), 966–988. <https://doi.org/10.1037/pspp0000192>
- Uchida, Y., & Kitayama, S. (2009). Happiness and Unhappiness in East and West: Themes and Variations. *Emotion (Washington, D.C.)*, 9, 441–56. <https://doi.org/10.1037/a0015634>
- Ueda, Y., Nagoya, K., Yoshikawa, S., & Nomura, M. (2017). Forming Facial Expressions Influences Assessment of Others' Dominance but Not Trustworthiness. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.02097>
- Vacharkulksemsuk, T., Reit, E., Khambatta, P., Eastwick, P. W., Finkel, E. J., & Carney, D. R. (2016). Dominant, open nonverbal displays are attractive at zero-acquaintance. *Proceedings of the National Academy of Sciences*, 113(15), 4009–4014. <https://doi.org/10.1073/pnas.1508932113>
- Valentine, K. A., Li, N. P., Meltzer, A. L., & Tsai, M.-H. (2020). Mate Preferences for Warmth-Trustworthiness Predict Romantic Attraction in the Early Stages of Mate Selection and Satisfaction in Ongoing Relationships. *Personality and Social Psychology Bulletin*, 46(2), 298–311. <https://doi.org/10.1177/0146167219855048>
- Valentine, K. A., Li, N. P., Penke, L., & Perrett, D. I. (2014). Judging a Man by the Width of His Face: The Role of Facial Ratios and Dominance in Mate Choice at Speed-Dating Events. *Psychological Science*, 25(3), 806–811. <https://doi.org/10.1177/0956797613511823>
- Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, 43(2), 161–204. <https://doi.org/10.1080/14640749108400966>
- Valmori, A., Meral, E. O., Hale, M.-L., Rusconi, P., & Brambilla, M. (2023). On the influence of implicit race attitudes on explicit trustworthiness judgments: An investigation of the perceivers and targets' race and gender intersection. *Visual Cognition*, 31(8), 584–598. <https://doi.org/10.1080/13506285.2024.2315800>
- van Rijsbergen, N., Jaworska, K., Rousselet, G. A., & Schyns, P. G. (2014). With Age Comes Representational Wisdom in Social Signals. *Current Biology*, 24(23), 2792–2796. <https://doi.org/10.1016/j.cub.2014.09.075>
- Van Dessel, P., Ye, Y., & De Houwer, J. (2019). Changing Deep-Rooted Implicit Evaluation in the Blink of an Eye: Negative Verbal Information Shifts Automatic Liking of Gandhi. *Social Psychological and Personality Science*, 10(2), 266–273. <https://doi.org/10.1177/1948550617752064>
- Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*, 111(32). <https://doi.org/10.1073/pnas.1409860111>
- Verosky, S. C., Zoner, K. A., Marble, C. W., Sammon, M. M., & Babarinsa, C. O. (2020). Implicit responses to face trustworthiness measured with fast periodic visual stimulation. *Journal of Vision*, 20(7), 29. <https://doi.org/10.1167/jov.20.7.29>

- Vignoles, V. L., Owe, E., Becker, M., Smith, P. B., Easterbrook, M. J., Brown, R., González, R., Didier, N., Carrasco, D., Cadena, M. P., Lay, S., Schwartz, S. J., Des Rosiers, S. E., Villamar, J. A., Gavreliuc, A., Zinkeng, M., Kreuzbauer, R., Baguma, P., Martin, M., ... Bond, M. H. (2016). Beyond the 'east-west' dichotomy: Global variation in cultural models of selfhood. *Journal of Experimental Psychology: General*, 145(8), 966–1000. <https://doi.org/10.1037/xge0000175>
- Walker, M., Jiang, F., Vetter, T., & Sczesny, S. (2011). Universals and Cultural Differences in Forming Personality Trait Judgments From Faces. *Social Psychological and Personality Science*, 2(6), 609–617. <https://doi.org/10.1177/1948550611402519>
- Walker, M., & Wänke, M. (2017). Caring or daring? Exploring the impact of facial masculinity/femininity and gender category information on first impressions (J. Cloutier, Ed.). *PLOS ONE*, 12(10), e0181306. <https://doi.org/10.1371/journal.pone.0181306>
- Wang, A., Quinn, B. P., Gofton, H., & Andrews, T. J. (2024). No evidence for an other-race effect in dominance and trustworthy judgements from faces. *Perception*, 53(9), 632–644. <https://doi.org/10.1177/03010066241258204>
- Wang, H., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2016). The Motivational Salience of Faces Is Related to Both Their Valence and Dominance (K. Lidzba, Ed.). *PLOS ONE*, 11(8), e0161114. <https://doi.org/10.1371/journal.pone.0161114>
- Wang, H., Han, C., Hahn, A. C., Fasolt, V., Morrison, D. K., Holzleitner, I. J., DeBruine, L. M., & Jones, B. C. (2019). A data-driven study of Chinese participants' social judgments of Chinese faces. *PLOS ONE*, 14(1), e0210315. <https://doi.org/10.1371/journal.pone.0210315>
- Wang, M., Zhang, J., Chen, J., & Zhang, L. (2022). An investigation of the influence of skin colour on the perception of femininity, masculinity and likeable. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1044505>
- Wang, Q. (2016). Why Should We All Be Cultural Psychologists? Lessons From the Study of Social Cognition. *Perspectives on Psychological Science*, 11(5), 583–596. <https://doi.org/10.1177/1745691616645552>
- Wang, Q., Chen, G., Wang, Z., Hu, C. S., Hu, X., & Fu, G. (2014). Implicit Racial Attitudes Influence Perceived Emotional Intensity on Other-Race Faces (T. Zalla, Ed.). *PLoS ONE*, 9(8), e105946. <https://doi.org/10.1371/journal.pone.0105946>
- Wang, Y., Dong, Y., Meng, X., Wang, J., & Luo, Y. L. (2024). Impacts of Facial Cues on Trust Judgments in Online Trust Games Among Children and Adolescents. *Proceedings of the 2024 2nd International Conference on Information Education and Artificial Intelligence*, 383–389. <https://doi.org/10.1145/3724504.3724567>
- Watkins, C. D., & Jones, B. C. (2016). Competition-related factors directly influence preferences for facial cues of dominance in allies. *Behavioral Ecology and Sociobiology*, 70(12), 2071–2079. <https://doi.org/10.1007/s00265-016-2211-2>
- Watson, T. L., Otsuka, Y., & Clifford, C. W. G. (2016). Who are you expecting? Biases in face perception reveal prior expectations for sex and age. *Journal of Vision*, 16(3), 5. <https://doi.org/10.1167/16.3.5>

- Wen, F., Zuo, B., Ma, S., Xu, Y., Coley, J. D., & Wang, Y. (2020). Do We See Masculine Faces as Competent and Feminine Faces as Warm? Effects of Sexual Dimorphism on Facial Perception. *Evolutionary Psychology*, 18(4), 1474704920980642. <https://doi.org/10.1177/1474704920980642>
- Wen, Y. F., Wong, H. M., Lin, R., Yin, G., & McGrath, C. (2015). Inter-Ethnic/Racial Facial Variations: A Systematic Review and Bayesian Meta-Analysis of Photogrammetric Studies (P. J. Hills, Ed.). *PLOS ONE*, 10(8), e0134525. <https://doi.org/10.1371/journal.pone.0134525>
- Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric Evidence that Sexual Selection Has Shaped the Hominin Face (T. Fitch, Ed.). *PLoS ONE*, 2(8), e710. <https://doi.org/10.1371/journal.pone.0000710>
- Wildman, A., & Ramsey, R. (2022). Estimating the effects of trait knowledge on social perception. *Quarterly Journal of Experimental Psychology*, 75(5), 969–987. <https://doi.org/10.1177/17470218211047447>
- Wildman, A., & Ramsey, R. (2023). Investigating the automaticity of links between body perception and trait concepts. *Visual Cognition*, 31(5), 341–362. <https://doi.org/10.1080/13506285.2023.2250505>
- Willis, J., & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, 17(7), 592–598. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, 113(1), 59–80. <https://doi.org/10.1037/pspi0000092>
- Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*, 26(8), 1325–1331. <https://doi.org/10.1177/0956797615590992>
- Windmann, S., Steinbrück, L., & Stier, P. (2023). Overgeneralizing emotions: Facial width-to-height revisited. *Emotion*, 23(1), 163–181. <https://doi.org/10.1037/emo0001033>
- Wood, A., Martin, J., & Niedenthal, P. (2017). Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance (D. Reby, Ed.). *PLOS ONE*, 12(8), e0183811. <https://doi.org/10.1371/journal.pone.0183811>
- Workman, C. I., Humphries, S., Hartung, F., Aguirre, G. K., Kable, J. W., & Chatterjee, A. (2021). Morality is in the eye of the beholder: The neurocognitive basis of the “anomalous-is-bad” stereotype. *Annals of the New York Academy of Sciences*, 1494(1), 3–17. <https://doi.org/10.1111/nyas.14575>
- Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, 117(2), 364–385. <https://doi.org/10.1037/pspi0000160>
- Xie, S. Y., Flake, J. K., Stoller, R. M., Freeman, J. B., & Hehman, E. (2021). Facial Impressions Are Predicted by the Structure of Group Stereotypes. *Psychological Science*, 32(12), 1979–1993. <https://doi.org/10.1177/09567976211024259>

- Xu, F., Wu, D., Toriyama, R., Ma, F., Itakura, S., & Lee, K. (2012). Similarities and Differences in Chinese and Caucasian Adults' Use of Facial Cues for Trustworthiness Judgments (E. A. Stamatakis, Ed.). *PLoS ONE*, 7(4), e34859. <https://doi.org/10.1371/journal.pone.0034859>
- Xu, Y., Lee, A., Wu, W.-L., Liu, X., & Birkholz, P. (2013). Human Vocal Attractiveness as Signaled by Body Size Projection (C. Alain, Ed.). *PLoS ONE*, 8(4), e62397. <https://doi.org/10.1371/journal.pone.0062397>
- Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers & Graphics*, 36(3), 152–162. <https://doi.org/10.1016/j.cag.2011.12.002>
- Zabag, R., Azoulay, R., Rinck, M., Becker, E., Levy-Gigi, E., & Gilboa-Schechtman, E. (2023). You never get a chance to undo a negative first impression: Social anxiety is associated with impaired positive updating of social information. *Personality and Individual Differences*, 203, 111993. <https://doi.org/10.1016/j.paid.2022.111993>
- Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individuated impressions of other racial groups. *Journal of Personality and Social Psychology*, 65(1), 85–101. <https://doi.org/10.1037//0022-3514.65.1.85>
- Zebrowitz, L. A. (2004). The Origin of First Impressions. *Journal of Cultural and Evolutionary Psychology*, 2(1), 93–108. <https://doi.org/10.1556/JCEP.2.2004.1-2.6>
- Zebrowitz, L. A. (2017). First Impressions From Faces. *Current Directions in Psychological Science*, 26(3), 237–242. <https://doi.org/10.1177/0963721416683996>
- Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait Impressions as Overgeneralized Responses to Adaptively Significant Facial Qualities: Evidence from Connectionist Modeling. *Personality and Social Psychology Review*, 7(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01
- Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2007). Are Effects of Emotion Expression on Trait Impressions Mediated by Babyfacedness? Evidence From Connectionist Modeling. *Personality and Social Psychology Bulletin*, 33(5), 648–662. <https://doi.org/10.1177/0146167206297399>
- Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2010). Facial Resemblance to Emotions: Group Differences, Impression Effects, and Race Stereotypes. *Journal of personality and social psychology*, 98(2), 175–189. <https://doi.org/10.1037/a0017990>
- Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, 15(6), 603–623. <https://doi.org/10.1007/BF01065855>
- Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology*, 28(6), 1143–1152. <https://doi.org/10.1037/0012-1649.28.6.1143>
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, 2(3), 1497–1517. <https://doi.org/10.1111/j.1751-9004.2008.00109.x>

- Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First Impressions From Faces Among U.S. and Culturally Isolated Tsimane' People in the Bolivian Rainforest. *Journal of Cross-Cultural Psychology*, 43(1), 119–134. <https://doi.org/10.1177/0022022111411386>
- Zhan, J., Garrod, O. G. B., van Rijsbergen, N., & Schyns, P. G. (2019). Modelling face memory reveals task-generalizable representations. *Nature Human Behaviour*, 3(8), 817–826. <https://doi.org/10.1038/s41562-019-0625-3>
- Zhan, J., Ince, R. A., Van Rijsbergen, N., & Schyns, P. G. (2019). Dynamic Construction of Reduced Representations in the Brain for Perceptual Decision Behavior. *Current Biology*, 29(2), 319–326.e4. <https://doi.org/10.1016/j.cub.2018.11.049>
- Zhan, J., Liu, M., Garrod, O. G., Daube, C., Ince, R. A., Jack, R. E., & Schyns, P. G. (2021). Modeling individual preferences reveals that face beauty is not universally perceived across cultures. *Current Biology*, 31(10), 2243–2252.e6. <https://doi.org/10.1016/j.cub.2021.03.013>
- Zhang, D., Lin, H., & Perrett, D. I. (2020). Apparent Emotional Expression Explains the Effects of Head Posture on Perceived Trustworthiness and Dominance, but a Measure of Facial Width Does Not. *Perception*, 49(4), 422–438. <https://doi.org/10.1177/0301006620909286>
- Zhao, F., Mayer, R. E., Adamo-Villani, N., Mousas, C., Choi, M., Lam, L., Mukanova, M., & Hauser, K. (2024). Recognizing and Relating to the Race/Ethnicity and Gender of Animated Pedagogical Agents. *Journal of Educational Computing Research*, 62(3), 675–701. <https://doi.org/10.1177/07356331231213932>