City Research Online

City St George's, University of London

Citation: Williams, R. N. (2026). Investigating Factors that Influence Face Perception. (Unpublished Doctoral thesis, City St George's, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/36918/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

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

Investigating Factors that Influence Face Perception

Rochelle Nicole Williams

A thesis submitted to the

Department of Psychology and Neuroscience

City St George's, University of London

For the degree of

Doctor of Philosophy (PhD)

February 2026

Declaration

I submit this thesis to City St. George's, University of London for the degree of Doctor of Philosophy in Psychology. I confirm that the work contained within this thesis is my own, and none of this work has been submitted for another degree.

Acknowledgements

First, I am immensely grateful to my first supervisor Dr Lúcia Garrido who has come alongside me during my research journey since my Master's degree. I have thoroughly enjoyed working with Lúcia and honestly perceive that my experience has almost been a breeze. Despite the (self-imposed) setbacks, Lúcia has been extremely patient, kind and encouraging throughout, motivating me to get back on track whenever I slowed down or became discouraged. Lúcia has enriched my research journey and has helped me realise I can learn some of the many things I didn't think I could. Lúcia's constant support, even relating to life in general as well as work and career, have been extremely helpful. Lúcia and I have had a great working relationship, and I have enjoyed our weekly meetings. I am honestly blessed to have had Lúcia as my supervisor. I often describe her as 'The Mother of the PhD Department'. I thank my second supervisor Professor Tina Forster for her advice, ideas and suggestions on where I could take my research, and for her encouragement on writing up my thesis.

I also thank the Department of Psychology and Neuroscience for support in matters pertaining to any queries or concerns I have had, and for the provision of resources to allow me to disseminate my findings at national and international conferences. I am grateful to the Experimental Psychological Society for awarding me with Grindley Grants, which provided financial support for these same endeavours on multiple occasions. I am also immensely grateful to the Department of Psychology and Neuroscience for awarding me the Black British Scholarship to fund my PhD. It has been a great source of financial support. I thank Professor Beatriz Calvo-Merino and Dr Andreas Kappes for overseeing my upgrade exam and providing extremely insightful feedback, and for creating a space for us to have a very nice conversation about my projects.

The PhD journey has, ironically, enriched my social life and brought me out of my shell more, and I am grateful for the experiences I have had along the way. I thank all those who I have had very interesting and fun conversations with, including my office mates and friends in rooms D415 and D426. In particular, I am grateful to Datin for coming alongside me throughout all stages of my PhD. Datin's wisdom and sense of stability has really helped steer me as I navigated the technicalities of research, and he has become a great companion and friend. I am also grateful to Gözde and Kristina for their friendship and companionship during my research journey, from heart-to-hearts to spontaneous social outings - I have enjoyed them all.

I would also like to thank my local church community and leaders, whom I have been a part of for all my life. In recent years, I have built friendships and connections with people who have supported me mentally, emotionally and spiritually. I am grateful for the prayers and encouragement as I navigated this stage of my life.

I thank my friends and family for always believing that I could do something like a PhD. Thank you for being a continuous source of support, love and encouragement. I thank my Mum, who has always supported me and would always remind me of the 'why' of my journey, among many other things, and I thank my Dad for his prayers and nuggets of inspiration. I'm also grateful for my siblings as they have encouraged me along the way to keep going.

Finally, I thank my God, who has always sustained me.

Table of Contents

Declaration **2**
Acknowledgements **3**
List of Tables **7**
List of Figures **8**
Peer Reviewed Contribution **10**
Abstract **10**
Chapter 1: General Introduction **13**
1.1 Face Perception 14
1.1.1 Information extraction during face perception 14
1.1.2 Social Traits 16
1.1.3 Kernel of Truth in social trait judgements 21
1.1.4 The Halo Effect 26
1.1.5 Social trait models 27
1.2 Neural correlates of face perception 35
1.2.1 Hierarchical processing of face perception 36
1.2.2 Neural pathways for face perception 39
1.2.3 Identifying the FFA and its computations 40
1.2.4 Identifying the OFA and its computations 43
1.2.5 Identifying the pSTS and its computations 46
1.3 Familiar Face Perception 48
1.3.1 The Mere Exposure Effect (MEE) 49
1.3.2 Models of familiar face perception 53
1.4 Individual differences in face trait perception 57
1.4.1 Individual differences in the three-dimensional model of face-trait perception 58
1.4.2 Relative contributions of individual and shared preference in face trait perception 59
1.4.3 Individual differences in face perception: anxiety, gender and personality 62
1.4.4 Individual differences in the brain: personality similarity predicts brain similarity at the dyadic level 65
1.5 Summary and Thesis Outline 67
Chapter 2: Methodology **71**
2.1 Introduction 72
2.1.1 Representational Similarity Analysis 72
2.1.2 Application of RSA in the present thesis 75

Chapter 3: Investigating how social trait representations of faces change with familiarisation 83
3.1 Introduction 86
3.2 Methods 96
3.3 Results 102
3.4 Discussion 119
3.5 Conclusion 127
Chapter 4: Investigating how brain representations of faces change with familiarisation 130
4.1 Introduction 133
4.2 Methods 142
4.3 Results 151
4.4 Discussion 162
4.5 Conclusion 168
Chapter 5: Dyadic Person Similarity Predicts Similarity in Face Judgements 170
Peer Reviewed Contribution 171
5.1 Introduction 174
5.2 Methods 178
5.3 Results 184
5.4 Discussion 191
References 194
Chapter 6: General discussion 206
6.1 Thesis aims and summary 207
6.1.1 Main findings 207
6.2. Overall conclusions, limitations and future directions 221
6.2.1 Overall conclusions 221
6.2.2. Limitations 223
6.2.3 Future directions 225
6.3 Conclusion 234
References 236
Appendices 260

List of Tables

Table 1. Analysis A: Wilcoxon signed-rank tests: z- and p-values between pre- and post-familiarisation sessions for each trait across face identities.

Table 2. Analysis B: One-samples rank tests for within-participant analyses: means, standard deviations, and z-values comparing RDM correlations for each trait to zero.

Table 3. Analysis C: means, standard deviations, and z-values for one-samples and paired samples rank tests comparing inter-subject RDMs per trait across sessions.

Table 4. Analysis D: Means, standard deviations, and z-values for one-samples and paired-samples rank tests, for each pair of trait RDMs, for each session, and across sessions.

Table 5. Analysis D: Means, standard deviations, one-samples and paired samples Wilcoxon tests for correlations between likeability and trustworthiness RDMs.

Table 6. Analysis A: One-samples rank test: z-scores for RDM correlations within participants for each ROI across sessions, compared to zero.

Table 7. Analysis B: Means and SDs (and medians and ranges) for inter-subject correlations of face-identity RDMs, separately for pre- and post-familiarisation.

Table 8. Analysis C: Means and SDs for correlations of face-identity RDM across ROIs, separately for pre- and post-familiarisation.

Table 9. Regression analyses using Self-Perceived Trait Similarity to predict Face Judgement Similarity.

Table 10. Regression analyses using Personality Similarity to predict Face Judgement Similarity.

List of Figures

Figure 1. Conceptual Diagram indicating the Kernel of Truth of a Person's Personal and Social Dispositions.

Figure 2. Comparison between Haxby et al. (2000) and Duchaine & Yovel (2015) models of face perception in the brain, containing regions and their functions.

Figure 3. Example diagram depicting the comparison between trait RDMs pre- and post-familiarisation.

Figure 4. Example diagram depicting the comparison between ROI RDMs pre- and post-familiarisation.

Figure 5. Example diagram depicting vectors of ratings carried out for Face Judgements, Personality traits and Self-perceived traits.

Figure 6. Sample video stimulus from the experiment. [Redacted]

Figure 7. Analysis A: Mean ratings of all face identities, for all traits, across sessions (pre- and post-familiarisation)

Figure 8. Analysis A: Trajectory of mean trait ratings across 12 episodes collapsed across all 8 face identities, for dominance, likeability and trustworthiness.

Figure 9. Analysis A: Trajectory of mean trait ratings for each of the 8 face identities across 12 episodes, for dominance, likeability and trustworthiness.

Figure 10. Analysis B: Example diagrams depicting the within-participant comparison between trait RDMs pre- and post-familiarisation.

Figure 11. Analysis B: Mean correlation between pre- and post-familiarisation RDMs within participants.

Figure 12. Analysis B: RDMs depicting the Euclidean distance between ratings of each of the face identities, for each trait both pre- and post-familiarisation, for Participant 1.

Figure 13. Analysis C: Example diagram depicting the agreement in trait RDMs between participants separately for pre- and post-familiarisation sessions.

Figure 14. Analysis C: Inter-subject correlations for each trait both pre- and post-familiarisation.

Figure 15. Analysis D: Example diagram depicting the similarity between trait RDMs, separately for pre- and post-familiarisation sessions.

Figure 16. Analysis D: Mean RDMs depicting pairwise correlations between Attractiveness, Dominance, Likeability and Trustworthiness, both pre- and post-familiarisation (across participants).

Figure 17. Analysis D: Additional analyses for likeability and trustworthiness: diagram depicting correlations between averaged pre-familiarisation RDMs with individual post-familiarisation RDMs in a criss-cross manner.

Figure 18. Analysis D: Additional analyses for likeability and trustworthiness: correlations between likeability and trustworthiness RDMs (and correlations within the traits).

Figure 19. Sample video stimulus from the fMRI experiment. [Redacted]

Figure 20. Example diagram depicting the comparison between ROI RDMs pre- and post-familiarisation.

Figure 21. Analysis A: within-subjects: correlations between RDMs pre- and post-familiarisation.

Figure 22. Analysis A: RDMs depicting the distance between faces for the FFA, OFA and pSTS, both pre- and post-familiarisation, for Participant 1.

Figure 23. Example diagram depicting ROI RDM correlations between subjects, for each session, then comparing pre- and post-familiarisation.

Figure 24. Analysis B: Inter-subject Correlations for each ROI both Pre- and Post-Familiarisation.

Figure 25. Example diagram depicting the similarity between ROI RDMs, separately for pre- and post-familiarisation sessions.

Figure 26. Analysis C: Inter-ROI correlations both pre- and post-familiarisation.

Figure 27. Analysis C: Mean RDMs depicting pairwise correlations between the FFA, OFA and pSTS, both Pre- and Post-Familiarisation (across participants).

Figure 28. Individual differences in self- and other-judgements: correlations between average self-perceived trait ratings and averaged face trait judgements.

Figure 29. Individual differences in self- and other-judgements: correlations between average self-perceived trait ratings and averaged face trait judgements.

Figure 30. Scatter plot: Individual differences in self- and other-judgements: correlations between average self-perceived trait ratings and averaged face trait judgements.

Figure 31. Familiar face perception model by Kovács (2020).

Peer Reviewed Contribution

[Publication]

Chapter 5 has been published in the Quarterly Journal of Experimental Psychology.

Williams, R., & Garrido, L. (2026). Dyadic Person Similarity Predicts Similarity in Face Judgements. *Quarterly Journal of Experimental Psychology*.
https://doi.org/10.1177/17470218261420

[Conference presentations]

Williams, R., Korolkova, O., Lavan, L., McGettigan, C., & Garrido, L. (2023, April 19-21). How do Social Trait Judgements Change with Familiarity? [Poster Presentation]. Experimental Psychology Society, Plymouth, UK.

Williams, R., Garrido, L. (2024, May 22-25). Investigating the relationship between dyadic person similarity and face judgement similarity. [Poster Presentation]. European Society for Cognitive and Affective Neuroscience, Ghent, Belgium.

Williams, R., Garrido, L. (2024, July 3-5). Investigating the relationship between dyadic person similarity and face judgement similarity. [Oral Presentation]. Experimental Psychology Society, York, UK.

Williams, R., Garrido, L. (2024, Aug 25-29). Investigating the relationship between dyadic person similarity and face judgement similarity. [Poster Presentation]. European Conference on Visual Perception, Aberdeen, UK.

Williams, R., Garrido, L. (2024, Sep 10-11). Investigating the relationship between dyadic person similarity and face judgement similarity. [Poster Presentation]. British Association for Cognitive Neuroscience, London, UK.

Abstract

The present thesis contains three empirical studies investigating factors that influence face perception in humans. We first investigated how face representations change with familiarisation based on social trait judgements, as research shows that the way we make judgements about familiar and unfamiliar faces is distinct (Mileva et al., 2019; Kim et al., 2022), though we did not yet know about how these representations *change* during the process. Specifically, we measured changes in the way a set of identities was perceived based on each of these traits separately, after participants had become familiarised with them over two to three weeks by watching a television show. Chapter 3 explores how explicit social trait representations of face judgements change with familiarisation both within and between participants, as well as changes in the associations between the traits themselves. We found that representations do indeed change for some of the traits, and that representations for trustworthiness in particular become substantially similar across participants. Results also show that trait representations are significantly inter-associated, particularly post-familiarisation. Chapter 4 investigated changes in the representations of faces in the brain, also within and between participants, as well as changes in the representational similarity between the ROIs themselves. We found that representations appear to change for each of the face-selective ROIs with familiarisation, and that representations in the OFA becomes substantially more idiosyncratic across participants. We also found that the FFA and OFA remain inter-associated with familiarisation. The final empirical chapter turns its focus to individual differences in face perception and assessed the effect of person similarity on similarities in face judgements. Previous work has demonstrated that pairs of people with similar personalities also have similar brain responses to various stimuli (Matz et al., 2022). This effect could also extend to similarities in the way we perceive our social world. Other research has demonstrated that similarities in our personality and social trait dispositions appear to relate to similarities in our judgements about others and appears to generalise across the globe (Oh et al., 2022). However, there appears to be some variability by country. Thus, Chapter 5 investigates whether person similarity predicts similarities in face judgements within a specified world region (i.e., the United Kingdom). We assessed the relationship between dyadic similarities in Big 5 personality traits as well as self-reported social trait dispositions, and face judgement ratings, from all over the U.K. Results showed a significant association between the self-reports and trait judgements of other faces, unaccounted for by similarities in ethnicity, gender, age, and geographical location within the U.K. While similarities in the latter demographic dispositions also relate to similarities in face judgements, this variance is separate from that of personality and social trait dispositions. These findings may have implications for how we may gravitate towards others and form friendships, given that we may perceive our social world in similar ways. Together, findings reveal the effects of

familiarisation and also our own dispositions on the way we make judgements about others' faces, namely that representations of faces appear to change in a relatively short period as we get to know them, and that some of the variance in our perception of others can be attributed to perceptions of ourselves, and this is consistent at the dyadic level. Findings have theoretical implications for models of face perception and challenge theories of face perception in the brain.

Chapter 1: General Introduction

1.1 Face Perception

Face perception is a highly developed visual process in humans (Haxby et al., 2000). We are highly attuned to perceiving faces and are constantly exposed to them in everyday life, whether we are walking in the street, watching TV, or scrolling through social media. From looking at a face, even for the very first time, we very quickly make intuitive judgements about the person based on characteristics such as attractiveness, trustworthiness and dominance (Willis & Todorov, 2006). These characteristics are socially relevant, in that we may rely on them to make decisions about how to interact with someone. For example, we may be more inclined to show respect to those that are older or approach a person for romantic reasons based on their sex or even avoid someone because they look dangerous. While we may use holistic body cues to make judgements about others, it is of particular interest to explore face perception as studies show that much of the time that we spend looking at a face, particularly the eye area, can be attributed to making inferences about person-related characteristics such as age (Nguyen et al., 2009; Kwart et al., 2012), attractiveness (Kwart et al., 2012) and emotion (Buchanan et al., 2007), a feature of which is commonly perceived as 'the windows to the soul'. Findings, however, are mixed regarding the proportion of time spent looking at faces during social interactions versus looking away from the face, where Varela et al. (2023) show that face fixation sits at a rate of around 90% on average, while Kendon (1967) shows that, when accounting for individual differences, the rate ranges between 28% and above 70%. In any case, findings demonstrate that we perceive faces as an important source of social information when interacting with a person.

1.1.1 Information extraction during face perception

Some of the information we initially extract from faces seem to revolve around the visual aspects, such as a person's age, gender, ethnicity, and mood or emotional state. A study by George & Hole (1998) analysed responses from 80 people aged 18-40 who freely described 6 face images, which ranged in age (i.e., two were younger, two were older, and two were age-matched to the participants). Participant descriptions ranged between one word and 38 words (with an average

of 11.24 words) and made judgements about 6 main categories: age, gender, physical appearance (i.e., hair, face features), emotion, attractiveness, and other descriptions relating to a trait). Findings showed that the frequency of age descriptions were around 70% (with 46% of descriptions referring to the face's apparent chronological age and 40% related to age in general; 17% included both descriptions). These age descriptions accounted for the highest percentage of face descriptions in comparison to the other categories. This was followed by descriptions of gender, physical appearance, and some trait about the person, accounting for around 60% of descriptions, then emotion at around 40%, and attractiveness at around 35%. Without any prompting on what participants should judge about the face specifically, participants naturally described the faces based on these characteristics, and to these respective proportions. The findings reveal the extent to which these features appear obvious to the observer, and how noticeable they are quite early on during face perception. Further research has shown that we can make judgements like these, particularly of a person's age, gender and identity, in as quickly as 100ms, especially if we are familiar with the person (Dobs et al., 2019).

Identity recognition is another important aspect of social interaction allowing us to individuate one person from another. In Bruce and Young's (1986) theoretical model of (familiar) face perception, it is proposed that we learn about nuanced person-related information during interactions, including name, age, sex, and other relevant information, and that these 'person identity codes' can be abstracted from the face. Logically, these codes would be less established upon first encounter, but as we become acquainted with someone, these codes become refined and robust in tandem with and through face recognition units (FRUs), the facial elements by which we are able to identify a person. While the nature of identity- and face-recognition codes are thought to be distinct (i.e., the former involves abstract processing while the latter involves visual-based perception), the former is activated by the latter to a larger degree than other aspects such as a person's voice, body form and appearance, as faces generally seem to be the most widely used cue to person identity (Bruce & Young, 1986). Further details of the model are

discussed in the context of familiar face perception and related empirical research in **section 1.3.1** of this chapter.

1.1.2 Social Traits

Research shows that we also make judgements in the social domain. Allport and Odbert (1936) found that about 18,000 words in the English language were attributed to trait impressions of people, indicating the importance of these judgements in our understanding of relationships and everyday interactions. Studies show that some of the traits we often form opinions about revolve around warmth, competence, and two that appear to underlie most of our judgements: trustworthiness (i.e., the *intention* to help or harm) and dominance (i.e., the *ability* to help or harm; Oosterhof & Todorov, 2008). The latter two trait dimensions will be discussed in **Section 1.1.3**. We also seem to be able to form these opinions about faces in as quickly as 100ms (Willis & Todorov, 2006; Olivola & Todorov, 2010), even from emotionally neutral faces, indicating that there are likely some subtle face characteristics that might convey relevant social information. Oosterhof and Todorov (2008) also showed that the most frequently mentioned word during free recall descriptions of faces was attractiveness, which appears to be another important dimension for face evaluations (Sutherland et al., 2013, also discussed in **Section 1.1.3**).

Social consequences of face trait perception: why our judgements matter

Considering how quickly we can make social judgements about others, it is important to consider the social consequences relating to these judgements, as they have the potential to change the course of a person's life when made in certain contexts. While there is wide range of social traits, the present thesis will focus on four in particular, those of which appear to consistently underlie the way we form impressions about others (Oosterhof & Todorov, 2008; Sutherland et al., 2013): attractiveness, which is crucial for mate-finding and bonding, serving as a function for biology

and reproduction (Thornhill & Gangestad, 1993; Foo et al., 2017), trustworthiness, which dictates whether we interact with others in the first instance (Oosterhof & Todorov, 2008), dominance which determines whether a person is a potential threat, or even competent in other contexts (Oosterhof & Todorov, 2008), and likeability, a type of preference (Asch, 1946) which may implicitly serve as a key marker in whether we form and maintain social bonds with others, akin to trustworthiness (Oosterhof & Todorov, 2008).

1.1.2.1 Attractiveness and the Beauty Premium

Findings show that people who are more attractive are more likely to win court cases, and baby-faced individuals are more likely to pay a cheaper fine than those who look more mature (Zebrowitz & McDonald, 1991). Children in schools who are more attractive than their peers are judged as smarter and popular by their teachers (Clifford, 1975) while unattractive children have a higher chance of being punished (Berkowitz & Frodi, 1979) and more attractive teachers in universities received higher ratings for teaching quality (Hamermesh & Parker, 2003). People who are "too" attractive or "too plain" reduce booking intake when working for a company, but moderately attractive people increase bookings, indicating an 'optimal' level of attractiveness that appears to be associated with booking outcome success (Li et al., 2022). So, attractiveness appears to influence many decisions that we make even in situations where it may not appear to be explicitly relevant. What does the effect of face attractiveness look like at the global level? It appears that people mostly associate positive outcomes with good-looking features, where beauty is associated with words including "success", "good" and "trust" to some extent, while ugliness is associated with negative words such as "distrust" and "bad", an effect also known as the Halo Effect (Thorndike, 1920; Kohler & Mill, 2025). This effect was found in 63 out of 68 measured countries (Kohler & Mill, 2025), where two significant deviations from this effect include Hebrew and Amharic. The language of Ethiopia, both languages and cultures of which may be heavily influenced by religious concepts of beauty (Friends of the Academy of the Hebrew Language in Jerusalem, 2025; Bible Hub, 2025). These concepts encompass both the

visual and non-visual components of a person (The Holy Bible: The Amplified Bible, 1 Peter 3:3-4). This may explain the variance found in the descriptions of beauty in these countries. Overall, the results from Kohler and Mill (2025) show that our ideas of attractiveness may collectively be informed by our environments at the societal and cultural levels, as well as religious beliefs. We seem to associate the trait with positive outcomes, which may also lead to higher expectations when it comes to success and perceived propensity for generosity (Solnick & Schweitzer, 1999), and the Halo Effect (Thorndike, 1920).

1.1.2.2 Trustworthiness

Trustworthiness has previously been defined as the willingness to be vulnerable to the actions of another person based on the *assumption* that they intend to carry out a particular behaviour that is important to the trustor, despite their actual ability to do so (Mayer et al., 1995, p. 712). Trust is therefore related to perceived risks in contexts of uncertainty (Sondern & Hertel, 2023), such as when meeting an unfamiliar person or making financial decisions that require a reliance on someone you do not know. Trustworthiness also encompasses perceiving a person's intention to help or harm (Oosterhof & Todorov, 2008).

Additionally, scientific researchers who are perceived to be more trustworthy are perceived to be carrying out "good science" (Gheorghiu et al., 2017) and the faces of people who had received a death sentence for murder were perceived to be more untrustworthy looking, including those who were initially sentenced to death but were eventually exonerated and found to be innocent (Wilson & Rule, 2015). In both cases, images of their faces were subsequently judged by a set of participants who were unaware of the judicial outcomes. In corporate contexts, trust has been associated with transparency by way of accuracy, clarity, and higher disclosure of important information (Tomlinson & Schnackenberg, 2022).

Across the world, trustworthiness may be perceived in vastly different ways, or associated with various concepts, one reason being that it is not strictly related to the visual aspects of a person, leaving room for higher variability. Work involving trustworthiness perception around the world

posits that countries such as Brazil associate trustworthiness with perceived friendliness and an attempt to mitigate any uncertainty between themselves and the other person, and may not be widely related to group-based motivations (Kwantes & Kuo, 2021, pp. 24-28). On the other hand, in various African countries including Nigeria and South Africa, trust is associated with dependency, hope and reliance, on others as well as on God, and countries heavily affected by global events such as the Transatlantic Slave Trade show differences in trust dynamics and, thus, show differences also in the use of language compared to countries that were not directly impacted in the same way (Kwantes & Kuo, 2021, pp. 33-36). Work in Taiwan concludes that trust, while multifaceted, is strictly related to religious ideas such as that of Confucius, including interpersonal mutuality, reciprocation, and the quality of close familial and friend relationships (Kwantes & Kuo, 2021, pp. 62-63). Other countries show differences such as Iran, where trust is associated with keeping secrets, honesty, collectivism, fearing God, and other religious principles of trust as abstracted from the Quran (Kwantes & Kuo, 2021, pp. 72-73), and the United States where trust appears to relate to individualism and how much confidence can be placed on people at the more individual level (Kwantes & Kuo, 2021, p. 107). It can be suggested, then, that on a global scale, common themes across world regions understand trustworthiness and related concepts to be associated with positivity and good intentions, the ability to rely on another person or group, and is something that facilitates bonding between the parties involved. Across cultures, it is also apparent that trustworthiness can be abstracted from the face based on perceived face typicality, so there appears to be elements of the face that might commonly convey good intentions relative to each culture (Sofer et al., 2017).

1.1.2.3 Dominance and Competence

People appear to use slightly distinct cues to determine how dominant a person is when seeing a face for the first time, despite these perceptions being based, at least initially, on physical and visual characteristics (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Features such as competence (Oosterhof & Todorov, 2008), masculinity (Sutherland et al., 2013), face width-to-

height ratio in men (Carré & McCormick, 2008), physical strength (Toscano et al., 2016), and emotionally neutral as well as angrier-looking faces (even if actually neutral; Oosterhof & Todorov, 2008; Sutherland et al., 2013) and even aggression (Francis, 1988; Carré & McCormick, 2008) all contribute to perceived dominance. While research on dominance perception and social outcomes requires further investigation, previous findings indicate that important social outcomes are influenced by perceptions of dominance and competence. For example, CEOs who appear to be competent are more likely to be recruited and receive higher pay (Graham et al., 2017), and political leaders who had won elections were also perceived to be more competent (including dependable and self-disciplined) than runners-up, when their images were presented to individuals who were unfamiliar with the politicians (Todorov et al., 2005; Olivola & Todorov, 2010). Despite not knowing about whether a person is truly capable of stewarding teams or leading a nation, people may assume that the face of someone who merely looks 'dominant' may be a reflection of their capability to do so. It is also apparent that dominance is associated with both positive and negative traits (i.e., competence and aggression), which may explain why abstractions of perceived dominance from various face features depend on context, particularly when considering *physical* dominance (Archer, 1988) and *social* dominance (Chagnon, 1988; Pérusse, 1993), both of which are related but distinct (Panideirada et al., 2022).

1.1.2.4 Likeability

Beyond some of the social judgements previously introduced (Oosterhof & Todorov, 2008; Sutherland et al., 2013), it is important to consider the fact that we tend to make somewhat holistic judgements about others, particularly related to perceived warmth (Asch, 1946; Cuddy et al., 2008). The perception of these traits, and the integration thereof, might develop into a more general, and possibly affective, *preference* for a person (Asch, 1946). For the purposes of the present thesis, I will refer to this type of judgement as liking. It has been concluded that we appear to aggregate various traits and characteristics (and possibly their relative weightings)

when forming an overall impression about someone, which are primarily associated with the sense of 'warmth' about the person (Asch, 1946; Oosterhof & Todorov, 2008). It is axiomatic that liking serves as a positive social disposition allowing for bonding and pro-sociality between people. Likeability as perceived in the face in particular has been highly associated with trustworthiness (Oosterhof & Todorov, 2008) and valence, characterised by how positively (or negatively) evaluated a person is (Rosenberg et al., 1968; Ryali et al., 2020). Rosenberg et al. (1968) assessed associations between overarching dimensions with 64 personality trait terms as sub-dimensions and reported an extremely high correlation between the like-dislike dimension and valence (i.e., good-bad) dimension, while intercorrelations between these dimensions and other overarching ones (e.g., active-passive, hard-soft) were much less related. Therefore, while likely more ambiguous in terms of its explicit relation to both physical features (Ryali et al., 2020) and higher-level affective representations (Asch, 1946), the exploration of face likeability as a trait is key to understanding how more latent factors of social evaluation are associated with person preference.

While comparably under-investigated in face trait perception research, face likeability can be predicted by how typical a face looks, as can the cardinal traits previously mentioned (Ryali et al., 2020). Therefore, the face appears to carry information that makes a person appear likeable, and the current ambiguity in face perception research around the visual representations of the trait warrants further investigation.

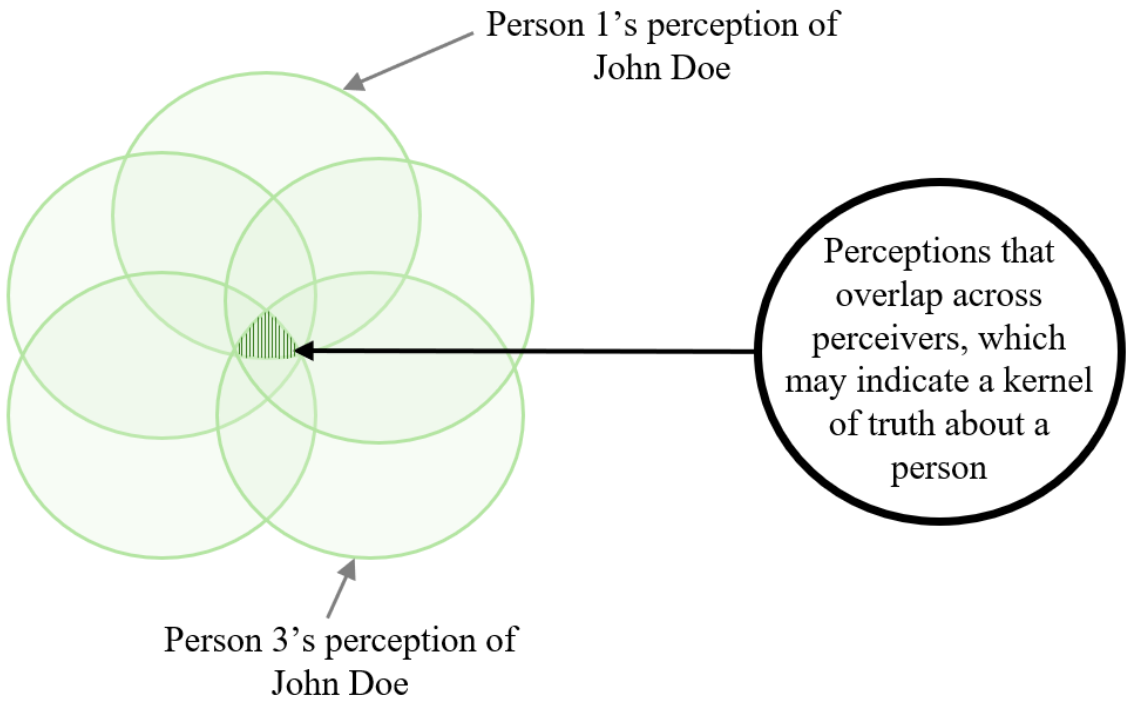
1.1.3 Kernel of Truth in social trait judgements

It is clear from the previous findings that we rely on the information we extract from faces to make decisions about others, sometimes leading to momentous outcomes, but how true are these judgements? The Kernel of Truth hypothesis posits that there is some element of truth among perceivers about an entity, even if the perceptions are largely different overall (Allport, 1954). In other words, the perceptions between perceivers may differ, but overlap to some significant degree, where the commonality may reflect some element of truth about the same person or thing

(see **Figure 1** for illustration). This might especially be the case if we consider that we might have the implicit intention to make judgements about people with an orientation toward objectivity, based on observable characteristics (Asch, 1946).

Figure 1

Conceptual Diagram indicating the Kernel of Truth of a Person's Personal and Social Dispositions



Note. A conceptual diagram of perceptual overlap across perceivers about a single person (i.e., John Doe), where common perceptions may reflect a kernel of truth about said person. Each overlapping green circle represents perception from a single hypothetical perceiver. The shaded area in the centre represents the common characteristics that shape people's perceptions of John Doe, potentially indicative of some true aspect about John Doe.

One possible determinant of the kernel of truth to judgements may involve physical manifestations. The nature of physical characteristics themselves reflect some element of objectivity (i.e., the true characteristics of a physical feature *is* the feature itself in its physical form and with its relative dimensions, etc.).

For perceived attractiveness, it is commonly understood that judgements are based on more observable, and therefore objective, characteristics such as face symmetry and averageness (Rhodes et al., 2001) and larger eyes and lips, characterising baby-facedness (the perception of which varies across the lifespan; Zebrowitz et al., 1993). Attractiveness has been associated with measurable intrinsic factors such as the quality of physical health relating to immune response (i.e., parasite resistance), as mediated by face symmetry and averageness or prototypicality (Thornhill & Gangestad, 1993), as well as fertility in men (Foo et al., 2017). Therefore, visual features may reflect some true element of perceived positive intrinsic qualities, particularly relating to biological functions. The objectivity relating to perceived attractiveness is likely implicitly assessed and demonstrated by the agreement among perceivers when measuring correlations across raters, which appear to be the highest compared to other traits (Oosterhof & Todorov, 2008).

Trustworthiness, on the other hand, is a more visually ambiguous trait and therefore likely rests upon more implicit processing. Trustworthiness appears to manifest, be interpreted, and be measured in various ways. For instance, Jenkins et al. (2013) demonstrate that even different photos of the same person can be rated differently in trustworthiness levels. On the other hand, Rhodes et al. (2013) show that, when assessing more stable representations within a given context, sexual unfaithfulness (which may be related to, but is not itself, trustworthiness), can be consistently predicted in the faces of men, while Bond et al. (1994) show that people who are perceived as more dishonest were more likely to actually take part in research where deception was required on their part. Slepian and Ames (2016) show that this effect is mediated by how the target persons feel the observers *expect* them to behave, which in turn produces the behaviour

related to being perceived as trustworthy. It appears that this effect might relate to a self-fulfilling prophecy, and a few points may be deduced from this. Firstly, it may be the case that the target persons are truly trustworthy through inclining themselves to behave in a way that is socially acceptable and expected of them. On the other hand, these target persons may be merely masking any significantly untrustworthy behaviour and may be displaying trustworthy behaviour based on perceived expectations.

In a similar vein, other findings have found that perceived trustworthiness does not indeed predict true trustworthiness (Rule et al., 2013). Rule et al. (2013) found that people rated both convicted criminals and non-criminals in similar ways, and rated those in the military similarly to those who were not in the military, which may have more than one implication. For example, it could be interpreted that the non-criminals and non-military individuals only had not (yet) manifested 'untrustworthy' behaviour in the way that the criminals or military individuals had done, but that they may still possess the tendency for similar behaviour. It may therefore (and even alternatively) mean that the mechanisms by which untrustworthy behaviour is best explained is by *non-physical* or *non-behavioural* dispositions (i.e., mental state, personality) rather than behavioural outcomes alone, such as conviction rates. This means that we may likely make evaluations about others both based on the way that a face is perceived (i.e., face trait) and also based on predictions about how they might behave (i.e., behavioural trait), leading to trait-behaviour associations (Srull & Wyer, 1989).

When assessing dominance, it appears that judgements of *social* dominance (including propensity for violence; Francis, 1988; Carré & McCormick, 2008, and getting ahead; Mileva et al., 2014) can be correctly identified by *physical* dominance characteristics in the face, as well as the voice (Wolff & Puts, 2010), but this is particularly exclusive to men and the effect is not consistent across many contexts, nor found in women (Mileva et al., 2014). It seems that there is some bearing on the truth of a man's 'dominance' in the social domain when assessing his voice, as he is likely to have optimal levels of testosterone and a higher number of reported sex

partners (which would, thus, be perceived as mating success; Wolff & Puts, 2010). On the other hand, perceptions of other characteristics of social and physical dominance, particularly from the face, may have no significant bearing on the actuality of the judgement, such as prestige (social dominance; Mileva et al., 2014). Face dominance, then, may serve similar functions to that of attractiveness in the context of mating and reproduction, but may not provide a true basis in other social contexts.

Naturally, there would not typically be a measure of ontological 'likeability' due to the dispositional nature of the trait. Preference for a person ultimately relies on the perceiver and it could be argued, therefore, that dispositions of a target face or person would not culminate into an objective 'state' of likeability, as would be the case with attractiveness (i.e., by way of physical appearance, serving as a biological marker for a potentially good mate; Thornhill & Gangestad, 1993; Sutherland et al., 2013; Foo et al., 2017), trustworthiness (i.e., relating to honesty, intention to help (Oosterhof & Todorov, 2008) and sexual faithfulness (Rhodes et al. (2013), and dominance (relating to physicality, masculinity; Oosterhof & Todorov, 2008; Sutherland et al., 2013) - all of which have some objective, measurable basis that relate to important biological functions, independent of preference.

An important takeaway concerning the study of the kernel of truth in face perception is that, irrespective of whether attributions made to a person are true or otherwise, the outcome of these judgements is largely similar. In other words, a person will be treated based on the way that they are perceived, regardless of whether judgements about them are true or not, and thus, there is likely not yet a concrete way to measure a person's true tendency aside from assessing some element of a person's physicality. As previously discussed, research does show that people who merely looked more untrustworthy were given harsher sentences for similar crimes (Wilson & Rule, 2015), and people who looked more competent had been elected to the US House of representatives (Todorov et al., 2005). As a result, we appear to generally make decisions about and interact with people based on how *we perceive* them, often without conscious awareness of

any errors in our perception. As researchers, then, it is all the more important to investigate the underpinnings of social trait perception and their functions during social interactions (for a Meta-Analysis, see Foo et al. 2022).

1.1.4 The Halo Effect

Leading on from the previous point, the variability in social trait perception that exists despite agreement in traits across both individuals and whole nations tells us that our judgements are strongly influenced by aspects separate from the target face. For instance, personal preference appears to play a role in our judgements to a similar extent to that of contributions from the target face (Hehman et al., 2017), but as briefly aforementioned, we also appear to associate certain traits with others (Thorndike, 1920; Mileva et al., 2019; Kohler & Mill, 2025). Mileva et al. (2019) demonstrate that we tend to associate positive traits with each other, namely attractiveness with trustworthiness to a significantly high degree, and extraversion and distinctiveness to a lesser degree. On a global scale, when accounting for 45 countries across 11 regions, people rated faces higher in positive traits such as confidence, emotional stability, trustworthiness, reduced weirdness, caringness, responsibility, intelligence and sociability when they also rated these faces as more highly attractive (Batres & Shiramizu, 2023). The faces that were rated lower in attractiveness were also rated higher in negative traits like aggressiveness, meanness and unhappiness (Batres & Shiramizu, 2023). Despite a lack of personal knowledge and experience, it is a widely universal phenomenon that we still tend to make judgements about whether a person might have good intentions, be more sociable, and even more responsible if they are visually attractive. We seem to use visual attractiveness as a cue for determining non-visual goodness about a person's character, possibly as a mental shortcut to forming impressions about others, as there is no established way in everyday life to ascertain all the information we think we need to know about a person at any given time. In other words, if someone were to approach us unprovoked, we might want to know that they are safe to talk to and mitigate any potential uncertainty. So, we may assume that if they appear to be communicating with us because they

have good intentions towards us (trustworthy), they might be non-threatening (inverse aggressive), or that they are associating with us for pro-social rather than anti-social reasons (extraverted and sociable).

1.1.5 Social trait models

A Two-Dimensional Model of Face Evaluation

So what is the function of face trait perception? What are we ultimately trying to learn about a person we have met for the first time? Oosterhof and Todorov (2008) derived a two-dimensional model of face trait perception that reveals the underlying social information we aim to extract from a face upon first encounter, a model which has become one of the most influential in face trait perception research. The Trustworthiness-Dominance model was derived from descriptions of a set of neutral faces. The opportunity to provide descriptions with no context allowed for capturing as many relevant social inferences that were natural and intuitive to the participants as possible. Participants freely described faces using a range of terms, of which fourteen explained 68% of the proportion (out of a total of 1,100, including attractive, unhappy, sociable, emotionally stable, mean, boring, aggressive, weird, intelligent, confident, caring, egoistic, responsible, and trustworthy). Three-hundred and twenty-seven new participants then rated emotionally neutral faces using the 14 trait dimensions. Twelve of these trait dimensions had high Cronbach's alpha, showing that inter-rater agreement was high ($\alpha > .9$). Two traits were removed from subsequent analyses because agreement was low (egoistic and boring, $\alpha < .17$). The remaining 12 traits were submitted to a principal components analysis (PCA) to determine the traits that contributed most to the face judgements.

Two principal components (PCs) were determined: valence and dominance (the latter dimension was not, however, derived from free-recall responses). Only positively-valenced social traits positively correlated with the first PC and correlated negatively with the second PC. Thus, the first PC was interpreted to be related to positive or negative judgements of a face (i.e., the

goodness or badness of a person). This PC correlated highest with trait trustworthiness, indicating that this trait can approximate and is analogous to valence. The second PC correlated highest with mean dominance ratings, followed by aggression and confidence, and was thus interpreted to be the Dominance PC. In order to measure the variance explained by both trustworthiness and dominance on both PCs respectively, by omission, both traits were removed from analysis and the remaining traits were resubmitted to PCA. Trustworthiness was then correlated with the first PC while dominance was correlated with the second, and results showed significantly high correlations for both traits (trustworthiness x first PC: r = .92; dominance x second PC: r = .87). When correlating trustworthiness with the second PC instead, and dominance with the first, results revealed negative correlations (trustworthiness x second PC: r = -.1; dominance x first PC: r = -.2).

A third trait dimension was determined when combining trustworthiness and dominance at their poles (i.e., low trustworthiness, high dominance). Thus, while not independent from the first two PCs, threat perception could be derived from decreasing perceived trustworthiness whilst increasing dominance. This implies that we perceive faces low in trustworthiness and high in dominance to be threatening.

In another segment of the study, 3-dimensional face models with no emotional expression were constructed and manipulated only using minimal information given by the participants' implicit judgements. In other words, the slight differences in ratings for the emotionally neutral faces were used to morph the 3D models to produce image depictions of the face traits in their extremes. For example, when the neutral computer-generated face model was exaggerated to depict what the participants perceived to be a trustworthy face, the face model showed an expression akin to happiness, while a face modelled to look untrustworthy showed an expression akin to anger. Additionally, a face modelled to look submissive looked more fearful, whereas a face modelled to look dominant also looked angrier. Thus, two new dimensions were derived from the computer-generated faces: frightening (inherently negatively valenced) and likeable

(inherently positively valenced). The 3D model creation successfully emulates and visually depicts the social information that is naturally derived from our implicit judgements.

Overall, Oosterhof and Todorov's (2008) work demonstrates that the various judgements we make about others likely serve the function of learning about a person's propensity for good-will, namely, whether they are trustworthy (described as the intention to help us and "get along"; Markey, 2002; Abele et al., 2020), as well as whether they are dominant (i.e., their propensity for physical harm or "getting ahead"; Markey, 2002; Abele et al., 2020). It could be argued, however, that while the Dominance PC mostly highly correlated with a negative trait (i.e., aggression), perceptions of dominance may also relate to other non-inherently negative traits such as competence, as this trait was also highly correlated with the Dominance PC. Dominant faces were also perceived to be more angry-looking, when generating computerised faces based on participant ratings.

A key feature of the trustworthiness-dominance model is that judgements of faces were obtained from people within the United States, making it susceptible to what is called the Western centrality assumption (Kline et al., 2018), leading to conclusions about judgements being based on a Western perceptual framework rather than a more universal one. In order to assess whether the model was generalisable across the world, Jones et al. (2021) tested an extensive sample of 11,570 participants across 41 countries in 11 world regions, where participants within these world regions were presented with faces that were ethnically representative of their world region and country. Specifically, participants from Africa, Europe, Asia, and South America were each presented with thirty face images of black, white, Asian, and Latin ethnic backgrounds respectively. Procedures across research laboratories in all regions and countries were standardised, and face rating trial blocks were carried out twice for all participants for test-retest reliability, which were averaged during analyses. Using PCA to reduce face judgement dimensionality across world regions, results showed that the two dimensions from Oosterhof and Todorov (2008) also emerged for most other countries, revealing the universality of the social

framework, and possibly the functions, by which we make these social judgements. However, when using other analysis techniques to assess the universality of these dimensions, particularly the emergence of the third threat dimension (derived from low trustworthiness/high dominance), there appeared to be some region-specific models that emerged, highlighting some of the variation and individuality of some world regions in face trait perception, particularly in Asia and Eastern Europe.

Nevertheless, the previous findings provide strong evidence to suggest that human societies across the world might rely on a similar representational framework when making judgements about faces within their own social environments.

A Three-Dimensional Model of Face Evaluation

Further developments were made to the two-dimensional model using a highly varied set of real-life face stimuli (Sutherland et al., 2013) as well as a richer set of demographic-related and social traits, including age, baby-facedness, and sexual dimorphism, to name a few. This breadth allowed for the possible development of other dimensions underlying our judgements based on elements of social perception that may not have been captured in the previous model. From this new set of stimuli, a third statistically orthogonal dimension was derived, labelled 'youthful attractiveness' (Sutherland et al., 2013).

For the first part of the study, however, face trait ratings at their extremes, as well as at intermediate levels of these extremes, were visualised to depict what our face-trait representations might look like on a face. Participants viewed a variety of naturalistic face images of various identities (varying between the sexes, accessories and clothing, hairstyles, age, and skin texture) for each of three experiments. For the first experiment, 12 participants rated 1,000 images based on seven social traits: trustworthiness, dominance, attractiveness, confidence, intelligence, sexual dimorphism (facial femininity/masculinity), and age. There was high inter-rater reliability for the ratings of all 1,000 images (Cronbach's $\alpha > .7$). Ratings were

averaged across participants, and all face images that were rated the lowest and highest for each trait were averaged using image morphing software (i.e., all face images that were rated the lowest in confidence were morphed together to create a 'lowest confidence' face image; all face images rated highest for confidence were morphed together to create an averaged 'highest confidence' face image). The face photographs were morphed by aligning the faces based on 179 fiducial points. This defined the shape and texture of the face that was modelled in subsequent procedures. Eleven face images were produced from each of the two averaged face images from either ends of a dimension (e.g., averages of face images rated lowest and highest in face trustworthiness). These 11 images represented a continuum of the extreme images, where each photo along the continuum was increasingly morphed by 10% in the direction of the other extreme. For example, the first face image in the continuum depicted what the least confident face would look like, then each subsequent face image would 'look' 10% more confident, ending with the face image that depicts the highest confidence level, based on the ratings and stimuli set within the experiment. This was computed for each of the 7 traits, thus producing a total of 77 face images including their continua. Participants then rated each of the 77 faces one at a time from 1 (e.g., very unconfident, unattractive, and so on) to 7 (very confident, attractive, and so on), based on all 7 traits.

The visualisation of the face-trait ratings provided insights into what each of the social judgements might look like based on a varied set of our social trait ratings; both in their extremes as well as the intermediate levels of these extremes. These traits are consistently extracted from a large set of naturalistic face images, which appear to be associated with certain face characteristics. For example, trustworthy faces were depicted as smiling, dominant faces were depicted as masculine, and unintelligent faces were depicted with eyes that appeared to look less focused.

For the next part of the study, principal axis factor analysis was used to reduce the dimensionality of these traits, replicating Oosterhof and Todorov's (2008) two-dimensional model whilst

accounting for potential co-variance, or the intercorrelation of any social traits. Additionally, the aim was to potentially derive as many dimensions as can emerge from this richer set of stimuli. Results replicated that of Oosterhof and Todorov (2008): for the first factor, characteristics akin to that of the original trustworthiness PC explained a significant proportion of the factor loadings and was thus labelled Approachability/Trustworthiness. Specifically, loadings for this factor consisted of perceived approachability, trustworthiness, and degree of smile. For the second factor, proportions of the variance were explained by low age, high attractiveness, health, and baby-facedness and was this labelled youthful-attractiveness. And the third factor was determined by sexual dimorphism towards the masculine end, dominance, increased age, confidence, and intelligence, which contributed substantially to the loadings, and was thus labelled Dominance.

Results indicate that these dimensions are orthogonal to each other, namely that that they are three distinct social traits that underlie our social judgements, but that there might still remain some conceptual relation between a pair or multiple traits or factors. Indeed, as Oosterhof and Todorov (2008) have demonstrated, trustworthiness and dominance each negatively correlate to some degree to the Dominance and Trustworthiness PCs respectively. Other research also indicates intercorrelations between some of the social traits (Mileva et al., 2019). The third factor itself even speaks to the interrelatedness of face traits and characteristics (i.e., age and attractiveness). Therefore, while distinct social domains can be statistically determined, there naturally remains some conceptual relation between them.

In sum, the central implication of these models stipulates that by making various judgements based on a variety of social traits, we are ultimately trying to understand a person's intentions toward us, their capacity for carrying out those intentions, and the assessment of the suitability of a potential mate when we meet them for the first time. In other words, we seem to use relevant information displayed on the face as a function of deciding whether and how to interact with

people based on the three underlying characteristics: approachability or trustworthiness, dominance, and (youthful) attractiveness.

A Four-factor model of face perception: How far does it go?

Multiple studies have extensively relied upon the two- and three-dimensional models of social trait perception in faces (Oosterhof & Todorov, 2008; Sutherland et al., 2013), but much of these studies have utilised a relatively restricted set of face and trait stimuli. Thus, in recent years, Lin, Keles and Adolphs (2021) explored an expansive set of adjectives and nouns used to describe faces, as well as a highly varied set of faces, spanning a breadth of facial features. The authors used a Deep Neural Network (DNN) to derive a set of words that are commonly and frequently used to describe faces from a total of 600 billion semantically associated words. A total of 100 words were retained after removing words that were rarely used, were already similar to other words, or were unclear. DNN was also used to produce visualisations of face stimuli using a model that was trained to classify millions of faces across facial aspects and races. Faces were then classified according to 128 facial features using multidimensional scaling, and 100 faces with the highest variability in face features were retained. Sixty participants rated the 100 faces using the 100 trait words including the cardinal traits (trustworthy, competent and beautiful), and others such as reasonable, feminine and affectionate. The results showed that when reducing the dimensionality of the face trait ratings using 5 distinct methods of exploratory factor analysis (EFA), 4 (i.e., Horn's parallel analysis, Cattell's scree test, optimal co-ordinates and empirical BIC) indicated that the most appropriate number of trait factors was 4, and included those of the two- and three-dimensional models: warmth, competence, femininity and youth (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Findings are in line with previous work (Oosterhof & Todorov, 2008; Sutherland et al., 2013) indicating that even with a broader range of face trait concepts, the approachability and warmth (i.e., trustworthiness) of a person, as well as a person's competence or dominance are fundamental to social perception. It also appears to be the case that a person's age and gender are important features, in line with previous work that lead to

similar conclusions (George & Hole, 1998). This highlights the prominence of making judgements about potential co-operative, pro-social and biological motivations about others, which seem relevant to important factors including mate-selection and reproduction (i.e., youth and sexual dimorphism), friendship formation (i.e., trustworthiness) and physical ability to help and enact positive behaviours (dominance/competence).

Given the extensive set of trait and face stimuli utilised in the previous study (Lin et al., 2021), and the plateau of trait dimensions derived after further reducing dimensionality, it appears that there are, at most, 4 explicit social dimensions that underlie the social judgements that we make about others when we see a face for the first time. While further latent dimensions may remain undiscovered, particularly during familiar face perception and other context-specific environments, it can be concluded that general face perception of unfamiliar faces relies on judgements that, at the minimum, relate to a person's trustworthiness and dominance (Oosterhof & Todorov, 2008), and at most, attractiveness and age in addition (Sutherland et al., 2013).

Even when accounting for the breadth of social contexts, including individual, inter-individual, group, and inter-group dynamics (Abele et al., 2021), and considering the various purposes, contexts, intentions and individual differences between people (i.e., cognitive processing speed), when evaluating social evaluation at these levels using 5 different models of social dynamics (including the Stereotype Content Model (Fiske et al., 2002), the Dimensional Compensation Model (Yzerbyt et al., 2005), Dual Perspective Model (Abele & Wojciszke, 2007), the Behavioral Regulation Model (Leach et al., 2007; Ellemers et al., 2017), and the Agency-Beliefs-Communion Model (Koch et al., 2016), two of the consistently derived fundamental dimensions remain: trustworthiness and dominance. It is apparent, then, that the two are crucial for making various inferences about others in social situations that require co-operation and task-related competence (Abele et al., 2021).

Much of the work that is carried out in face perception research has revolved around unfamiliar faces (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Nguyen et al., 2009; Kwart et al., 2012; George & Hole, 1998; Bond et al., 1994; Slepian and Ames, 2016; Rule et al., 2013; Wilson & Rule, 2015; Todorov et al., 2005; Olivola & Todorov, 2010), and while researchers have extended this work to include that of familiar people (Mileva et al., 2019; Kim et al., 2022), much still remains to be investigated. Some questions remain as to the way that perceptions change as a function of face familiarisation, as we go from being strangers to becoming more acquainted. Familiar face perception will be discussed in **section 1.3**.

1.2 Neural correlates of face perception

This thesis will also investigate face perception on the basis of its underlying neural substrates. In the 1960s, the term "grandmother cell" was coined to describe a hypothetical neuron that responded to someone familiar to us, such as our own grandmother (Gross, 2002). It was posited that grandmother cells would respond to faces that looked similar to our grandmother based on visual characteristics such as white hair, mature skin, and femaleness. This concept led to empirical investigations around the possibility that there are cells in the brain that are tuned for processing various individual face characteristics. Konorski (1967) explored these populations of cells that he called 'gnostic neurons', which we call 'regions' today.

Within the last 30 years, technological advancements have provided neuroscientists with the ability to capture the inner workings of the brain during face perception, using instruments such as functional magnetic resonance imaging (fMRI) to visualise the relevant underlying structures and neural activity in response to faces. Numerous studies have demonstrated the presence of whole regions that selectively respond more to human faces than to other visual objects, treating faces as 'special' (Kanwisher et al., 1997; Gauthier et al., 2000; Rossion et al., 2003; Frith &

Frith, 2003; Yovel & Kanwisher, 2004; Isik et al., 2017), studies of which have been pivotal in our understanding of the domain-specificity of face processing (Yovel & Kanwisher, 2004). Thousands of subsequent studies have replicated these findings in an attempt to identify further characteristics by which these face-selective regions distinguish faces, both from other objects as well as from other faces. Based on many of these studies, models of face perception in the brain have emerged, most since the early 2000s, suggesting the presence of an interconnected system of brain regions implicated in processing faces, each region with distinct but possibly overlapping functions (Haxby et al., 2000; Gobbini & Haxby, 2007; Haxby & Gobbini, 2011; Duchaine & Yovel, 2015; Kovács, 2020; Pitcher & Ungerleider, 2021).

1.2.1 Hierarchical processing of face perception

According to the Bruce and Young's (1986) model of (familiar) face perception, the brain processes the identity of a face in a particular order: first, face recognition units consisting of the visual and structural characteristics of the face are extracted. From these, pre-stored representations associated with the person's identity (i.e., 'codes') are recruited, subsequently triggering the recognition of the person's name. This implies that the brain engages in hierarchical processing during face perception, and that the initial processing of the visual components of a face is crucial for the subsequent recruitment of more abstract person-related representations.

Haxby et al. (2000) first proposed a model which serves as a notable neuroscientific account for Bruce and Young's (1986) model, outlining the hierarchical processing of faces in the brain, consisting of three core posterior face-selective regions residing in the occipito-temporal area. These regions include the lateral fusiform gyrus (FG), containing the fusiform face area (FFA), the inferior occipital gyri (OGs), which houses the occipital face area (OFA), and the face-posterior superior temporal sulcus (pSTS). These regions constitute what was called the core system and was proposed to be primarily implicated in processing the more visual elements of faces. Specifically, the OFA computed lower-level, basic image properties (i.e., face shape and

structure, including the T-formation of eyes, nose and mouth; Gauthier et al., 2000), while the FFA was suggested to be implicated in both higher- and lower-level information processing (i.e., unique face identity, static face features; Kanwisher et al., 1997), and the face-selective STS was suggested to be involved in processing dynamic face features (i.e., eye gaze, lip movement, facial expressions; Haxby et al., 2000).

Haxby and colleagues (2000) also proposed that there was an extended system, which interacts with but is independent from the core system. The extended system would consist of the intraparietal sulcus (IPS), implicated in directed attention, the amygdala and limbic system for emotion-related processes, and the anterior temporal lobe (ATL) involved in the recruitment of a person's name, identity and biographical information. The premise was that regions in the extended system would receive abstract, person-related information that was first extracted from a face via the core face regions, specifically the FFA and pSTS, which themselves may receive relevant visual and structural information from the inferior OG (and, thus, the OFA). This integration was thought to be indicative of a hierarchical system for face perception in the brain.

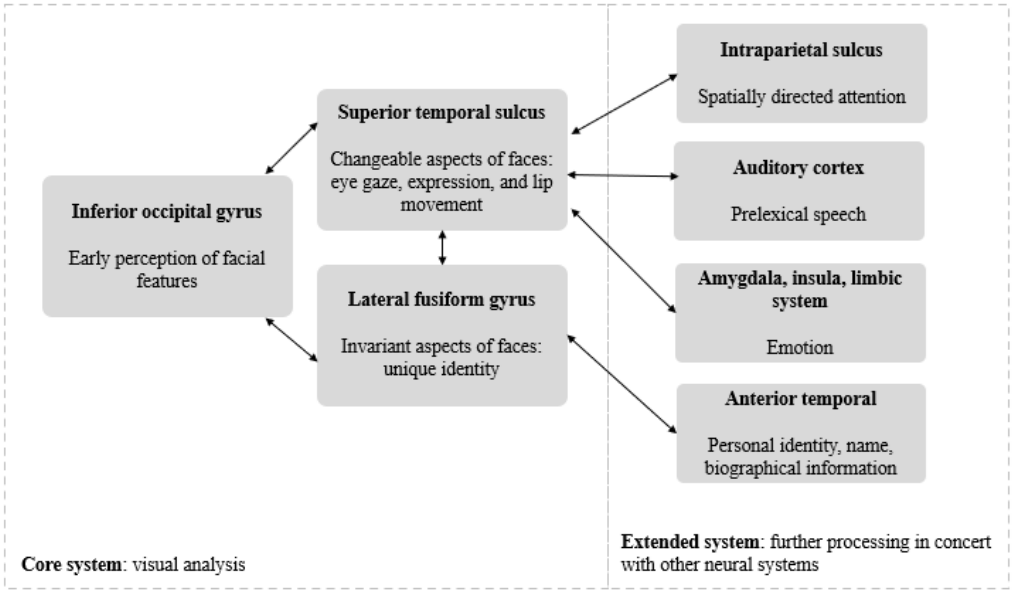
Duchaine & Yovel (2015) propose revisions to Haxby et al.'s (2000) model specifying the core visual ROIs involved in early, intermediate and later face processing. Instead of the broader inferior occipital gyrus, the early visual cortex has now been proposed to be implicated in face processing. Regions that are proposed to play a role in face perception have been specified as the FFA, OFA and pSTS, as well as subregions of the FFA and pSTS, rather than the broader areas of cortex such as the STS. Moreover, there appears to be potential connectivity among the core areas *as well as* direct connections from all core posterior ROIs to more anterior dorsal regions of the brain (i.e., direct connections between the OFA and the face-ATL), rather than a strict succession of connectivity from, for instance, the OFA, to the FFA, then to the ATL) implying a more highly interconnected system for face processing. We also learn that sub-regions have been identified within two of the core regions (i.e., posterior FFA; pFFA, middle

FFA; mFFA, aSTS, and so on), evidencing specified locations of highly attuned face-related activity (see **Figure 2**).

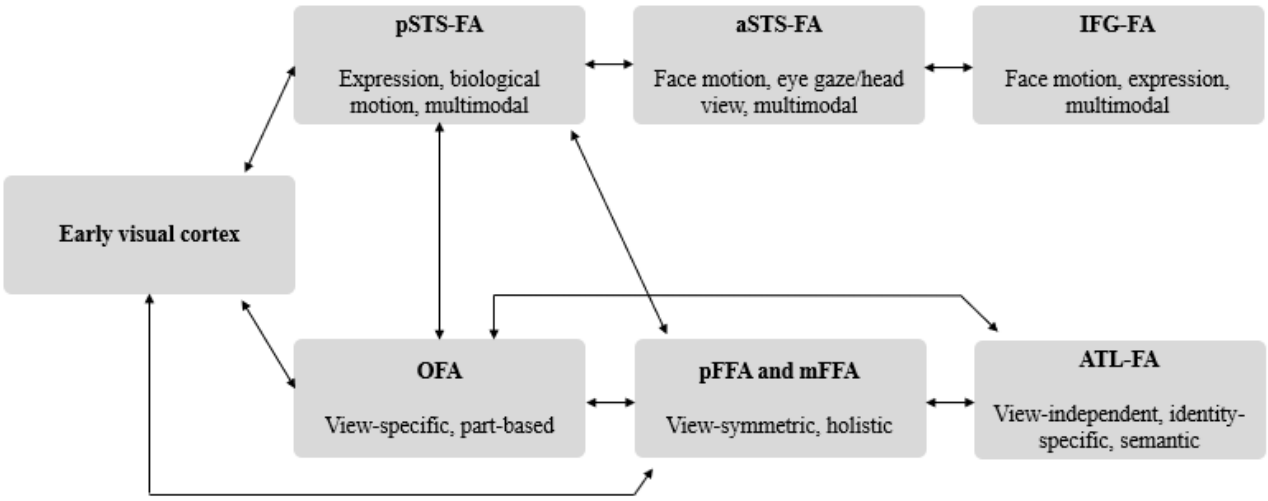
Figure 2

Comparison between Haxby et al. (2000) and Duchaine & Yovel (2015) models of face perception in the brain, containing regions and their functions.

A



B



Note. **A.** Model by Haxby et al. (2000) containing regions within the core system (inferior occipital gyrus akin to the OFA, lateral fusiform gyrus akin to the FFA, and superior temporal sulcus containing the pSTS), and extended system (containing IPS, auditory cortex, amygdala, insula and the limbic system, and ATL). **B.** Revised model by Duchaine and Yovel (2015) specifying each core ROI and sub-region, with an addition of the early visual cortex and more direct connections between ROIs in the core and extended systems.

1.2.2 Neural pathways for face perception

Other models of visual perception in the brain propose that there are multiple pathways through which elements of a stimulus are computed (Duchaine & Yovel, 2015). However, for many years, it was initially proposed that there were two main pathways known as the dorsal and ventral visual pathways (Mishkin and Ungerleider, 1982; Mishkin et al., 1983). The functions of the dorsal pathway were thought to be implicated in visuo-spatial information (i.e., *where* the stimulus is) as well as vision leading to related actions (i.e., throwing, grasping; Milner & Goodale, 2006; Pitcher & Ungerleider, 2021), whereas the ventral pathway likely involves processing object stimuli (i.e., *what* the stimulus is). More recent models suggest that, regarding face stimuli specifically, the ventral visual pathway contains the three core face ROIs (Haxby & Gobbini, 2011) and involves a feedforward flow of face information from early visual areas to the OFA, and subsequently the FFA (then possibly to the pSTS, though some researchers propose the pSTS is instead part of the dorsal pathway; O'Toole et al., 2002; Bernstein & Yovel, 2015). More recent models propose that there are in fact multiple neural pathways implicated in face perception (Duchaine & Yovel, 2015), and there is empirical evidence to suggest a third pathway, where input from early visual areas bypass the OFA and FFA entirely and propagate face-related information directly to the pSTS (Pitcher & Ungerleider, 2021). The authors suggest that this third pathway might be implicated in familiar face processing, though it is not known whether it is exclusive to this type of face processing.

Since the early 2000s, there has been ample research that has successfully identified the relative functions of the core visual areas. These regions appear to be dominant in the right hemisphere (Rossion, 2014) and larger in the right hemisphere (Bukowski et al., 2013). Right-hemisphere dominance for face perception was found to be associated with handedness (Willems et al., 2010) which is consistent with the universal knowledge that around 85-90% of people are right-handed (Masud & Ajmal, 2012). While questions remain as to the representational segregation and specificity of functions in these regions, both individually and as a network, much research over the decades has identified multiple functions that the core face ROIs play in face perception.

1.2.3 Identifying the FFA and its computations

Some early studies in face perception marked a breakthrough in identifying the spatially distinct regions that selectively responded to faces and distinguished between face exemplars based on distinct features. For example, Kanwisher et al. (1997) discovered the FFA (within the fusiform gyrus; FG) using fMRI and functional localisers, which is located in the occipitotemporal area. A series of experiments investigated whether the FFA responded to lower-level face features or to a broader set of stimuli, including animate human body parts and objects.

Firstly, Kanwisher et al. (1997) localised each participant's face-responsive region. Participants passively viewed grey-scale images of the faces of 90 students as well as common objects (e.g., spoon, hammer). The authors examined whether there was any region in the occipitotemporal visual pathway that was 1. active in each participant, and 2. whether the region(s) was consistently activated across participants. Each participant's own 'face area' was located, which was measured within each respective participant in subsequent experiments. Results showed significantly higher activation for faces vs objects in 12 out of 15 participants: 5 showed bilateral activation in this region, and 7 also showed activation in the middle temporal gyrus/rSTS.

Participants were also tested using two-toned face images to rule out the effect of image-based properties (e.g., luminance). When comparing activity patterns to intact versus scrambled two-tone faces, results indicated the FG responded significantly more to intact faces, implying that activity can indeed be attributed to characteristics of the face itself rather than lower-level non-face features. A subsequent experiment in Kanwisher et al. (1997) measured whether the FG merely represented various stimuli of the same category, such as objects with features that are shared with faces (i.e., shape) rather than treating faces as a 'special' type of stimuli, thus assessing the domain-specificity of the FG (Yovel & Kanwisher, 2004). Participants were presented with images of 90 faces and 90 houses, which typically displayed face-like features (i.e., two windows and a door, akin to eyes and a nose/mouth). Results showed that the FG responded significantly more to the faces compared to houses, suggesting that the fusiform gyrus *selectively* responds to faces and not merely different exemplars of the same category. In other words, there is something 'special' about faces, even if face and object processing share a visual pathway during the earlier stages of perception. It could be argued that increasingly complex objects could be represented as a 'special' type of object, including faces. However, activity in this region selectively attends to the unique combination of features that a human face exhibits and, thus, points toward face selectivity. The final experiment attempted to rule out whether activation in the FG may be related to: 1. the human body more holistically (i.e., animate, human body parts), 2. more general attentional processes, and 3. subordinate-level classification. Participants were asked to match images of faces and match images of hands, where only three quarters of the face was shown in the face matching task. Results showed that, again, the localised FG was significantly more activated across participants when viewing the faces compared to the hands, in line with that of the previous experiments pointing towards face-selectivity in the FG, particularly based on the internal features of the face.

There has been much subsequent investigation in support of face-selectivity of the FFA (Puce et al., 1999; Mundel et al. 2003; Földiák et al., 2004; Tsao et al., 2006; Jonas et al., 2018; Jonas &

Rossion, 2021; Tsantani et al., 2021; for a review, see McKone & Kanwisher 2005). However, there has still been a long-standing question among researchers as to whether the FFA is merely involved in categorising stimuli for which we have increased expertise (e.g., chess pieces, Bilalić, 2016; Greebles, Gauthier & Tarr, 2002) or whether the FFA is indeed specifically attuned to faces. In other words, is the FFA exclusively face-selective, or could it be merely face-responsive? Xu et al. (2009) found that the FFA elicits a stronger response when distinguishing between faces with increased visual similarity when compared to various exemplars of the same identity that appear to be more distinct, in line with results from another study (Verosky et al., 2013).

However, face-selectivity in the FFA may relate to abstract representations of a person that are assigned to a face (Bruce & Young, 1986). It could be the case that the FFA is attuned to the identity behind the face and uses face information in order to perceive who the person is (Gilaie-Dotan & Malach, 2007). A study by Gilaie-Dotan and Malach (2007) found that while the FFA showed adaptation (Grill-Spector & Malach, 2001) in the form of reduced activity to the repeated presentation of the same face image, these neurons made a full recovery once presented with a morphed image of two different exemplars of the same face identity. This would be indicative of the direct implication of FFA in face identity processing, where highly specified identity-related information is used to distinguish between faces. These conclusions are also supported by more recent findings (Tsantani et al., 2021), who showed that representations of the FFA was related to categorising faces based on both higher-level perceptions such as social trait information, and lower-level image-based information including perceived face similarity and gender. Though, the FFA appears to be implicated in the former to a larger extent (Tsantani et al., 2021).

It can be argued, then, that the FFA plays a key role in face identity processing but may also moderate tendencies to anthropomorphise non-face stimuli, which is supported by a study showing that significant activation in this region was related to descriptive judgements of car

fronts resembling faces (Kühn et al., 2014). The premise may also be supported by a finding showing that people with face-Paraedolia (the perception of face-like features extracted from inanimate objects) show increased activation in the FFA (Akdeniz et al., 2018). It seems, then, that whether a face is real or merely perceived, the FFA seems to play a significant role in computing some basic, visual face-like properties and assembling parts of face-like images into wholes in order to identify a face (Bruce & Young, 1986). This premise is in line with Liu et al. (2010), who demonstrated that the FFA is sensitive to images of face *parts* as well as faces as a whole configuration.

While there is still an open question as to the exact functions of the FFA, its function(s), seem to involve assembling parts of non-face stimuli that merely resemble faces, which still speaks to its face-*selective* role. What can be concluded, however, is that due to a dual role in processing image-based information as well as higher-level perceptual information (Gilaie-Dotan & Malach, 2007; Tsantani et al., 2021), the FFA could be placed at a more intermediate level of face processing, supporting previous theories and models of face perception (Haxby et al., 2000; Gobbini & Haxby, 2007), and may explain findings that appear to be in support of domain-general processing (Bilalić, 2016; Gauthier & Tarr, 2002).

1.2.4 Identifying the OFA and its computations

A few years after the FFA was discovered, a study carried out by Gautier et al. (2000) sought to further explore face-selective responses in this region. In search for the FFA in a set of participants, the authors happened to locate another face-selective region which was named the occipital face area. During the experiment, 20 participants viewed 10 images of various capitalised letters from the English alphabet, and 40 grey-scale human faces (20 male; 20 female). Images in the face presentation condition consisted only of the internal feature of the faces. Face images were randomly allocated based on gender (i.e., 8 random images of women

constituted one set of images). These images were increasingly morphed together until all images created a final averaged image. The same technique was done for images of letters with various fonts (i.e., 8 images of the letter 'M' in various fonts and so on). These averaged images as well as the originals were presented during trials. During the task, participants were asked to search for stimuli that were presented consecutively in the same location, for all conditions.

Results showed that both the FFA and the newly discovered OFA showed higher activation when presented with the images of faces than of letters, revealing that the FFA and OFA may be implicated in face categorisation. The ability to do this with a variety of face exemplars shows they are highly attuned to features that are common in faces, particularly the internal features (i.e., T-formation of eyes, nose and mouth). Thus, both the FFA and OFA seem to be implicated in exemplar-level face classification, though the FFA responded significantly more to faces than does the OFA. However, the FFA may also play a role in the sub-categorisation of other types of stimuli including non-human animals, as the classification of birds (i.e., a pelican) were associated with activity here. This may explain findings from Kanwisher et al. (1997), namely that sub-classification in the FFA is related to other domains rather than human body parts. As for the OFA, findings from Gautier et al. (2000) show that this region is most likely implicated in processing more image-based features of the face.

Recent research by Tsantani et al. (2021) is in line with this finding, demonstrating that when comparing between representations in the OFA and pixel-level and Gabor jet frequency information, representations were similar, indicating that the OFA appears to distinguish between face identities based on lower-level image properties. These properties likely include facial viewpoint, as research carried out by Kietzmann et al. (2015) demonstrated the causal role of the OFA in perceiving the viewpoint and angle of a face. Participants underwent transcranial magnetic stimulation (TMS) which temporarily stimulates the region so as to temporarily render neurons in this area unavailable to respond to stimuli. Participants received TMS on the right OFA (rOFA) while passively viewing faces from various viewpoints and angles on the

contralateral visual field (i.e., left), to which they reported deficits in both viewpoint and angular perception. When presented with faces of various viewpoints and angles on the ipsilateral side (i.e., right), only deficits in viewpoint perception were reported, indicating the causal role of rOFA in computing facial viewpoint.

An investigation into a patient with bilateral lesions to the occipito-temporal region provides evidence that the processing of image-based properties in the OFA (and FFA) would be isolated to faces, effectively by elimination (Rossion et al., 2003). Patient P. S. had endured face perception deficits after an incident causing lesions to the left middle fusiform gyrus (i.e., left FFA) and right inferior occipital cortex, while leaving the right FFA intact. Patient P. S. was able to recognise a list of objects perfectly and was also able to classify unfamiliar faces but showed a deficit in recognising famous faces she had known prior to the incident, recognising only 14 out of 60 famous faces. This sheds light on the role the OFA plays in face perception and implies the potential occurrence of feedback processes from the FFA to the OFA, for familiar face recognition in particular (Gobbini & Haxby et al., 2007; Dobs et al., 2019).

A comprehensive review of studies by Pitcher et al. (2011) noted high consistency in the face-selectivity of the OFA where a total of 14 studies at the time were able to successfully identify the right OFA and found significant activation to faces when compared to various non-face stimuli including objects, houses, places, scrambled images of faces, letter strings, textures, cars, and Fourier noise images. Taken together, it is clear that the OFA most likely plays a specified role in distinguishing between faces based on physical information.

A recent finding, however, found the OFA to be implicated in identity information linked to semantic information (i.e., name and occupation; Eick et al., 2020). TMS was applied to either the right or left OFA, and participants were asked to recall previously associated names and occupations of the faces. Participants were unable to recall this information when TMS was

applied to the OFA in either of the hemispheres, challenging previous models of face processing (Haxby et al., 2000; Gobbini & Haxby, 2007).

Nevertheless, while further research is required to identify further functions of the OFA, findings consistently show that this region is most likely implicated in image-based computations.

1.2.5 Identifying the pSTS and its computations

More anterior face-selective regions appear to serve increasingly multifunctional purposes. Some of the earliest work on the functions of the face-selective posterior superior temporal sulcus include processing of eye gaze and emotional expressions (Engell and Haxby, 2007). Engell and Haxby (2007) discovered that portions of the pSTS, particularly in the right hemisphere, played an important role in processing emotional expressions (i.e., fear, anger, disgust, and surprise), as well as averted gaze (as opposed to direct gaze, with gazes directed fully to the left, partially to the left, fully to the right, and partially to the right). Thirteen participants viewed either emotionally expressive or emotionally neutral faces, both with direct eye gaze, and also viewed neutral faces with averted eye gaze. Results showed that there was a portion in a lower anterior region of the STS that responded significantly more to emotionally expressive faces than emotionally neutral faces, whereas a more dorsal posterior region of the STS responded more to images of faces with averted eye gaze than images with direct gaze. These findings reveal that this region is implicated in more than one element of face processing. This finding in particular is also in line with that of an experiment involving a person with prosopagnosia, who showed ability to perceive averted gaze, but not emotional expressions, revealing the functional distinctiveness of the two in face perception (Campbell et al., 1990). Later research shows the face-pSTS seems to be involved in processing social interactions more generally, as neural responses were stronger when participants viewed videos of social interactions between people (Landsiedel et al., 2022) and even when viewing video animations

of simulated interactions between visual objects (i.e., shapes), where the human element was removed from the context (Isik et al., 2017).

Researchers have also suggested that this region is involved in the audio-visual integration of face and voice information during social interactions (i.e., when the target person is speaking). While one finding showed no distinction in the way the region computes visual, auditory, or audio-visually integrated information (Hocking & Price, 2008), more recent findings found that part of the pSTS responded exclusively to faces, while the anterior portion (aSTS) responded exclusively to auditory information (Landsiedel & Koldewyn, 2023). Other work investigating the pSTS ascribes more abstract roles such as mentalising and theory-of-mind processing (Frith & Frith, 2003).

Collectively, findings reveal the extent to which the face-selective pSTS plays a multifaceted role in face processing, where it appears to be consistently involved in computing dynamic face characteristics, including eye gaze and emotional expressions (Engell & Haxby, 2007). It seems that face-selective pSTS uses these elements to contextualise social interactions, which may serve the purposes of determining whether a person intends to help or hinder a person (Isik et al., 2017) as well as detecting threat and whether someone seeks our wellbeing (Perrett & Mistlin, 1990). These characteristics are akin to the types of social information we deem fundamentally important when we make judgements about a face (Oosterhof & Todorov, 2008).

While the pSTS appears to serve multiple general functions, there also seems to be sub-regions within the pSTS that each have their own distinct functions. This was directly investigated by Deen et al. (2015) who showed that there is a sub-region that selectively responds to dynamic faces and voices, while other portions respond to combined face and voice information as well as combined face and biological motion. The whole STS also appears to show functional *intra*-connectivity, where sub-regions serve similar functions and are activated as a network during social perception. This has implications for the way face perception researchers investigate face-

related activity in the pSTS, where greater caution is required in order to isolate face-selective sub-areas and, thus, greater caution in assigning meaning to neural activity here.

The key theme underlying the present thesis is to address factors that affect the way people perceive faces. Previous work demonstrates that our everyday experiences involve constantly forming perceptions about people, which may influence the ways in which we interact with them (Berkowitz & Frodi, 1979; Zebrowitz & McDonald, 1991; Graham et al., 2017) whether they are true or not. An important question seems to revolve around how dynamic these perceptions are, and how they might change over time as we get to know people. Much of the previous work revolves around unfamiliar face perception, or face perception in general, without accounting for the social information we might learn about others with repeated exposure and interaction. Familiar face perception research has become increasingly studied within the fields of social psychology and social cognitive neuroscience, and it almost goes without saying that familiar face perception would be distinct from that of a face we have only seen for the first time, hence the two should not be conflated in face perception research (Ramon & Gobbini, 2018). The next section discusses models of familiar face perception from two standpoints, both from the behavioural as well as the neuroscience domain.

1.3 Familiar Face Perception

Face perception is not always a one-time event, but also involves repeated exposure to faces over time, and we may become more acquainted and familiar with a person's character, mindsets, personality, attitude, and other socially-relevant tendencies (Kovács, 2020). The concept of familiarity can be defined or described as a 'feeling' or 'sense' of knowing something or someone; a sort of acquaintance with something or someone (Peskin and Newell, 2004). Familiarity essentially describes one's relationship to a stimulus, and generally ranges from 'no prior knowledge', to 'knowing very well'. It has also been proposed that there is an optimal level of familiarity, where familiarisation, particularly with a person, occurs up to the point where

further interaction with them would no longer significantly increase one's knowledge of them (Popova & Wiese, 2022), be it semantic, emotional, or another type of knowledge. Becoming familiar with someone would mean that one has spent time 'getting to know' them to some significant degree and would involve processes including memory and recognition (Bruce & Young, 1986; Kovács, 2020). Familiarity is not directly or completely quantifiable, and thus, defining the term may presently seem arbitrary.

Though, familiarity can be considered a catalyst for perception, in that it 'speeds up' the processing of a particular stimulus (Manahova et al., 2020), where detection of a familiar face among unfamiliar faces is faster (Visconti di Oleggio Castello & Gobbini, 2015), and recognition of characteristics extracted from a familiar face is faster than that of an unfamiliar face (i.e., age, gender; Dobs et al., 2019). Thus, we may find that we are very quick at recognising a piece of music we know, our favourite object, or our best friend.

1.3.1 The Mere Exposure Effect (MEE)

Findings show that shortly after meeting someone for the first time, people tend to develop a liking or a preference for them over others (Zajonc, 1968). This is known as the Mere Exposure Effect (MEE), where slight familiarity by way of repeated rapid exposure to a stimulus influences the perception of them (Bornstein & Craver-Lemley, 2022). This effect has also been alluded to as a "glow of warmth", a "feeling of intimacy", or possibly even a "release from tension" associated with an unfamiliar or potentially threatening stimulus (Titchener, 1909, p. 408), which is in line with research showing that the MEE was stronger in those who reported higher levels of social anxiety (Siegel et al., 2019). Bornstein and Craver-Lemley (2022) reviewed over 400 studies that investigated the effects of merely being repeatedly exposed to various types of stimuli, including standard and nonsense words, drawings and images, pictures, and music. The empirical findings demonstrated that brief exposure to these stimuli increased, including a liking for people. However, it was apparent that liking for a stimulus decreased after a specified number of exposures as a result of boredom or habituation, and was stronger when a

stimulus was implicitly perceived (Van Dessel et al., 2017). This depletion of interest, along with the distinction in implicit and explicit processing, indicates an important difference between mere exposure and a longer-term familiarity with a stimulus. Mere exposure appears to be related to lower-level perception and may occur without conscious awareness (Van Dessel et al., 2017), whereas familiarity through more thorough acquaintance with a stimulus may be related to higher-level explicit processing, where this type of information is gained with repeated exposure to a stimulus over time, thus potentially changing or solidifying the more concrete representations of that stimulus and not necessarily leading to habituation or boredom.

1.3.1.1 The Mere Exposure Effect on Social Trait Judgements

At what point, however, does mere exposure transpose into actual familiarisation with a person? And does our perception of people differ between the two? To date, there appears to be no study conducted on the mere exposure effect on social trait perception, except that of Peskin and Newell (2004). The authors sought to investigate the effects of two distinct types of face familiarisation on the perception of attractiveness, one type including mere exposure (described as 'episodic familiarity') and the other type based on the affective experience of familiarity (described as 'perceived familiarity'). This is an important distinction given the distinct processes involved in merely seeing a face on multiple occasions versus repeated exposure coupled with other psychological and (inter)personal experiences with the person behind the face. While the authors, as do others, treat both conditions as a type of 'familiarity', I discuss the present study to highlight the important distinctions between the longer-term, higher-level familiarisation, and exposure at the lower-level perceptual plane.

In the first experiment, which explored the 'placebo' type of face familiarity (i.e., perceived familiarity), 31 participants rated 84 monochromatic images of smiling female faces on attractiveness (from 1 – very unattractive, to 7 – very attractive). Another set of participants (*N*

= 30) rated the same faces on how familiar each face was (on a scale of 1 (very unfamiliar) to 7 (very familiar). For the perceived familiarity condition, participants were told prior to rating the faces that the images were of people that were currently attending the college, however, the images were actually of people who had attended the college 10 years prior. Yet another group of participants rated the faces on perceived distinctiveness. The scores for attractiveness and familiarity were correlated, showing a significant positive relationship, indicating that faces that were perceived as more familiar were also perceived as more attractive, indicating that familiarisation itself may lead to a halo effect (Thorndike, 1920).

In the second experiment, the authors explored the effect of *episodic* familiarisation by way of repeated face exposure and its effect on perceived face attractiveness. Yet affective or implicit familiarity was also explored in the present experiment by manipulating face distinctiveness. Face distinctiveness was used as it was proposed that any newly seen face deviating from a perceived 'averageness' of all faces in our memory would be perceived as less familiar, while a newly seen face that is more similar to this 'average' face representation would be perceived as more familiar. Distinctiveness ratings from the first experiment were used to sort 16 faces into the 'distinctiveness' group, and another 16 into the 'typicality' group. A further 8 faces for each of the conditions were sorted. The 8 faces from each condition (distinctive vs typical) were each to be presented to the participants once during the subsequent judgement phase, while 16 from each condition were to be presented 6 times each prior to and during the judgement phase, thus producing 192 trials for the exposure condition and 48 trials for the judgement phase. In short, faces were either presented once or six times. Thirty-two participants rated these faces on perceived attractiveness, and results showed that faces presented 6 times were also rated as more attractive than faces presented only once. Additionally, typical faces were rated as more attractive than distinctive faces. Taken together, the results indicate that faces that are repeated visually and more typical looking are perceived as more attractive. Taken together, the two experiments suggest that more familiar faces are perceived to be more attractive. These findings

are in line with the mere exposure effect (Zajonc, 1968) which proposes that we tend to prefer things we are more familiar with and also supports the notion that more typical-looking faces are perceived as more familiar due to 'commonality' between face exemplars and the average face represented in memory.

So merely feeling familiar with and undergoing brief repetition of a face image appears to increase judgements of attractiveness, though the effects of repeated exposure were tested over a relatively short time period (i.e., within a session). In effect, a test on mere exposure was carried out, leading to effects that may not remain stable over a relatively extended period of time due to habituation (for a meta-analysis, see Bornstein, 1989). In fact, after 20 exposures to a stimulus, the effect of increased positive affect may taper off (Bornstein, 1989). Research also shows that the effects of 'mere exposure' appear to last over a week or 2-week period, where being exposed to a face rapidly more than 4 times (and up to 16 times), and with a delay as short as 5 minutes and as long as 2 weeks increased preference ratings for that face (Stang, 1975).

In real life, familiarisation increases with exposure and, thus, we may habituate to the extent where attractiveness perception stabilises. Results from Peskin and Newell (2004) make apparent the distinctions between the mere exposure effect and the more solidifying psychological experience of familiarisation, particularly on perceptions of attractiveness. Importantly, due to the brief nature of mere exposure, the effect of repeated exposure to faces may be mitigated by individual differences (Ladd & Gabrieli, 2015). For example, Ladd and Gabrieli (2015) show that both people with state and trait anxiety do not significantly develop a preference for faces they have been briefly exposed to.

Therefore, in this thesis, I refer to the terms 'familiarity' and 'familiarisation' as a process that both elicits and consolidates pre-stored, higher-level representations of a stimulus with repeated exposure and learning over a significantly longer period of time, ideally a minimum of two weeks.

1.3.2 Models of familiar face perception

Behavioural model of familiar face perception

Similarly to the previously discussed models, familiar face models outline the ways in which a face is computed once we have prior knowledge of the person. Bruce and Young's (1986) familiar face model focuses on the relative functions of certain perceptual and cognitive processes involved in perceiving a familiar face. The model stipulates that there are various codes or units of information, including pictorial (visual-based), structural (aspects of a face that allow for distinction between faces), face recognition units (FRUs; visually-derived semantic information), person identity nodes (identity-specific semantic information), and name generation. Pictorial codes are typically the basis upon which our perceptions of an unfamiliar face lie, but this is not the case for familiar face perception, as invariant characteristics beyond the (static) image of a person causes us to recognise a person, such as the information perceived from the structural codes, which are abstract and stable in nature. This indicates that with repeated exposure to various exemplars of faces (which may vary in their pictorial codes based on face viewpoint and angle orientation, for instance), stable identity representations develop. These would develop by learning about the person based on descriptions or perceptions about the face features, individually and as a whole. These descriptions produce face recognition units consisting of semantic information extracted from the visual elements of the face, which become stable over time and are constructed into identity-related codes (pieces of information related to more stable identity representations of that person). These identity codes, and then subsequently the person's name, are then elicited by the internal features of the face during social interactions. Thus, a familiar face is recognised when the pre-stored identity information (i.e., name, semantic information) about a person matches the presently encoded representation or exemplar of their face.

Neuroscientific models of familiar face perception

Familiar face perception was proposed to be distinct from the perception of unfamiliar faces (Ramon & Gobbini, 2018), implying the presence of distinct underlying cognitive processes. Gobbini and Haxby (2007) proposed an update to their previous model (Haxby et al., 2000) accounting for the neural processing of familiar faces, again serving as a neuroscientific explanation for Bruce and Young's (1986) model. A slight revision was made to account for the decoding of semantic information in familiar faces, which may occur in the ATL in the extended system, receiving input from the FFA in the core system, reflecting bottom-up processes during familiar face perception (Visconti di Oleggio Castello et al., 2017, 2021). Naturally, all core regions are suggested to be implicated in familiar face processing (Gobbini & Haxby 2007; Kovács, 2020), but there appears to be distinct identity-related activity between unfamiliar and familiar faces in these core regions (Kovács, 2020), where significant activity related to the identity of a face was found for unfamiliar but not familiar faces. This implies that people tend to rely on more visual-based information in order to identify an unfamiliar face, while relying on more abstract non-visual representations to distinguish between identities for familiar faces, supporting previous models (Bruce & Young, 1986). Duchaine and Yovel (2015) pose that there are feedback (i.e., top-down) processes at play during familiar face perception, where areas implicated in higher-level perceptual information about a face relay information to more ventral areas implicated in lower-level processing. This is supported by findings of enhanced neural activity in the occipito-temporal area with face familiarisation (Dobs et al., 2019; Ambrus et al., 2021). Behavioural work by Schwartz & Yovel (2016) suggests that we begin to rely on higher-conceptual information in familiar faces: faces that were repeatedly presented alongside names were recognised significantly more than faces without names, irrespective of face presentations from various viewpoints, indicating that conceptual information like names consolidate stronger invariant face identity representations. This raises questions as to how the brain reorganises the way it represents face information with increased person familiarisation. Do these ROIs rely on different pieces of information in order to represent a familiarised face?

Computational model of familiar face perception

Kramer et al. (2018) used computational modelling to simulate the bottom-up automatic processes (using PCA) as well as top-down feedback processes (using linear discriminant analysis; LDA) that likely occur during the ongoing process of face familiarisation. Bottom-up processes in face perception are typically implicated in computing visual face characteristics whereas top-down, higher-level processes during familiar face perception would typically be implicated in back-propagating pre-stored abstract information that would have been encoded from a face previously (i.e., semantic and non-visual identity-related information, Bruce & Young, 1986; Schwartz & Yovel, 2016; Dobs et al., 2019). Kramer et al. (2018) proposed that we tend to mainly *extract* information during unfamiliar face perception, often relying on visual face features (and thus engage more in bottom-up processing). It seems plausible that this would typically occur at first encounter given that the social function of the interaction would be to focus on learning about the person, and therefore extract as much information as we can about them. It also makes sense based on the finding that we quickly make judgements about the age and gender of a person (George & Hole, 1998) which are typically apparent in the physical features of face. However, during familiar face perception, incoming perceptual information likely elicits pre-stored person-related semantic knowledge and, thus, we rely on visual face information significantly less (Visconti di Oleggio Castello et al., 2017, 2021; Schwartz & Yovel, 2016; Dobs et al., 2019).

In Kramer et al. (2018), various exemplar images of face identities were submitted to PCA space alone as well as combined PCA+LDA space. The two simulated the encoding of bottom-up (i.e., unfamiliar face processing), and the interaction between bottom-up and top-down processes (i.e., familiar face processing) in humans respectively. The degree to which the model spaces became 'familiarised' with the face identities was determined by the number of face images submitted to the combined model space, reflecting the human experience of face familiarisation through repeated visual exposure. The combined space was repeatedly exposed to over 4,000 face images

to simulate face familiarisation (some images of which were famous, about half were female, and most were Caucasian, although the set of images contained a mix of ethnicities). Some face identity images were projected more than others, emulating various degrees of familiarity. To test the 'recognition' abilities of the model, a leave-one-out approach was used, where a test image was projected into the combined space after all other images.

Results in Kramer et al. (2018) showed that the model correctly clustered the test images with the projected face images based on the internal features of the face (eyes, nose and mouth), even when images were degraded by 25% and 50%. The combined PCA+LDA model also correctly match identities to some extent, though there was overlap: the model had mistaken some of the different identities as being the same people, though performance was almost as good as human face matching. The model failed to classify faces when PCA was tested independently from LDA, demonstrating the inability to 'recognise' a face without relevant identity information. In other words, LDA (or top-down processes in humans) appears to be necessary for the successful identification of familiar faces, paired with PCA (or bottom-up processes). This model serves as a useful demonstration of face familiarisation in terms of the way we integrate new information about faces with pre-stored information about a person's identity, and in cumulative fashion. It is important to note, however, that this integration of information is restricted to the visual modality and only speaks of the stable visual representations associated with a person's identity gained during face familiarisation. This is due to the fact that only images were tested in the models and, by nature, would not emulate the processing of abstract information. Nevertheless, the model demonstrates how the perceptual process of integrating information about a person becomes stable and thus increases recognition accuracy of the face identity.

The model in Kramer et al. (2018) also demonstrates that familiarity is not binary but exists along a continuum, in line with subsequent research (Karimi-Rouzbahani et al., 2021). This does not negate the premise that there is an optimal level of face familiarity, as other studies have proposed that 14 months is necessary to become *optimally* familiar with someone based on the

recency, frequency and closeness of contact (Popova & Wiese, 2022). In any case, Kramer et al. (2018) have successfully demonstrated the necessity for stable higher- or intermediate-level information in familiar face recognition and have also demonstrated that familiarity is best treated as a process rather than a categorical entity.

It is clear from the models previously presented that successful recognition of a face requires pre-stored visual and abstract person-related knowledge (Visconti di Oleggio Castello et al., 2017, 2021; Schwartz & Yovel, 2016; Kramer et al., 2018; Dobs et al., 2019), and that this information needs to be successfully matched with the presently encoded face information (Bruce & Young, 1986; Kramer et al., 2018). It is also apparent that these cognitive processes might be implicated in similar brain regions to that of processing unfamiliar faces, though in distinct ways (Kovács, 2020). Thus, the present thesis will examine the ways in which the representational structure of faces in face-selective ROIs change with familiarisation.

1.4 Individual differences in face trait perception

The current thesis aims to address another facet that may moderate the way we represent faces, namely, the individual differences in our face trait judgements. Researchers are increasingly interested in what drives our individual differences in the way we perceive faces (Finn et al., 2020; Sutherland et al., 2020a, 2020b), and, thus, are turning toward a focus on observer differences and the effects of the observer's traits and dispositions. Firstly, I will address the individuality of face perception based on the universal trustworthiness-dominance-attractiveness dimensions. This will shed light on the stability of face trait perception for each person when viewing the same faces. Then, I will outline empirical studies demonstrating the relationship between various dispositions based on social traits, personality, and also (sub)clinical conditions. By definition, individual differences encompass a vast range of traits, characteristics and tendencies and, thus, research in this domain is quite vast and includes a sparse range of categories. Some of the ones that will be discussed in the next sub-sections include personality (Mattarozzi et al., 2015; Oh et al., 2022; Matz et al., 2022), social trait dispositions (Hönekopp,

2006; Hehman et al., 2017; Oh et al., 2022), trait anxiety (Willis et al., 2013), gender (Mattarozzi et al., 2015), and private taste (Hönekopp, 2006; Hehman et al., 2017), all of which have been found to influence social perception.

1.4.1 Individual differences in the three-dimensional model of face-trait perception

In reference to the three-dimensional model from Sutherland et al. (2013), findings show that three fundamental social dimensions appear to underlie the judgements we make about others, and that this is universal to some extent (Jones et al., 2021). But what does this look like at the individual level? Namely, what is the extent of this consistency at the individual level?

In Sutherland et al. (2020a), 24 participants each rated 200 varied face images from 1-7 based on 12 traits (used in Oosterhof & Todorov, 2008). From this data, the authors explored two questions: 1. would the participants produce models that replicate the holistic three-dimensional model in Sutherland et al. (2013) individually? (that is, trustworthiness, youthful attractiveness and dominance as underlying dimensions of their face judgements). And 2. do the participants use the same facial cues to make these judgements?

In answer to the first question, results showed that PCA models from 18 out of 24 participants (75%) indeed produced the same underlying social dimensions as the universal model, showing that at the individual level, participants' perception of faces ultimately rely on trustworthiness, attractiveness and dominance, but also competence as an additional factor (i.e., the third factor was split: around half of the participants produced competence dimensions while the other half produced dominance/competence dimensions). It appears that while trustworthiness and attractiveness dimensions are more stable, the competence/dominance dimension may contain higher variability. It could be the case that individual differences have a stronger influence on the latter dimension in particular, or that participants are relying on a wider variety of face cues to make these judgements.

Regarding the second question, each participant's factor score was correlated with factor scores for the same faces from the universal model (in Sutherland et al. 2013). Results showed that the participants tend to rely on the same face cues to judge trustworthiness or warmth (with an average correlation of .79), and youthful/attractiveness (average correlation of .74), though to a lesser extent. However, participants seem to rely on largely different cues to determine dominance or competence, with an average correlation of .40. This could explain some of the variability found in the competence/dominant PCA loadings in the previous analysis. It could be that competence and dominance are two sides of the same coin, and, thus, people may interpret the concept of 'dominance' differently. Some of the variance could also relate to judgements being made in relation to the observers themselves: it could be that participants are making judgements about whether the faces appear to display 'physical strength' or the 'potential to harm' *them* specifically, which may vary from observer to observer due to a sort of polarity. In other words, one observer may feel that a person is able to physically harm them because they appear to display more physical strength in comparison, while another may view the same face as less dominant while perceiving themself to be physically stronger in comparison. A later study by Sutherland et al. (2020b) found that the variance found in individual observers appear to be substantially related to our personal experiences, while a small proportion can be accounted for by genetics (based on experiments in the study using twin samples), and an even smaller proportion may be attributed to a shared environment. These results open up further questions about other factors that may shape our perceptions of others. For example, to what extent do our own dispositions contribute to our judgements?

1.4.2 Relative contributions of individual and shared preference in face trait *perception*

Research shows that our individual *preferences* play a role in our judgements to an equal extent as target face characteristics. These personal preferences could be responsible for some of the variability that is present in face perception. With a particular focus on face attractiveness,

Hönekopp (2006) investigated the extent to which private taste and shared taste contribute to face attractiveness judgements. The study consisted of three parts: in the first part, 31 participants made judgements about the attractiveness of 77 emotionally neutral faces. Participants rated face attractiveness on a scale of 1 (not attractive) to 7 (very attractive) and had rated these faces twice, with a week's interval between rating sessions to ascertain the stability of each participant's rating for each face. Shared taste was derived from the average correlation between participants in attractiveness ratings. On the other hand, private taste was derived from the difference between all participants' average ratings for each face and an individual's own average rating for each face (between the two sessions). The results showed that both shared taste and private taste contributed about equally to face attractiveness judgements. Part one of this study was replicated in the second part to explore the effect of ethnicity. In other words, the authors explored whether participants within the same ethnic group as the face they rated would rate those faces as more attractive than faces of another ethnic group, and results revealed this to be the case. Whether this effect is due to in-group favouritism or exposure was not explored, however. In their second experiment, results also showed that the effect of private taste was slightly stronger than shared taste on face attractiveness. This implies that private taste relies significantly on the sameness of ethnicity between the target face and the observer. This is likely due to familiarity with the ethnicity, and the fact that the target face is similar to the self (Zebrowitz et al., 2007), or even preferences shaped by the culture and environment that is common between the two. Lastly, the authors further explored the extent of private and shared taste for gender. Shared taste and private taste were analysed for women and men separately. Results showed that while, again, private taste and shared taste contribute roughly equally to face attractiveness perception, private taste contributed significantly more to face attractiveness in men than in women. This may relate to mate-seeking behaviours such as the careful visual assessment of physical attractiveness, which are particularly pronounced in men (Marcus & Miller, 2003), which, however, may only be relevant for potentially short-term relationships (Little et al., 2014) and may therefore manifest

differently when rating familiar female faces, or long-term female friends or partners. Taken together, the findings from all three experiments strongly suggest that unfamiliar face perception, particularly when judging face attractiveness, relies on both commonly extracted face information as well as private taste, and to similar proportions.

Hehman et al. (2017) extended the investigation to assess whether the effect of private and shared taste was found using multiple face traits and not only face attractiveness. The authors gathered 698,829 ratings of faces from 6,593 participants in 39 different countries, which would inform us of the generalisability of this effect. The authors explored a wide range of traits including creativity, intelligence, competence, attractiveness, physical power, health, warmth, friendliness, likeability, social power, aggressiveness, femininity, happiness, race-typicality, assertiveness, youthfulness, wisdom, anger, meanness, smartness, trustworthiness, dominance, and care. The quantity of ratings varied between 857 (for ratings on anger) to 79,379 (for ratings on physical strength). The results revealed that while there is much variance yet unaccounted for, each set of trait ratings were somewhat related to both shared and private taste. Namely, raters relied on their own preconceived representations to some degree as well as commonly perceived characteristics in order to make various judgements about others. However, another notable finding showed that each trait varied in terms of the quantity to which respective ratings were related to the perceiver or target. In other words, some traits were more visual based (e.g., happiness, assertiveness, anger) while others were more perceiver based (e.g., creativity, intelligence, aggressiveness). Some traits were roughly equally self- and target-based (e.g., meanness, trustworthiness, dominance). It is also worth noting that while there was a large discrepancy between the number of ratings for some traits (such as anger and physical strength), due to the relatively large quantity of ratings within each trait, there is likely some stability that remains within the ratings. This is especially considering that the ratings were collected across 39 countries, which would increase 1. the generalisability of findings and 2. the reliability of results, as the authors note. Findings from both Hönekopp (2006) and Hehman et al. (2017)

demonstrate that face perception, more generally, may rely on both the contributions of representations that come from individual perceivers as well as characteristics extracted from faces. The characteristics commonly extracted from the face are related to characteristics that are most obviously decipherable, even if judgements are made based on more non-visual attributes such as character. In sum, the findings tell us that, across people, our judgements are not only influenced by the faces we look at, but also by our own preferences.

1.4.3 Individual differences in face perception: anxiety, gender and personality

Individual differences in anxiety and trustworthiness perception

Some individual differences in face perception may relate to traits related to mental health and (sub)clinical conditions (Willis et al., 2013). For example, Willis et al. (2013) investigated whether trait anxiety in the general population related to face trait perception. The authors assessed whether people who reported higher levels of trait anxiety using the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1971) would perceive emotionally neutral faces as more untrustworthy compared to those who reported lower levels of trait anxiety. For the approachability judgement task, participants were asked to imagine being on a crowded street in a hurry to meet a friend. An emotionally-neutral face was presented to them, and they were asked to rate whether they would approach this person for directions. For trustworthiness judgements, participants were asked to imagine they were on holiday. Again, they were presented with emotionally-neutral faces and were asked to imagine whether they would trust the person to hold their camera for them. Results showed that those who reported higher trait (but not state) anxiety were more likely to rate faces as less trustworthy, but this was not the case for approachability. Findings suggest that anxious people tend to generally be more untrusting of a face, even if there doesn't appear to be significant signs of untrustworthiness in the face. This may reflect a type of projection of previous expectations or unpredictability to the face identity and demonstrates the

importance of accounting for individual differences in face perception based on cognitive and clinical traits in the general population (Willis et al., 2013).

Individual differences in personality and gender on trustworthiness perception

However, this finding was not replicated by Mattarozzi et al. (2015), though they suggest that it is possibly due to methodological differences, indicating the necessity for providing context in order to experimentally induce feelings of anxiety, which the authors did not do. Nevertheless, Mattarozzi et al. (2015) investigated other subgroups and the contribution of these individual differences to face judgements, particularly on face trustworthiness. Specifically, the authors investigated whether gender and personality traits, as well as anxiety and aggression, affect face perception. Four-hundred and ten participants rated 60 emotionally neutral faces that were either high, neutral or low in levels of trustworthiness. Participants then completed a battery of questionnaires, including the 60-item Big Five Questionnaire (BFQ), a questionnaire on both state and trait anxiety (STAI-Y), a questionnaire on aggressiveness, measuring verbal aggression, anger and hostility (AQ), and the rational-experiential inventory (REI) which assesses cognitive styles including 'need for cognition' and 'faith in intuition'. The results showed that gender had an effect on face trustworthiness judgements, but only for women: women who scored low on aggression (verbal, anger and hostility) significantly rated trustworthy faces as more trustworthy. Thus, there was an effect of gender on face trustworthiness perception, but only if the face was perceived to be trustworthy. There was no effect of gender, however, on trustworthiness judgements for those who specifically rated themselves higher in agreeableness. Both the women and men who rated themselves higher on agreeableness also rated the trustworthy faces significantly higher than those who rated themselves as less agreeable. In fact, agreeableness was the only personality trait to significantly relate to face trustworthiness judgements. In other words, there was no significant relationship between personality ratings from the other subdomains and face trustworthiness ratings, and none between anxiety and trustworthiness judgements either. This effect was absent both overall

and when measuring the effect by gender. The results do indicate, however, that both gender and personality independently have an effect on trustworthiness perception to some degree. Though, as the authors note, it would be more informative to explore this effect with dominance as the other fundamental social dimension (Oosterhof & Todorov, 2008) and explore with a wider range of other social traits, which would provide a more comprehensive understanding of the effects of gender and personality on face judgements.

Personality and self-perceived trait structure predicts the structure of face-trait perception

A broader set of personality traits, and social traits perceived within the self, do in fact seem to relate to our judgements about others. Oh et al. (2022) investigated the ways in which social traits were associated with each other, both in the perception of the self and, consequently, in the perception of others. In other words, it is possible that a person who associates their own intelligence, for instance, with their own attractiveness, and also projects this perpetual association onto others. The authors investigated this relationship across multiple world regions to understand the universality of this potential association. The work discussed in previous sections shows that certain social traits are important when it comes to making judgements about others (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Sutherland et al., 2020). Are these traits also important for self-perception? Could it be that we use our self-perception (based on these traits) as a framework to make the same judgements about others, or even the reverse?

The authors tested this at a global scale (42 world regions) in which they analysed 307,313 self-report responses (Johnson, 2014) and 13,671 social trait judgements of faces (Jones et al., 2021) from two online databases. Three hundred items from the Revised NEO-PI (NEO-PI-R) Big 5 personality scale including characteristics such as modesty, cheerfulness, self-consciousness, imagination, and self-efficacy which corresponded to either openness, neuroticism, agreeableness, extraversion, or conscientiousness dimensions (Costa & McCrae, 1999, 2000). These characteristics were then categorised into 13 social traits (attractiveness, dominance, trustworthiness, aggressiveness, emotional stability, unhappiness, intelligence, meanness,

weirdness, confidence, responsibility, care, and sociability; Oosterhof & Todorov, 2008) by 25 different coders (i.e., cheerfulness as a Big 5 characteristic would be assigned to the 'sociability' social trait category).

Self-report ratings and trait judgements were averaged across participants for each world region and were computed into separate dissimilarity matrices (RDMs), which depict the pairwise similarity between conditions or responses to stimuli, thus reflecting the whole structure of responses to a variable (e.g., personality). Each entry in the matrices marked the average correlation between pairs of trait ratings (i.e., attractiveness and intelligence, and so on). Thus, there was a self-perceived trait RDM and a face-trait judgement RDM, both constructed per world region. RSA was used to compare the two RDMs, and results showed a significant positive correlation between the two, indicating that co-occurrences in self-perceived personality and social trait dispositions are associated with the judgements we make about others, and this is the case on a global scale. Specifically, if people associate aggressiveness and intelligence within themselves and happen to perceive another person as more aggressive, they are also more likely to perceive the person as more intelligent. This could indicate that we consistently use the social self as a framework to make judgements about others in our environment, or that we may even learn to associate certain traits within ourselves as a consequence of what we perceive to co-occur in others. While this pattern occurs across the world, specified regions show slight variability in terms of the specific traits that are associated with each other in perception, showing that the unique personality structure of people within a given region is reflected in their judgements of others within that same region. In short, Oh et al. (2022) demonstrate that some of our individual differences, namely variability in personality and social dispositions, may affect our perceptions of others.

1.4.4 Individual differences in the brain: personality similarity predicts brain similarity at the dyadic level

Personality similarity at the dyadic level appears to be reflected in similar brain activity, which in turn might reflect perceptual similarity. Matz et al. (2022) investigated whether brain similarity (in terms of brain activity patterns and magnitude) was reflected in perceptual similarity, and whether this in turn related to an alignment in personality. In the first study, 66 participants completed a short 10-item personality questionnaire based on subdomains of each of the Big 5 traits (Ten-Item-Personality-Inventory; TIPI; Gosling et al., 2003) and provided information about their age, gender, ethnicity, nationality, and handedness. Participants then underwent functional magnetic resonance imaging (fMRI) while they watched a series of short videos including comedy, documentaries, and drama to evoke differences in opinions and perspectives. Similarities in both the magnitude and pattern of activation for pairs (dyads) of participants were computed for 200 cortical and 14 subcortical brain regions. Results showed that personality similarity, in one subdomain of each Big 5 domain, was significantly associated with neural synchrony, and was significant to a larger extent than the socio-demographic characteristics. This was the case for both the magnitude and pattern of activity in multiple brain regions, particularly in the frontal lobe, the default mode network, and the limbic system. However, personality similarity did not predict neural synchrony in visual areas, typically involving more ventral regions in primary and perceptual areas. Firstly, this indicates that the neural activity alignment in these regions may be associated with an alignment in perspectives and opinions, and not merely visual stimuli. Secondly, and importantly, it indicates that personality similarity is associated with brain synchrony, both in terms of the magnitude and pattern of neural responses. By extension, those with more similar personalities are more likely to share perspectives in a range of ways including sense of humour, issues they deem important and what they find interesting.

To more thoroughly explore the relationship between personality similarity and brain similarity, the authors carried out another experiment using electroencephalography (EEG), and an

extended (60-item) version of the personality questionnaire, as well as an addition of political ideology questions. In this experiment, 225 participants viewed 104 images of various concepts including people, nature, and art, while brain activity was being recorded using EEG. Eye-gaze patterns were also recorded using an eye-tracking device to analyse similarities in eye motion. Average whole brain activity and eye gaze patterns were correlated between dyads of participants. Results were replicated in the second study: neural synchrony in brain activity from EEG recordings significantly correlated with 15 personality facets (except for the aesthetics sub trait), and to a larger extent than gender, ethnicity and political ideology. These results suggest that an alignment in neural activity can indeed be predicted by an alignment in personality (except for the aesthetic sub-trait, which may be highly idiosyncratic). Could personality in turn predict perceptual similarity at the same level? Oh et al. (2022) demonstrate that this is the case at the global level, but what is the extent of this relationship at the dyadic level?

In the previous section, I have addressed some of the ways in which our individual differences manifest in our perception of faces. We see that we each have somewhat robust face-trait representations that ultimately rely on three important social traits (trustworthiness, dominance and attractiveness, Sutherland et al., 2020), and that our personal preferences play as much of a role in these judgements and others as do the face characteristic themselves (Hönekopp, 2006; Hehman et al., 2017). Additionally, when we look at the face-trait structure across the world, we see that this is related to the self-perceived personality and social trait structure, and that the patterns in our face judgements are associated with our patterns of perception in others (Oh et al., 2022) even at the more idiosyncratic level of subgroups (Mattarozzi et al., 2015) and possibly the dyadic level, as seen in the brain (Matz et al., 2020).

1.5 Summary and Thesis Outline

The overarching research question of this thesis is to investigate factors that moderate face perception. The present thesis will explore two major domains of research within the field that

seek to understand the ways in which we perceive others, which will further enriching our understanding of the way we interact with all kinds of people in our environment, including those close to us, new acquaintances and strangers. This includes 1. the effects of familiarity on face perception, both from a behavioural as well as a neuroscientific lens. It will then turn its focus to 2. individual differences in face perception to explore the extent to which those who are more similar in their personality and self-perceived social trait dispositions also perceive face traits in a more similar manner. The thesis employs representational similarity analysis (RSA) in all empirical chapters, including both standard RSA and inter-subject RSA (IS-RSA). Chapter 2 outlines the conception of RSA (Kriegeskorte et al., 2008a) and how it has been used to compare representations across domains in ways that have not been previously possible, such as between humans and monkeys (Kriegeskorte et al., 2008b). Chapter 2 also outlines the use of IS-RSA, particularly for assessing the relationship between representations at the dyadic level (Matz et al., 2020), maximising idiosyncrasy whilst assessing consistency simultaneously, also called 'idiosynchrony' (Finn et al., 2020).

Chapter 3 examines how social trait representations change as we become familiar with people. While prior research assesses representational differences between social trait judgements in unfamiliar and familiar faces, they do not inform about the *process* of change in representations over time. Moreover, previous studies have often compared representations between unfamiliar faces and faces that are already familiar to us (i.e., famous faces). These types of familiar faces have shown to be representationally distinct from those we actually *become* familiar with (Kovács, 2020). Thus, Chapter 3 explores the representational structure of social trait judgements for unfamiliar faces and then how these change as individuals become familiar with the same individuals through para-social (i.e., through media, TV) or contextual learning (Kovács, 2020; Ambrus et al., 2021). The fundamental social dimensions as well as trait likeability were used in the present experiment: trustworthiness, dominance, attractiveness and likeability (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Participants watched a television series (The Apprentice

UK, series 11) to learn about 8 different face identities. Face-trait representations for these identities will be measured for both unfamiliar faces and for these same faces after the participants had become familiar with them. Changes in social trait perceptions will be assessed as structures (i.e., how faces are perceived in relation to others), rather than with averages for each face identity.

Chapter 4 investigates how familiarity affects neural representations of faces in three core regions: the FFA, OFA, and pSTS. Previous work demonstrates that the FFA, OFA and face-selective pSTS are implicated in both unfamiliar and familiar face processing (Gobbini & Haxby, 2007; Duchaine & Yovel, 2015; Pitcher & Ungerleider, 2021; Kovács, 2020). However, there seems to be differential activity in identity representations in these regions for familiar faces, in that we may no longer rely on visual features in order to perceive a familiar face (Kovács, 2020). This implies that the representational structure within these regions may change with familiarisation, which Chapter 4 sets out to explore. Brain activity patterns were recorded from the same participants that took part in the behavioural experiment in Chapter 3. Thus, participants' brains were scanned using fMRI whilst they viewed faces of each identity from The Apprentice UK series, both before and after watching the show. Similar questions to that of Chapter 3 are of interest, thus, the re-organisation of face representations in each ROI will be assessed.

Chapter 5 turns its focus to investigating how individual differences in personality relates to face trait judgements at the more idiosyncratic level. This particular study leads on closely from Oh et al. (2022) who demonstrated that the structure of personality and social trait dispositions are associated with face trait judgements at the universal level. Importantly, the authors demonstrated that there is some variability by region, and it was reasoned that investigating this effect within a specified world region (i.e., the United Kingdom) and without the effects of averaging (hence, dyadic person similarity) would shed light on the robustness of this effect at a more personal level. Research in face perception is moving towards more idiosyncratic scales,

particularly based on dyadic situations (Finn et al., 2020), therefore we aimed to measure differences between individuals rather than differences at the regional level. Matz et al. (2022) demonstrate that pairs of people with similar brains also have more similar personalities. Could this be reflected in perceptual similarity of faces at the same level? For the experiment, participants were recruited from around the U.K. and were asked to rate faces based on 6 social traits (i.e., trustworthiness, dominance, attractiveness, aggressiveness, intelligence and sociability; Jones et al., 2021). These 6 traits appear to be of particular importance for face trait perception within the U.K., and thus, the intention was to employ a representative sample using social traits that were representative to the world region. Participants then rated themselves on the Big 5 personality traits (McCrae & Costa , 1985) as well as 13 social traits (including those used to rate faces and additional traits: meanness, weirdness, confidence, emotional stability, unhappiness, responsibility and caring). The 'idiosynchrony' (Finn et al., 2020) or similarities in personality and social trait dispositions between pairs of individuals, will be correlated with their respective social trait judgements of faces. Implications may relate to potential motivations for gravitating towards others and building friendships with those we are similar to, and, thus, whom we may share perceptions of our social world (Matz et al., 2022).

Chapter 6 will discuss the findings of all empirical chapters, as well as the efficacy and limitations of the methodologies used. It will conclude on the theoretical and practical implications of all main findings and will present suggestions for future research directions.

Chapter 2: Methodology

2.1 Introduction

In the present chapter, I will introduce a data analysis framework known as Representational Similarity Analysis (RSA; Kriegeskorte et al., 2008a), an increasingly utilised tool for analysing large and complex datasets particularly in the fields of neuroscience and computational modelling. I will outline the principles of RSA and how it is employed in a standard way, then I will detail the use of RSA for the datasets present in this thesis. I first use RSA to assess the representational structure of face identities from both behavioural and fMRI brain data, and how these representations may change with face familiarisation (i.e., from pre- to post-familiarisation). The behavioural data consist of ratings of social judgements from faces and the fMRI data consists of brain activity in response to faces in core face-selective regions. Finally, I turn my focus to individual differences in face trait perception and use inter-subject RSA (IS-RSA; Finn et al., 2020). I use IS-RSA to investigate how inter-subject similarities in the structure of face-trait are related to inter-subject similarities in personality.

2.1.1 Representational Similarity Analysis

RSA was founded on principles related to the fields of mathematics and psychology and aims to bridge the subdomains of neuroscience, behavioural research and computational modelling (Kriegeskorte et al., 2008a). A significant difficulty in researching the functions of the brain revolves around accurately assigning specified patterns of neural activity to the events that elicit this activity. Studies have previously measured brain activity patterns against patterns of stimuli presentations often by using correlation analyses at the first level (i.e., correlating face image presentation timings to corresponding brain patterns), whereas RSA advances *upon* first-level

analysis and aims to capture similarities between these variables at the second-level, in the form of what is called second-level isomorphisms (Kriegeskorte et al., 2008a). For example, similarities between conditions within a domain (i.e., behavioural ratings of a set of faces) are first computed, creating a computational structure, which is subsequently compared to that of another domain (i.e., similarities between conditions based on brain activity patterns elicited by a set of faces). This approach enables researchers to map similarities between different domains at the level of second-order representations. A representation is interpreted as an event-related output in some form (i.e., brain activity pattern elicited by a stimulus; Kriegeskorte et al., 2008a).

In other words, RSA entails that a structure or geometry of a representation of a set of conditions is obtained and statistically 'preserved', which is then compared with another representation of the same conditions from another domain (type or modality of data) to assess the similarity in the organisation of conditions. In RSA, the structure of these variables is computed as a representational dissimilarity matrix (RDM). In an RDM, each pairwise entry represents a dissimilarity measure between a pair of conditions (e.g., voxel activations elicited by two stimuli). The dissimilarity between the pairs may be measured using any dissimilarity measure, including reverse correlation (i.e., 1 minus the correlation between the two data values; Aguirre, 2007; Haxby et al., 2001; Kiani et al., 2007; Kriegeskorte et al., 2008a), Euclidean Distance (Edelman et al., 1998), Mahalanobis distance (Kriegeskorte et al., 2006), and others (Walther et al., 2016). Pairwise dissimilarities (or distances) are computed for all possible pairs. These pairwise distances are then constructed into a matrix (the RDM) and, thus, show the dissimilarity between each pair of conditions. The distances themselves are considered first-level analysis; the comparisons between the whole structures or RDMs are considered second-level analysis, or second-order isomorphisms. This is because the structures are directly analogous to the stimuli they represent in terms of the way they are organised. Thus, the RDMs are interpreted as representations or statistical depictions of the way relevant stimuli are mentally organised. For

instance, one may generate an RDM for activity patterns to faces in the fusiform face area (FFA), and another RDM relating to trustworthiness ratings of these same faces. The two RDMs can then be compared to assess whether the structure or organisation of the representations of these faces are similar. Similarities between the two RDMs would indicate that they are both isomorphic or analogous to each other to some significant extent, and it may in turn be interpreted that trustworthiness representations are related to face representations in the FFA. Essentially, RSA is a useful way of comparing the structure or geometry of representations of conditions using measurements that cannot be compared directly.

Representational similarity analysis can be used to compare the representational organisation or clustering of stimuli between various *groups* and at multiple levels of granularity. For instance, the organisation of neurons within Region A can be compared to that of Region B, and the same can be compared between individual subjects at the more general level, and even between species at the even broader level (Kriegestorte et al., 2008a, 2008b). One of the first demonstrations of RSA compared the representational categorisation of visual objects between monkeys and humans (Kriegeskorte et al., 2008b). Brain activity in the inferior temporal cortex (ITC) were obtained from monkeys using neuronal recordings, and from humans using fMRI. Both monkeys and humans viewed images of 92 animate and inanimate objects, including household items, as well as human faces and hands. The ITC is typically implicated in visual object processing in both monkeys and humans (Van Essen et al., 2001; Tsao et al., 2003; Tootell et al., 2003; Denys et al., 2004; Orban et al., 2004; Van Essen & Dierker, 2007), therefore, the aim of the experiment in Kriegeskorte et al. (2008b) was to investigate whether this region was implicated in categorising objects in a *similar* way between the species. While previous studies shed light on the general functions of this region, RSA allowed for the direct comparison of representations across species (Kriegeskorte et al., 2008b). Electrophysiological recordings of activity of neurons (in monkeys) cannot be directly compared with fMRI responses from voxels (in

humans). But RSA provided an innovative way of directly comparing brain activity from both species by comparing the representations. For monkeys and humans separately, dissimilarities of brain activity patterns from the ITC in response to each stimulus were computed into RDMs, where each entry depicted the pairwise distance between mean amplitude of activity relating to the stimuli (using reverse correlation, i.e., 1 minus the correlation between each pair of stimulus). All pairwise distances for all stimuli were computed into separate RDMs for monkeys and humans. Higher values for each entry indicated higher dissimilarity and lower values indicated more similarity. The resulting RDMs were then compared across species, and results showed that the two RDMs were highly correlated. These results showed that both monkeys and humans categorise objects, both animate and inanimate, in a similar way. RSA is revolutionary in terms of the level at which we can now understand the functions of and the information that is implicated in particular brain regions.

RSA has become increasingly utilised over the past decade to identify corresponding representational properties across disciplines, including social neuroscience and cognition, computer modelling and neuroscience (Ambrus et al., 2021; Tsantani et al., 2019, 2021; Oh et al., 2022). In the present thesis, I employ RSA at multiple levels, including the (neural) regional, inter-regional, and also the participant level. I also use RSA to compare representations within the behavioural domain.

2.1.2 Application of RSA in the present thesis

2.1.2.1 Face-trait RSA

In Chapter 3, I investigate the representational structure of social trait judgements made about faces, and whether there are changes in the structure of the representations of these judgements

as we become familiar with and learn more about people[1]. These representations are based on four social traits: trustworthiness, dominance, attractiveness and likeability. This effect of familiarisation could be evidenced by a change in the clustering of faces, based on the social traits. For example, if Face Identity A and Face Identity B are rated similarly on trait trustworthiness pre-familiarisation, and this significantly changes post-familiarisation, then this might suggest that the mechanisms by which one made the call on trustworthiness might change. RDMs for each social trait were created using people's face judgements on four social traits including the fundamental dimensions (i.e., trustworthiness and dominance; Oosterhof & Todorov, 2008), likeability (Oosterhof & Todorov, 2008) and attractiveness (Sutherland et al., 2013).

RDMs were created for each trait, where each entry in the RDM corresponded to the dissimilarity (using Euclidean distance) in the rating for each pair of faces based on that trait (see **Figure 3**). Distances in the range of values for pairwise vectors of face ratings were obtained, and each value in the resulting vector (representing the range between a pair of face identities) was squared. The vector of squares was then summed, and the square root of the resulting sum was obtained. This was carried out for all possible face pair combinations and, thus, all resulting values represent the dissimilarity between each condition or face pair (i.e., 8 x 8 dissimilarity matrices).

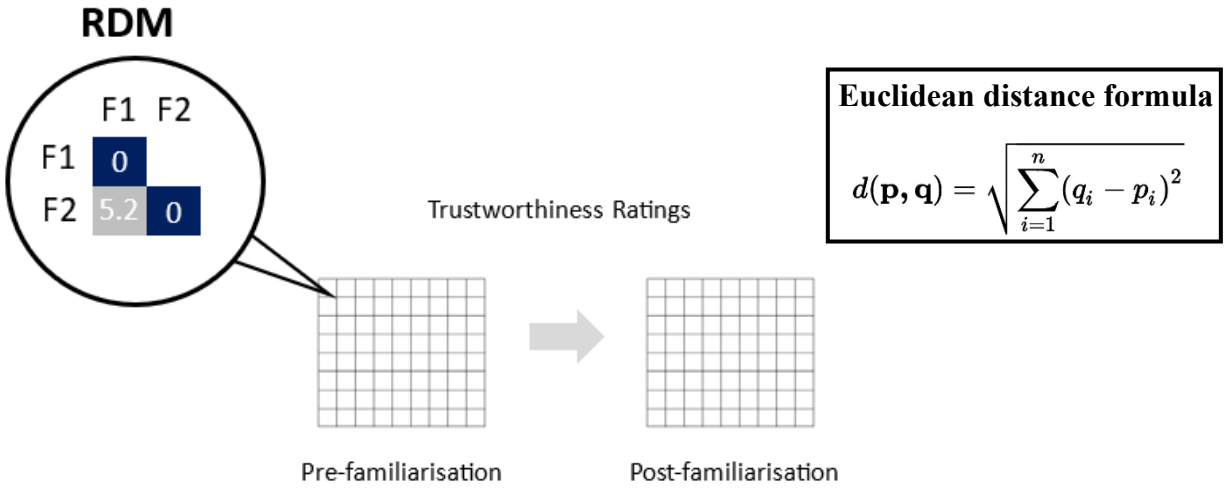
RDMs were computed for each participant and each trait, and separately for pre- and post-familiarisation. We aim to compare RDMs from pre- to post-familiarity to investigate how representations of each trait change with familiarisation. We believe that investigating changes

[1] Data from Chapters 3 and 4 are from the same experiment. Participants took part in both behavioural rating tasks and underwent fMRI scanning while viewing the same faces to obtain neural activity patterns. Data was pre-collected by Nadine Lavan and Olga Korokolva; all analyses and interpretation of data were carried out by the candidate.

at the level of representation allows for a more fine-grained analysis of potential changes as any changes can be assessed at once. Chapter 3 describes in full the methods that we used to compare these changes in representations.

Figure 3

Example diagram depicting the comparison between trait RDMs pre- and post-familiarisation.



Note. 8 x 8 RDMs were computed consisting of pairwise Euclidean distances between the faces based on all 4 social trait ratings. Analysis A: RDMs were compared between pre- and post-familiarisation sessions within each participant. This was carried out for each social trait, each participant and session. Analysis B*:* individual participants' RDMs were correlated with the average of all other RDMs. This was carried out for all 4 social traits and both sessions. RDM correlations were then compared across sessions. **C.** Analysis C*:* individual RDMs for each trait were correlated with each of the other traits, for each participant. Inter-trait correlations were then compared across sessions to assess whether the similarities in trait organisation changed. 'F1' = Face Identity 1; 'F2' = Face Identity 2.

2.1.2.2 fMRI - ROI RSA

Chapter 4 of the present thesis investigates the structure or geometry of brain representations of faces in face-selective ROIs and how these might change with face familiarisation. Brain fMRI activation patterns (vectors of fMRI activity in a group of voxels) were obtained from each ROI and were used to construct each participant's respective dissimilarity matrix (i.e., an RDM per ROI and per participant). The dissimilarity between the brain activation patterns for each pair of faces was computed using Linear Discriminant Contrast, a measure of similarity that uses multivariate noise normalisation (LDC; Nili et al., 2014; Walther et al., 2016), forming the entries in each RDM (see **Figure 4**). For example, the first entry in the RDM for the FFA was the dissimilarity in the response between Face Identity A and Face Identity B; all pair combinations were represented in the RDM. The RDM thus depicts the manner in which the FFA clusters or organises the set of faces.

RDMs were computed for each participant and each ROI, and separately for pre- and post-familiarisation. We aim to compare RDMs from pre- to post-familiarity to investigate how brain representations in each ROI change with familiarisation. There are multiple ways in which these ROIs potentially re-organise these faces as people become familiar with them and learn about them over a period of time. Chapter 4 describes the analyses that we performed to investigate these changes.

Figure 4

Example diagram depicting the comparison between ROI RDMs pre- and post-familiarisation.

Note. 8 x 8 RDMs were computed consisting of pairwise distances between the faces based on voxel activity patterns (VAPs) for each ROI and session, using Linear Discriminant Contrast (LDC). 'F1' = Face Identity 1; 'F2' = Face Identity 2.

2.1.2.3 Inter-Subject RSA

Chapter 5 utilises RSA in a different manner to that of Chapters 3 and 4, where similarities between individuals constitute the RDMs (see **Figure 5**). This has been called inter-subject RSA (IS-RSA; Finn et al., 2020), where the distances between individual participants are the main units of analysis for the RDMs. In Finn et al. (2020), the authors attempted to correlate between brain and behavioural RDMs which contain relative pairwise distances between participants. The aim was to ascertain whether people with similar brain activity patterns also exhibited similar personality trait dispositions and working memory patterns. In the fMRI scanner, participants watched 5 15-min movie clips of independent films, documentaries, and a montage of brief dynamic scenes with people and places. These contained some social and affective content so as to elicit related underlying neural activity. Outside of the scanner, participants completed self-report questionnaires on working memory and personality (60 items). The authors used two similarity measures to test personality similarity and working memory similarity.

First, the nearest neighbour (NN) model was used, going by the implicit assumption that those who score similarly in a domain would always 'be similar' to their 'neighbours' (i.e., pairs of people whose scores are close in, e.g., Euclidean space), irrespective of their absolute position on the scale. The second used Anna-Karenina (AnnaK) distance, which accounts for idiosyncrasies in the way that participants may be dissimilar from the rest of the sample. In other words, those who are different from others would be different in their unique way, which would be detected by AnnaK measures, and it was assumed that working memory differences would manifest in this manner and, thus, be best captured by this model. There were no strong predictions, however, about which would best capture personality. Pairwise similarities for personality and working memory were computed separately into RDMs, using the NN model and AnnaK similarity separately (resulting in 2 RDMs for each behavioural measure). These were each correlated with brain RDMs containing pairwise distances in brain activity patterns which were obtained during the fMRI experiment.

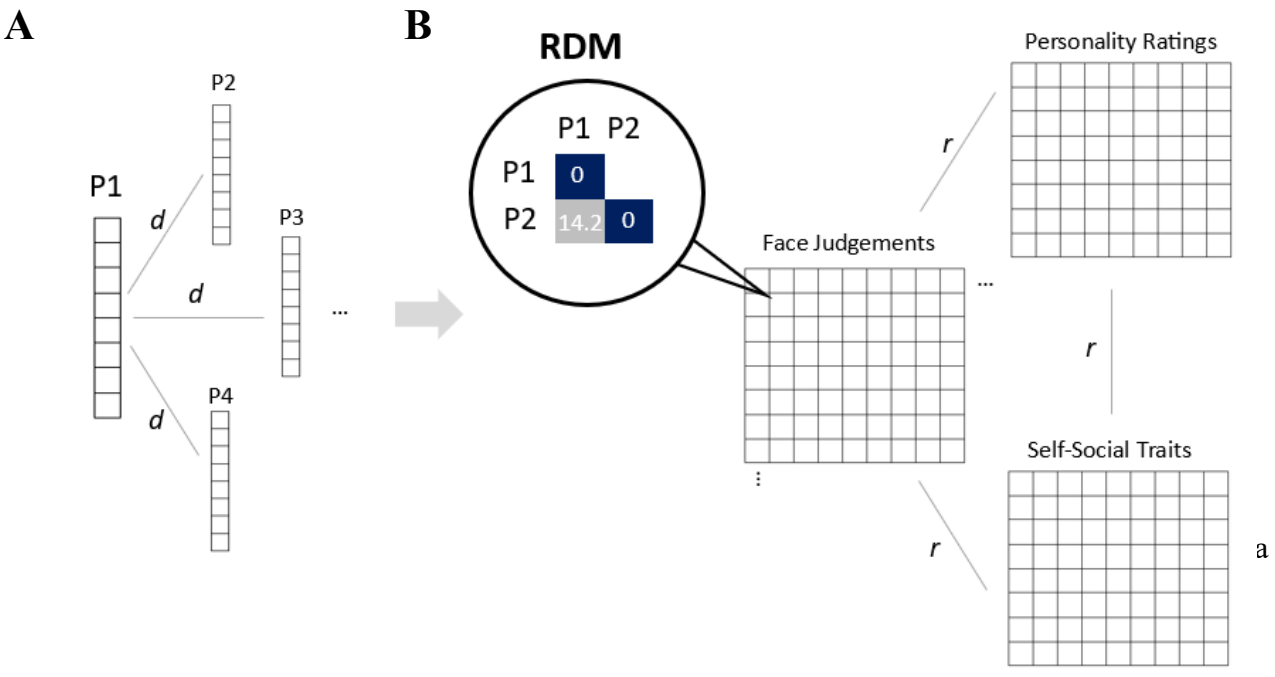
Results showed that working memory was indeed best predicted by the AnnaK model, showing that those who were more dissimilar were dissimilar from both high and low scorers on a scale. For the personality RDM computed using the NN approach, there was no significant correlation with the brain RDM, based on summary scores. However, when comparing based on specific items within the questionnaire, there was a highly significant correlation for 16 nodes in the brain. These include areas in the fusiform gyrus, right inferior frontal gyrus, and cerebellum – the first two of which are implicated in higher level face processing (Gobbini & Haxby, 2007) while the latter has been increasingly found to be implicated in social cognition (Van Overwalle et al., 2014). Though, only one node survived conservative correction measures. When using AnnaK distance, however, only 2 nodes were correlated at the uncorrected level. Finn et al. (2020) conclude that results for personality similarity might best be measured based on the more

subtle ways in which participants rate themselves, such as for specific items, rather than more general levels of the measure.

In sum, while personality may not be captured extremely well by either measures, it is best captured by the nearest neighbour approach (i.e., Euclidean distance). In Chapter 5[2] of the present thesis, we investigated whether similarities between participants in their personality/social dispositions were related to similarities in how they perceived social traits in faces. As the variables will be of a similar nature, some of the variance within these behavioural domains might be reduced when comparing across conditions. We built two types of RDMs: one for personality/social dispositions, and one for social judgments of others' faces. In both cases, each entry of an RDM corresponded to the pairwise distance between ratings from each participant, separately for face judgements, as well as personality/dispositions. The Euclidean distance (i.e., NN approach) between ratings was the dissimilarity measure we used.

Figure 5

Example diagram depicting vectors of ratings carried out for Face Judgements, Personality traits and Self-perceived traits.



Note. **A.** the Euclidean distance between vectors for each pair of participants were computed, producing **B**. an *N* x *N* RDM. This was carried out for Face Judgements, Personality Ratings and Self-Social Trait Ratings separately, resulting in 3 RDMs in total. 'P1' = Participant 1; 'P2' = Participant 2.

2.2 Conclusion

Overall, the studies carried out in this thesis aim to use RSA to investigate the representational structure of faces in multiple ways. RSA is a useful measure for ascertaining representations at the second level, which can be compared within domains, as well as between domains that otherwise could not be compared (i.e., brain and behavioural representations). Instead of observing face trait perception at the univariate level, we are able to see how the structure of these perceptions might manifest in certain contexts (i.e., behavioural ratings of face traits, brain activity) and even how these might change when compared across timeframes. IS-RSA is also a useful feature for capturing idiosyncratic patterns of similarity between people, particularly in domains where people might demonstrate high variability such as face trait perception and personality dispositions.

Chapter 3: Investigating how social trait representations of faces change with familiarisation

Abstract

Research shows that we tend to make social judgements about people immediately after seeing them for the first time. However, less is known about the way these impressions change as we become familiar with them. Chapter 3 investigates how social trait representations of faces change with increased familiarisation. Twenty-six participants rated 8 unfamiliar face identities from The Apprentice UK based on four traits: attractiveness, dominance, likeability, and trustworthiness. Participants then watched the series over 2-3 weeks, then re-rated the faces. We had 4 aims: first, we investigated whether there would be overall changes in trait ratings to assess whether opinions would merely increase due to familiarity, and results showed that dominance and trustworthiness ratings increased post-familiarisation. We then assessed the representational structure of these face traits for the next three analyses instead of using averaging across participants in order to understand how faces are perceived in relation to each other. When assessing within-person changes to the representational structure of these representations, only those of attractiveness and likeability remained stable post-familiarisation while dominance and trustworthiness representations did not, indicating a change in the perceptual clustering of faces based on who we trust and perceive to be dominant. We then assessed the agreement in trait representations between participants, both pre- and post-familiarisation, and then compared the agreement in each trait across sessions. Results showed that all trait representations except for

trustworthiness were significantly similar, but these remained stable across sessions, not significantly increasing or decreasing in similarity. For trustworthiness only, there was a significant increase in agreement between participants post-familiarisation, indicating a potential reliance on similar social information to determine who to trust. Finally, we investigated how the trait space (i.e., correlations between trait representations) changes with familiarisation: results showed that representations for likeability and trustworthiness became significantly more associated post-familiarisation, which may be due to post-familiarity trustworthiness representations becoming more similar to pre-familiarity likeability representations. In other words, we may begin to trust someone for the same reasons we initially liked them. Findings reveal how representations of faces based on social traits change as we become familiarised with them, indicating that changes may relate to newly ascertained person- and social-related information.

3.1 Introduction

Social trait evaluation of unfamiliar faces

Social judgements are important for everyday interactions and we seem to rely on a person's face as a source of social and personal information (Nguyen et al., 2009; Kwart et al., 2012; Varela et al., 2023). We also seem to develop these social perceptions immediately after seeing a face for the first time (Willis & Todorov, 2006). A study by Willis and Todorov (2006) measured social judgement ratings after being exposed to people's faces within various time constraints, based on a range of social traits, including trustworthiness, likeability, attractiveness, competence and aggressiveness. For one set of participants, 128 people rated the faces based on these traits using a scale from 1-9 with no time limits, while another set of participants rated the faces with different time limits, some rating with a limit of 100 ms, another of 500 ms, and another, 1000 ms. Participants also rated how confident they were in their judgements. To test the robustness of these judgements, ratings from the untimed condition were correlated with ratings from each of the timed conditions (i.e., ratings made in 100 ms, 500 ms and 1000 ms). No significant correlation would indicate that timed judgements may not particularly be reliable, as judgements in the untimed condition were made with enough time for a more genuine evaluation to take place. Correlations between the timed and untimed conditions would indicate the possibility to make robust judgements immediately upon seeing a face for the first time,

which was the case especially for timings made between 500 ms and 1000 ms. Results also showed that confidence levels had also remained stable within this timeframe. It appears that robust social judgements can be made in as quickly as 100 milliseconds, but also that people can be confident about these decisions in just half a second of viewing a face. This has implications for the importance of judgements made very quickly about people in various contexts and without prior information (e.g., choosing a mate [Walster et al., 1966], or determining prison sentences in court [Wilson & Rule, 2015]). But how do these judgements change over a longer period of time, after getting to know someone? Studies investigating the differences in the way we perceive unfamiliar and familiar faces based on social trait information are few.

Social trait evaluation from familiar faces

As discussed in the General Introduction, Peskin and Newell (2004) investigated the effects of face familiarisation on perceived attractiveness in two ways: 1. perceived familiarity, which can almost be likened to a sort of placebo or affective priming effect (Banse, 2001), and 2. Episodic familiarity, characterised by repeated visual exposure to the face thus provoking a sense of perceived face typicality (i.e., a 'feeling' of familiarity). This procedure thus explores face familiarity from two points of view, that is, based on low-level visual features of a face, and higher-level affective experiences or 'feelings' about a face. The aim was to explore whether initial ratings of face attractiveness would change with repeated exposure to faces, as well as based on feelings of familiarity. The authors found that attractiveness ratings had increased with repeated exposure to faces, as well as the affective experience of 'familiarity' whereby participants perceived that they had remembered images of students (which in reality whom were not presently attending their institution). In order to understand the true effects of familiarisation that goes beyond mere exposure, the actual *process* of familiarisation and acquaintanceship with a person needs to be studied, and the potential progression or change in perception thereof.

Familiarisation with faces often goes beyond processing the visual aspects of a face and involves encoding a person's identity, which typically relies on the internal features of the face (Bruce & Young, 1986). In order to emulate the process of face familiarisation using computational modelling, Kramer et al. (2018) submitted a series of face images to a model using combined principal components analysis (PCA; emulating bottom-up, lower-level processing) and linear discriminate analyses (LDA; emulating top-down, higher-level processing). The model was increasingly able to accurately cluster faces of the same identity together as faces were added to the model space, and was able to do so based on the internal face features. When PCA alone was tested, PCA failed to classify faces, indicating that the successful 'recognition' or classification of a face entails the reliance on inner face features to correctly match the encoded face with 'pre-stored' face representations.

What types of information beyond identity might we extract from familiar faces, and how do these differ between unfamiliar and familiar faces? Based on our own experience, we know that our first impressions about someone may change after getting to know them. A study by Mileva et al. (2019) investigated how our social trait perceptions might differ between unfamiliar and famous faces, particularly in the form of trait associations. The authors explored the inter-associations of five social traits: trustworthiness, dominance, attractiveness, extraversion and distinctiveness, and did this for male and female faces separately. For the first study, participants were presented with 40 unfamiliar face images of men and women and rated the faces on each of the five social traits. Results showed that across genders, there were many trait associations: attractiveness was significantly associated with trustworthiness and distinctiveness, and trustworthiness was significantly associated with extraversion. Participants were also presented with 40 famous faces (Hollywood stars, i.e., Scarlett Johansson and David Beckham) and rated these faces on each of the same 5 traits. For the famous faces, results showed that, also across genders, distinctiveness was significantly associated with all the other traits: attractiveness, trustworthiness, dominance, and extraversion. Attractiveness was also significantly associated

with trustworthiness and dominance. Firstly, results indicate that the face-trait structure (i.e., associations between traits) differs between faces we know and faces we do not. We may rely on physical face properties (Kramer et al., 2018) such as attractiveness to make fundamental judgements about unfamiliar people (Wilson & Eckel, 2006), including how trustworthy they are (Oosterhof & Todorov, 2008). While this was also the case for famous faces, there appears to be associations between all traits and distinctiveness; it could be that people perceive famous faces and make judgements about them based on some distinct quality(ies) they might have, which may relate to differentiated higher-level representations about these identities (i.e., conceptual rather than perceptual). This type of information is typically only accessible for familiar rather than unfamiliar faces (Bruce & Young, 1986; Kramer et al., 2018).

Further research is in line with the previous finding, namely, that the representational structure of social traits differs between famous and unfamiliar people. However, a particular focus was on the *concept* of various traits and how they might differ between unfamiliar and familiar people. Kim et al. (2022) investigated 13 traits (i.e., trustworthiness, dominance, intelligence, attractiveness; Oosterhof & Todorov (2008); Sutherland et al. (2013), the Big 5 traits; McCrae & Costa (1985); warmth and competence; Cuddy et al. (2008), and agency and experience; Gray, et al. (2007)). 1,985 participants were asked to imagine a person based on a particular valenced behaviour (e.g., "Imagine a person who spent a Saturday volunteering at a soup kitchen") and to then rate this person along a Big 5 trait dimension (e.g., "Please rate the openness to experience of this person"). The same procedure was carried out for unfamiliar emotionally-neutral *faces* with names, where a separate set of 654 participants were presented with 60 unfamiliar faces. Again, a face was presented with a face and name, and with a valenced behaviour description (e.g., Andrew spent a Saturday volunteering at a soup kitchen"), which was counterbalanced for positivity/negativity across participants. Results showed that ratings for faces and names versus those without faces and names were highly comparable (Kendall's $\tau = 0.857$, $p < .0001$), thus, the latter was used in subsequent analysis only to remove the variance in judgements associated

with visual face information (i.e., so as to maintain the variance associated with the abstract concept of the social traits). Participants were then asked to rate how similar unfamiliar people were based on the valenced behaviours.

For famous people, ratings were obtained from 869 people (taken from Thornton & Mitchell, 2018), and a separate set of participants rated how similar pairs of famous identities were.

Representational dissimilarity matrices (Kriegeskorte et al., 2008a) using the absolute difference in average similarity ratings based on all 13 traits (i.e., 13 x 13 RDMs) were computed. Each pairwise entry represents the difference between, for instance, trustworthiness and competence. Thus, there was one RDM for unfamiliar people and another for famous people. Pairwise trait ratings were then correlated within both unfamiliar and famous face RDMs independently and the two RDMs were subsequently compared using chi-squared tests to measure whether the two were conceptually represented in distinct ways. Results showed that the two RDMs were significantly distinct, where trait similarity ratings for unfamiliar people were more intercorrelated than for famous faces. Due to a large sample, and the consistency in the pattern of ratings between the different sets of participants, the findings illustrate the nature of our representations at the more societal level. Essentially, we appear to represent famous faces more idiosyncratically than unfamiliar people (Mileva et al., 2019). Additionally, valence appeared to explain much of the variance in the trait judgements for unfamiliar people, indicating that we may tend to rely on our perceptions of the moral goodness of people we don't know in order to make other social judgements about them, which is in line with previous models of face perception (Goodwin, 2015). Moral goodness appears to bear a strong influence on judgements about individuals, however, it has been argued that warmth may be more relevant for *interdependence* or social interactions (Goodwin, 2015), particularly when forming judgements about potential interactions between self and other, rather than judgements about another's acts of goodwill which may not necessarily involve the perceiver. Findings from Kim et al. (2022) may even be generalised to *faces* of the same people, as the correlations between ratings of the

people without names and faces, versus those with names and faces, were high. Nevertheless, the authors conclude that this distinction between unfamiliar and familiar people based on the representational structure of social traits may relate to relative differences in the *function* of face trait perception (Kim et al., 2022). In other words, we may rely on different social traits to make judgements about unfamiliar people compared to familiar people.

It is clear, then, that the way we more holistically perceive people, based on a set of social and personality traits, differs between unfamiliar and familiar individuals (Mileva et al., 2019; Kim et al., 2022). This difference is likely marked by a reliance on lower-level visual face properties for unfamiliar faces, while familiar face representations are likely to be primarily based on higher-level information, such as a person's identity and other related information (which is largely inaccessible for unfamiliar faces; Bruce & Young, 1986; Kramer et al., 2018). Unfamiliar faces are also consequently perceived to be more similar to each other than are familiar faces based on the valence of a face (Kim et al., 2022) and possibly other visual-based features (Tsantani et al., 2021). Familiar faces, on the other hand, may elicit more idiosyncratic pre-stored person knowledge, which subsequently moderates the information presently being extracted from the face.

Familiarity research into other modalities

Other recent studies have explored familiar person perception based on their voice (Lavan et al., 2021). Lavan et al. (2021) investigated how perceived trustworthiness and dominance might change with familiarisation to a person's voice. Are there characteristics in the voice that we might use to determine how trustworthy a voice is, and does this change with familiarity? A subset of 124 participants rated how trustworthy and dominant they perceived a set of unfamiliar voices to be, while another set of participants rated the set of voices after having been trained to recognise the speakers, during which they also learned semantic and/or biographical information about them. Some were trained to recognise voices with positively valenced information (i.e.,

"she is very patient when dealing with others' problems"), others with negatively valenced information (i.e., "she didn't apologise even when she knew she was wrong"), and yet others were trained with names only and no valenced information (i.e., "do you have two seagulls in the air?"). Results showed that trustworthiness ratings were influenced by the type of training group the participant belonged to: positively-associated voices were rated the highest in trustworthiness, then neutral voices, then negatively-associated voices were rated lowest. When comparing between the variability of trustworthiness ratings in the voices, results showed a difference between unfamiliar voices and both positively-associated as well as neutral voices. There was no difference, however, in the variability between unfamiliar and negatively valenced voices. Results indicate that we may generally find unfamiliar voices less trusting, or, at least, the variance associated with decreased trustworthiness during person perception might even be related in part to a person's voice.

Results for dominance voice perception only showed a significant increase in overall ratings with familiarisation to the voice; the variability in the ratings between the valenced conditions was not found for this trait. Even when testing participants by having them watch a show called Breaking Bad, there was no significant difference between the unfamiliar and familiarised voice identities in terms of trustworthiness and dominance; there were only differences detected that were related to the differences in identities. Lavan et al. (2021) conclude that, due to inconsistent findings across experiments, the influence of semantic knowledge on voice familiarisation should not be overestimated. Thus, further exploration into the effects of familiarisation on voices and other cross-modal representations is warranted.

Social trait representations, then, might be more highly associated with face rather than voice information. As previously stated, faces seem to provide useful information about a person's social tendencies and dispositions (Nguyen et al., 2009; Kwart et al., 2012; Varela et al., 2023), and this might occur to a larger extent than a person's voice. A study that found that participants recalled more personal details about a famous person (i.e., name and occupation) up to 92% of

the time when presented with their face compared to only 63% of the time when presented with a recording of their voice (Hanley et al., 1998). Thus, it could be the case that person-related information may also include social traits.

Based on the studies reviewed in the previous section, it is apparent that social trait representations are distinct between unfamiliar and familiar people (Kim et al., 2022), particularly for faces (Mileva et al., 2019), where we seem to extract more socially-relevant information, in comparison to other sources such as a person's voice (Hanley et al., 1998; Nguyen et al., 2009; Kwart et al., 2012; Lavan et al., 2021; Varela et al., 2023). Familiar faces likely elicit our pre-stored person knowledge about the identity (Bruce & young, 1986; Schwartz & Yovel, 2016), which appear to be more idiosyncratic than for unfamiliar faces (Mileva et al., 2019; Kim et al., 2022) and is likely to be composed of episodic experiences and semantic knowledge (Hanley et al., 1998). Notably, however, many studies of familiar face perception have often compared between unfamiliar faces and famous, or already-familiar faces (Hanley et al., 1998; Mileva et al., 2019; Kim et al., 2022). As a result, there is a gap in the literature regarding the ways in which these social trait representations *change* during the process of becoming familiar with a person. Much work on the mere exposure effect and similar concepts (Zajonc, 1968; Titchener, 1909, p. 408; Ladd & Gabrieli, 2015; Van Dessel et al., 2017; Siegel et al., 2019; Bornstein & Craver-Lemley, 2022) account for changes in ratings based on short-term, episodic exposure to faces. Findings show, however, that familiarisation is 1. a continuum (Kramer et al., 2018; Karimi-Rouzbahani, 2021; Ambrus et al., 2021; Li et al., 2022) and 2. eventually leads to consolidated face identity representations (Bruce & Young, 1986), even to an optimal point (Popova & Wiese, 2022, 2023). Kovács (2020) demonstrates that the brain represents different types of familiar faces distinctly (i.e., famous, personally familiar, contextually/para-socially familiar), which may be related to the ways in which these faces are learned. Therefore, investigation into the ways in which social trait representations of faces change over time is warranted.

The present study

The present study aims to investigate the way that social trait representations of faces change as we become familiar with people. Which trait representations actually change with familiarisation? Which first impressions stick? The present study employed a contextual/para-social familiarisation paradigm (Kovács, 2020, Ambrus et al., 2021) where participants watched a television series called The Apprentice UK (series 11) over the course of 2-3 weeks. Participants rated the faces of 8 identities that had remained on the show for longest (for context, one candidate from the show would leave at the end of each episode), and rated based on trustworthiness, dominance, attractiveness and likeability (Oosterhof & Todorov, 2008; Sutherland et al., 2013). We were interested in the change of social trait representations, both as a general trajectory but also as a *structure*. Specifically, we are interested in the way that faces are perceptually organised or clustered based on a set of traits (i.e., how are faces perceived in relation to each other based on trustworthiness, and how does this change with familiarisation?) (Kim et al., 2022). Thus, representational similarity analysis (RSA) was used to compare representations both prior to and after familiarisation in 3 of 4 of the analyses below (Kriegeskorte et al., 2008); there were 4 aims for the present study:

1. Measure overall changes in trait ratings across participants and for each of the traits, after two weeks of familiarisation. How much do first impressions matter? We do not expect there to be a significant change for trait attractiveness, specifically, as Tsantani et al. (2021) show that attractiveness representations are highly implicated in a brain region that deals primarily with image-based properties. Thus, it seems reasonable to assume that perceived attractiveness may not change over an extended period, especially as representations may become more stable over this period, in contrast to the MEE effect (Bornstein, 1989; Peskin & Newell, 2004; Ladd & Gabrieli, 2015; Van Dessel et al., 2017; Siegel et al., 2019; Bornstein & Craver-Lemley, 2022). We expect there to be an increase, however, in overall ratings for the other traits (trustworthiness, dominance and

likeability) based on Lavan et al. (2021). Lavan et al. (2021) demonstrated a change post-familiarisation for voices, and we expect the same to happen for the face modality where there is likely to be a richer set of person-related information extracted from the face (Hanley et al., 1998). We also expect a significant increase since familiarisation takes place over a longer period.

2. Investigate whether the representational structure of face traits change within participants. We expected the structure for all traits to change except for attractiveness (Tsantani et al., 2021) for similar reasons to the first aim, though we have no predictions about which of these trait representations will change more.

3. Explore consistencies *between* the participants, both initially and post-familiarisation. We expect that there will be initial agreement to a significant degree between participants but also that this will further increase post-familiarisation for all traits (except for attractiveness; Tsantani et al., 2021). While research indicates possible idiosyncratic representations of familiar faces (Mileva et al., 2019; Kim et al., 2022), given that participants would have access to the same social information, we predict that the representations would become more synchronised. According to our knowledge, this is a novel research aim.

4. Assess the representational structure of faces between the traits themselves (i.e., what do inter-correlations initially look like, and how does this structure change with familiarisation?). Based on Mileva et al. (2019), we expect that a) the relationship between attractiveness and likeability RDMs, and b) attractiveness and trustworthiness RDMs would decrease with familiarisation, as a reliance on more abstract representations would increase while the predominant reliance on face features decreases (Bruce &

Young, 1986). We also expect that c) the relationship between trustworthiness and likeability RDMs would initially be high, based on high correlations between the two traits seen in Oosterhof and Todorov (2008). We had no other predictions regarding the other pairwise trait correlations.

3.2 Methods

3.2.1 Participants

Thirty-one participants aged 18 to 27 were recruited from Royal Holloway, University of London and Brunel University London (*male* = 5, *female* = 26; *Mean age* = 21.35, *SD* = 2.68; *age range* = 18-27). The data for five participants was removed due to missing data. Thus, data for a final sample of 26 participants was analysed (*male* = 3, *female* = 23; *Mean age* = 20.85, *SD* = 2.17; *age range* = 18-27). Ethics was approved by the Research Committee at Brunel University London (see **Appendix A**). All participants provided written consent before participating (see **Appendix B**).

Inclusion criteria. Participants were required to be between 18 and 30 years of age and to have lived in the UK for at least 10 years. This ensures the participant sample and stimuli set (based on a UK television series) are matched and representative of the geographical region where the present study had taken place. All data was collected by Nadine Lavan and Olga Korolkova.

Participants took part in the study that involved behavioural testing and structural and functional MRI (results for the brain data are reported in Chapter 4). Participants also watched The Apprentice UK (Series 11), a British series about people who complete tasks in teams with the

aim to become the next apprentice for multi-millionaire Lord Alan Sugar. This was our familiarity task in which we reasoned that participants would become familiar with the characters in the show. Participants watched the whole of series 11 over the course of two weeks at home. Participants took on average 20.76 days (*SD* = 1.81) to watch the series. The testing sessions happened before watching the show (pre-familiarity) and after watching the show (post-familiarity).

After watching each episode, participants completed questions online on Qualtrics (www.Qualtrics.com) about events that had happened in the show (e.g., "What was the reward back at the house for the winning team?"). This was to test whether participants had paid attention to each episode. Participants were also asked whether they recognised each person in the show and were asked to rate them on social traits.

This chapter focuses on the behavioural data collected in the two sessions in the lab (i.e. testing happened in person), i.e., the sessions that happened pre- and post- watching the show (pre- and post-familiarity).

3.2.2 Design

A within-subjects design was used for the present study looking at the effects of face familiarisation (pre- and post-familiarisation) on the perception of four social traits (attractiveness, dominance, likeability, and trustworthiness). Ratings on each trait and identity were acquired pre- and post-familiarity. Change in representations of these traits will be measured in different ways, as explained for each of the analysis below.

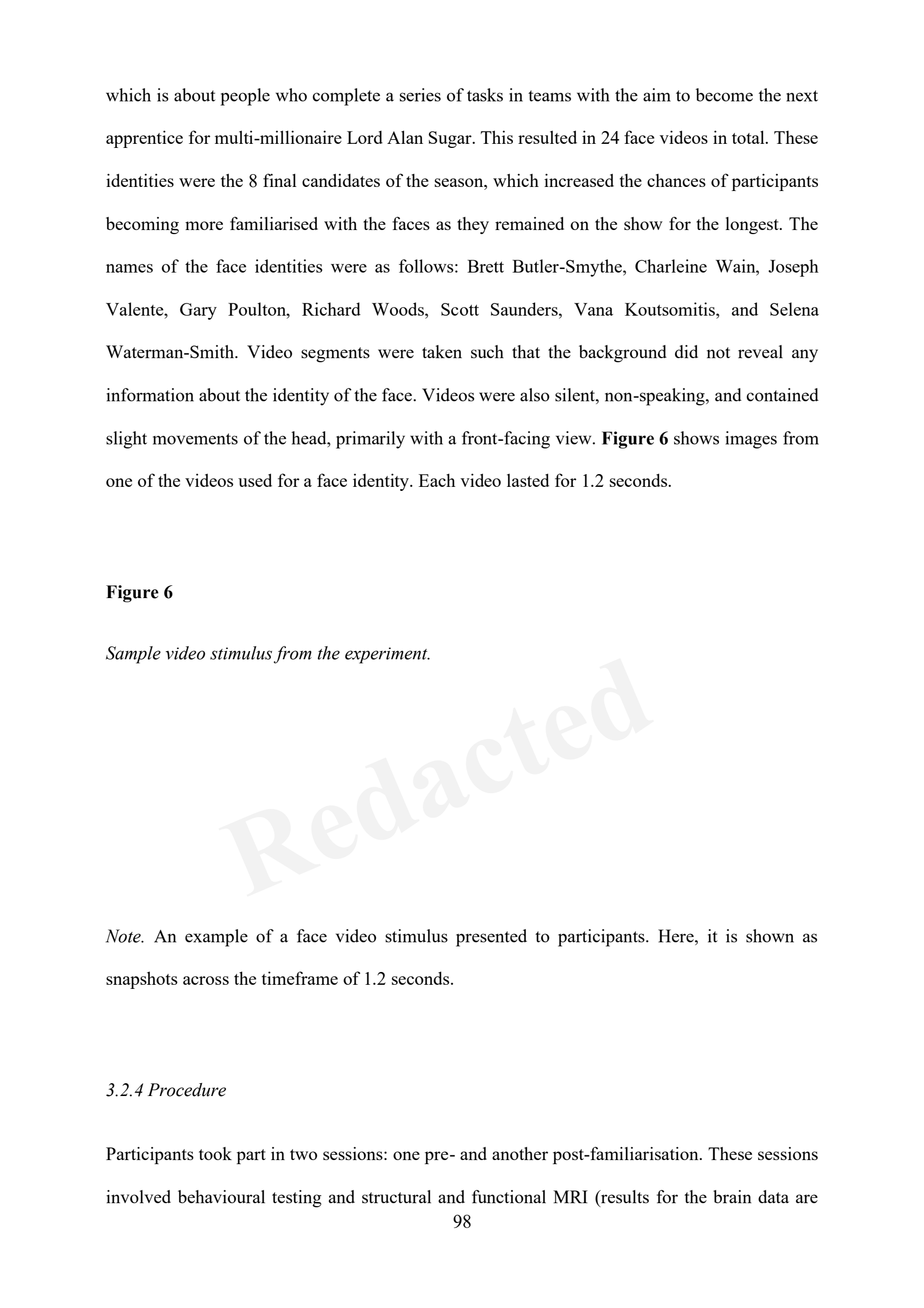
3.2.3 Materials

Three individual face videos (i.e., 3 tokens) of each face identity were taken of 8 adult identities (*Gender*: 5 men, 3 women; *Ethnicity*: 7 white, 1 mixed) from the British series 'The Apprentice',

which is about people who complete a series of tasks in teams with the aim to become the next apprentice for multi-millionaire Lord Alan Sugar. This resulted in 24 face videos in total. These identities were the 8 final candidates of the season, which increased the chances of participants becoming more familiarised with the faces as they remained on the show for the longest. The names of the face identities were as follows: Brett Butler-Smythe, Charleine Wain, Joseph Valente, Gary Poulton, Richard Woods, Scott Saunders, Vana Koutsomitis, and Selena Waterman-Smith. Video segments were taken such that the background did not reveal any information about the identity of the face. Videos were also silent, non-speaking, and contained slight movements of the head, primarily with a front-facing view. **Figure 6** shows images from one of the videos used for a face identity. Each video lasted for 1.2 seconds.

Figure 6

Sample video stimulus from the experiment.



Note. An example of a face video stimulus presented to participants. Here, it is shown as snapshots across the timeframe of 1.2 seconds.

3.2.4 Procedure

Participants took part in two sessions: one pre- and another post-familiarisation. These sessions involved behavioural testing and structural and functional MRI (results for the brain data are

reported in Chapter 4). For the behavioural segment, participants provided social trait ratings of face and voice stimuli of 8 identities from a British television series called The Apprentice UK (Series 11).

First session: pre-familiarisation

Behavioural task. In the first session, participants provided informed consent. For the face rating segment, participants were presented with face stimuli using Psychtoolbox (version 3; Brainard, 1997; Pelli, 1997) on MATLAB on a computer screen. Participants viewed face videos that were presented for 1.2 seconds each. There were 3 silent videos for each face identity (i.e., 3 tokens per identity) which consisted of the face slightly moving. Videos were cut to mostly display a front-facing view of the face. Participants viewed one face at a time and were presented with a question about a trait (e.g., "How Trustworthy is… Vana?"), to which they rated the face on one of 4 social traits, including trustworthiness, dominance, attractiveness and likeability, using a scale of 1 (i.e., not trustworthy at all) to 7 (i.e., very trustworthy). Face videos were randomised across all identities and tokens within each block of traits (i.e., all 3 face video tokens for all 8 identities were randomly presented first with a question about, e.g., trustworthiness, then for dominance, and so on). Thus, each face and trait question constituted 1 trial each. Each participant completed a total of 96 face trials (3 videos, 8 identities, 4 traits, 1 run for face presentation). The order of trait blocks was randomised across participants.

For the voice rating segment, the same procedure was carried out using tokens of the same voice identities and question prompts using the same traits.

fMRI task. Participants then underwent structural and functional magnetic resonance imaging (fMRI) where face and voice stimuli were presented in separate runs. The same procedure as in the behavioural task was carried out for both faces and voices in the scanner: one run for viewing

the face stimuli, and another run for listening to the voice stimuli of the same identities. Runs were counterbalanced across participants.

Interim session: watching the TV series

Participants were provided with a USB stick containing instructions for the study, links to all 12 episodes of the show, and questionnaires about the identities from the show (see **Appendix C**). Participants watched the whole of The Apprentice UK series at home over 2 weeks. This was our familiarity task as we reasoned that participants would have become familiar with the characters in the show during this time. To ensure that participants were watching the show during the 2 weeks, they were asked 5 simple trivia multiple choice questions using a questionnaire platform called Qualtrics (www.qualtrics.com). Example questions include "Which market did the contestants go to buy most of their ingredients?", to which participants would select the correct answer. Participants also provided 8 interim questions for each of the 12 episodes asking how familiar they were with each identity (i.e., "How familiar are you with… Gary?") and 24 trait questions for 3 traits, including trustworthiness, dominance and likeability (i.e., "How dominant is… Richard?"). Attractiveness was not included here as changes in perceived attractiveness were not expected to occur (Tsantani et al., 2021). Participants rated the identities on familiarity and social traits using a 7-point scale. Participants took on average of 20.76 days (SD = 1.81) to watch the show.

Second session: post-familiarisation

After watching the show, participants were invited back into the lab to provide ratings and undergo fMRI scanning while viewing the same faces and listening to the same voices for the second time. This session was identical to the first. Participants were then debriefed and compensated for their time.

Data for neural activity acquisition are presented in Chapter 4. Data for voice trait ratings are not presented in this thesis.

3.2.5 Data Analysis

Mean Ratings

We first measured the overall change in average ratings from pre-familiarisation to post-familiarisation, for each trait. We were interested in whether trait ratings would increase and become more positive (or negative) with familiarisation. In other words, how much do first impressions matter? For each participant, trait ratings were averaged for each of the three video stimuli per identity, thus resulting in an average rating per identity and per trait. Means were calculated for ratings collected both before and after watching the show (pre- and post-familiarisation). Mean ratings were then compared between sessions using non-parametric Wilcoxon signed-rank tests. Non-parametric tests are common in RSA as it is assumed that data structures are non-normally distributed, due to interdependent pairwise values within the matrices (Diedrichsen & Kriegeskorte, 2017). We also computed mean ratings for each trait (except attractiveness) after each of the 12 episodes across face identities, as well as for each identity separately, to observe any variance in the trajectory of ratings over time.

Representational dissimilarity matrices (RDMs)

Subsequent analyses used the Representational Similarity Analysis framework (Kriegeskorte et al., 2008). We measured representations for each trait as the dissimilarities between the eight identities in that trait space. The dissimilarities were computed using the Euclidean distances between pairs of average ratings for each identity (for example, the distance between the average attractiveness rating for Brett (e.g., 5) and for Vana (e.g., 6.5) was computed and displayed as a single value in one entry of the attractiveness matrix). The resulting representational dissimilarity matrices (RDMs) for each trait and each participant were used in the following analyses.

Analyses were carried out in MATLAB, version R2021b.

3.3 Results

3.3.1 Analysis A: *Mean ratings: overall change with familiarisation.* Mean ratings for each trait are shown in **Figure 7**. We used Wilcoxon signed-rank tests to compare differences in ratings across sessions for each trait. Results showed that ratings significantly increased for dominance and marginally for trustworthiness. Ratings for attractiveness and likeability did not significantly change from pre- to post-familiarity, suggesting that these ratings were quite stable (see **Table 1** and **Figure 7**).

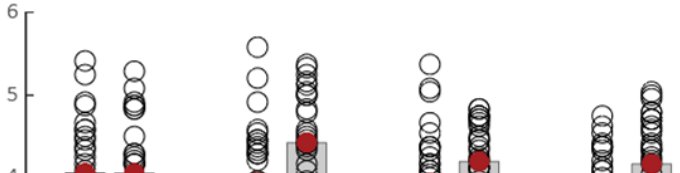
Table 1

Analysis A: Wilcoxon signed-rank tests: z- and p-values between pre- and post-familiarisation sessions for each trait across face identities.

Trait	Pre-Familiarisation	Post-Familiarisation		
	Means (SD)	Means (SD)	*z*-values	*p*-values
Attractiveness	4.06 (.78)	4.05 (.85)	.11	.91
Dominance	3.94 (.93)	4.24 (.74)	-2.35	.018
Likeability	3.94 (.78)	4.20 (.78)	-1.47	.14
Trustworthiness	3.88 (.60)	4.17 (.86)	-1.99	.046

Figure 7

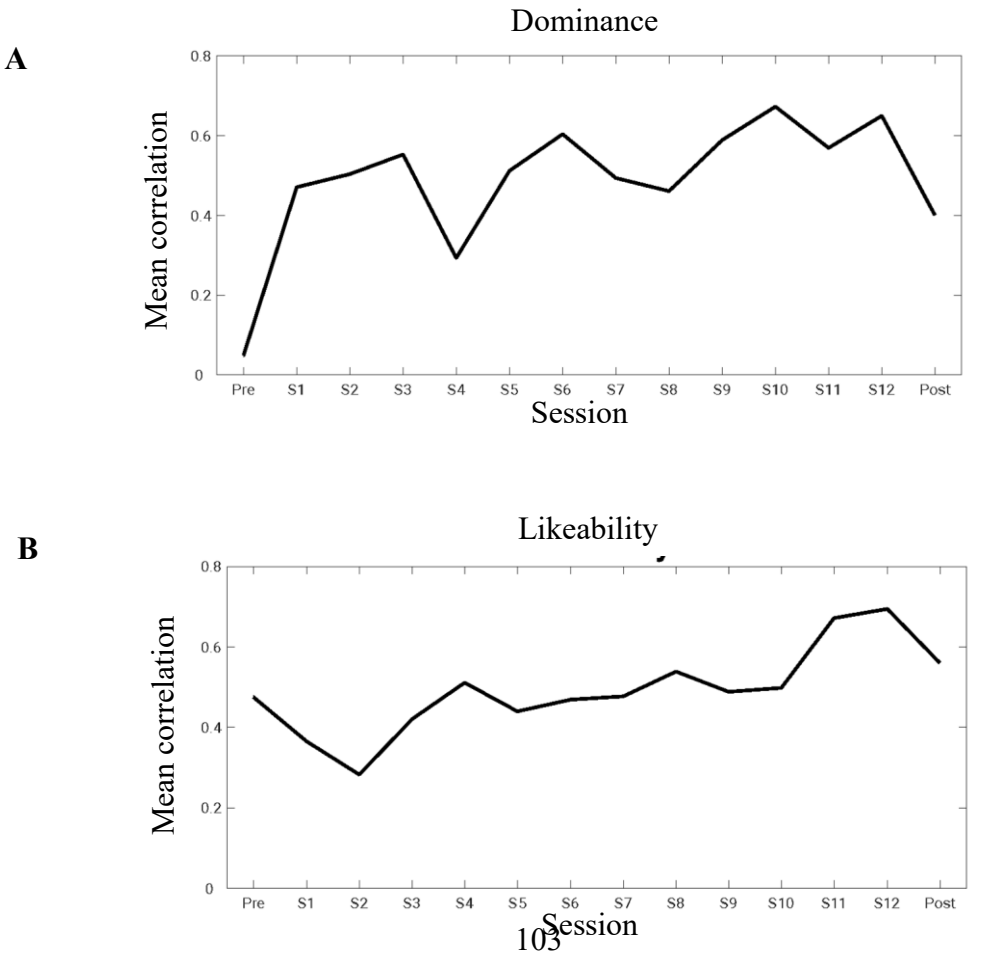
Analysis A: Mean ratings of all face identities, for all traits, across sessions (pre- and post-familiarisation).



Note. Red dots indicate the mean rating across participants for that trait; black circles indicate mean ratings for each participant for that trait.

Figure 8

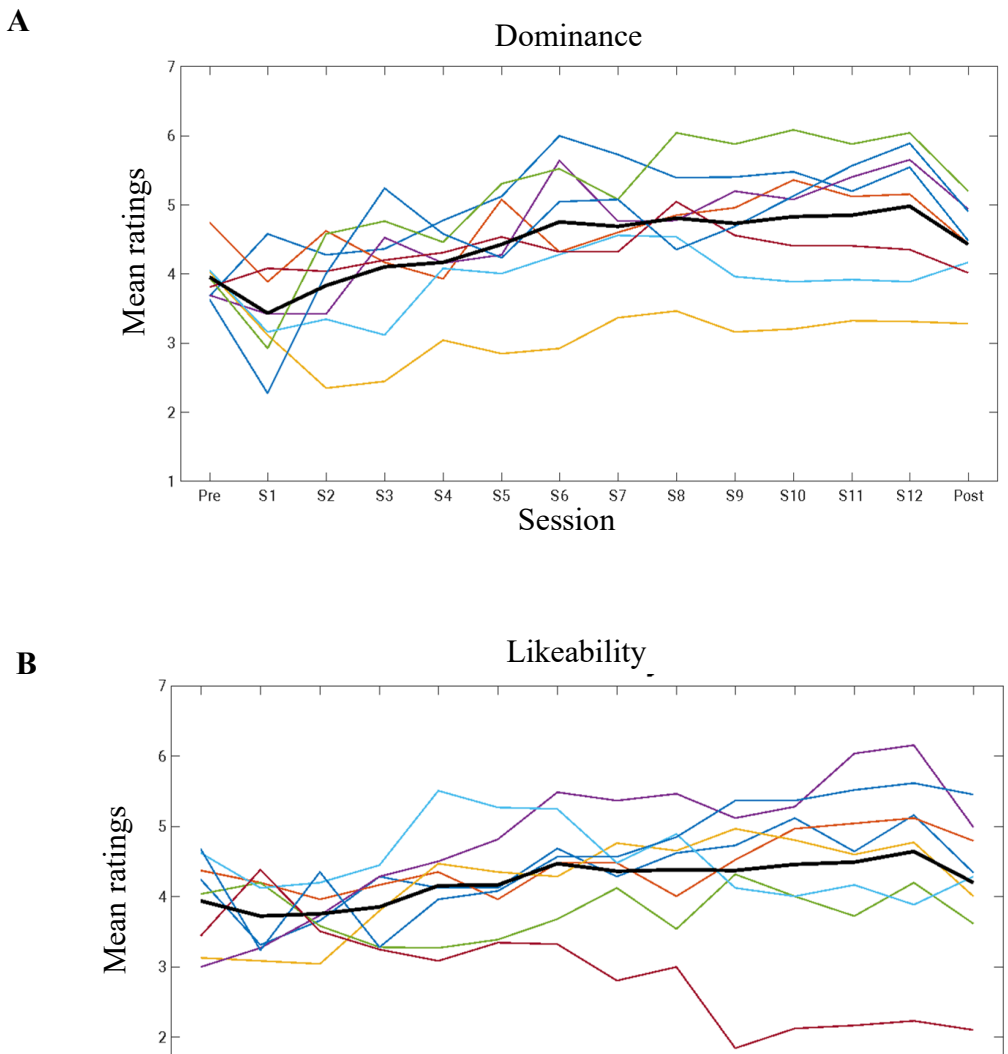
Analysis A: Trajectory of mean trait ratings across 12 episodes collapsed across all 8 face identities, for dominance, likeability and trustworthiness.



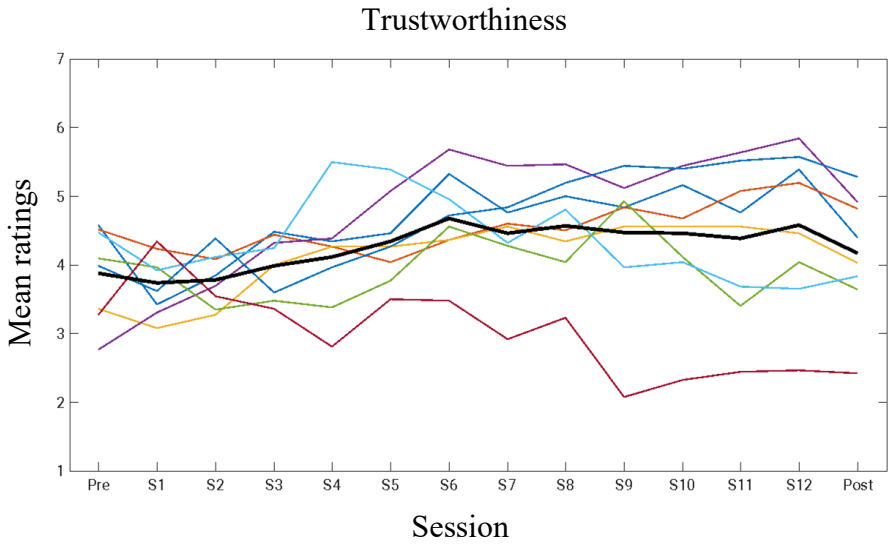
Note. Lines represent the mean correlation across face identities for **A.** dominance, **B.** likeability and **C.** trustworthiness over 12 episodes of The Apprentice UK (series 11).

Figure 9

Analysis A: Trajectory of mean trait ratings for each of the 8 face identities across 12 episodes, for dominance, likeability and trustworthiness.



C



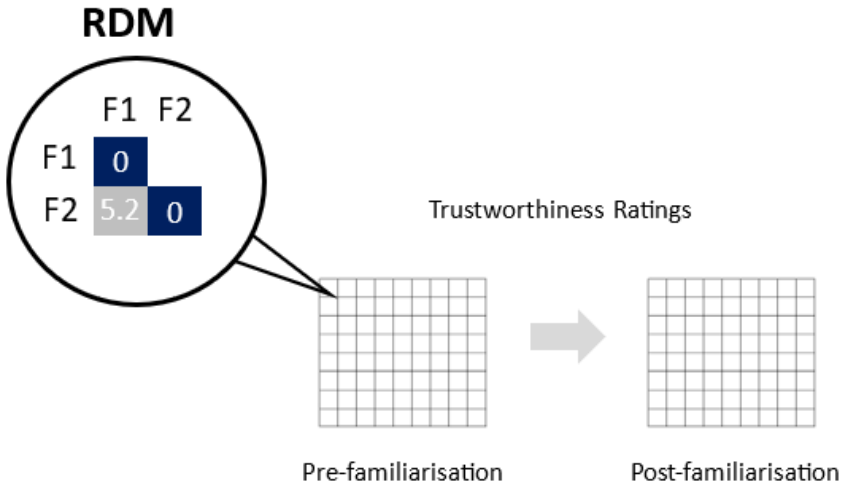
Note. Each coloured line represents the mean rating for each of the 8 identities; the black line represents the mean of all ratings collapsed over identities, each for **A.** dominance, **B.** likeability and **C.** trustworthiness. Ratings represent means over 12 episodes of The Apprentice UK (series 11).

3.3.2 Analysis B: *Comparing representations within-participants pre- and post-familiarisation.*

Our second analysis compared the representations of identities for each trait between pre- and post-familiarisation (see **Figure 10**). RDMs depicted the Euclidean distances between ratings of each identity, per trait, and were computed for each participant, separately for pre- and post-familiarity. Then, for each participant and each trait, we correlated the pre-familiarity RDM with the post-familiarity RDM. Mean correlations across participants for each of the traits are shown in **Table 2** and **Figure 11**: higher correlations indicate more stability in the representations of face-traits post-familiarisation. Wilcoxon signed-rank tests were used to compare participants' correlations for each trait against 0. The highest correlation was for attractiveness, followed by likeability, which were both significantly higher than zero, suggesting that representations of attractiveness and likeability were relatively stable from pre- to post-familiarity. For trustworthiness and dominance, correlations pre-and post-familiarisation were not significantly different from 0, suggesting a change in representations across sessions. See **Figure 12** for a visualisation of RDMs from Participant 1.

Figure 10

Analysis B: Example diagrams depicting the within-participant comparison between trait RDMs pre- and post-familiarisation.



Note. **Analysis B:** RDMs were computed for each social trait and participant consisting of pairwise distances between ratings of the faces. RDMs were compared between pre- and post-familiarisation sessions within each participant and for each trait. 'F1' = Face Identity 1; 'F2' = Face Identity 2.

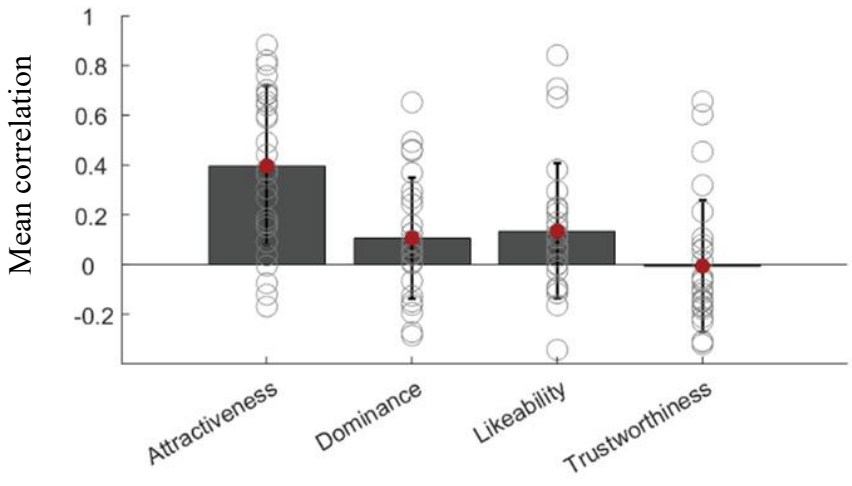
Table 2

Analysis B: One-samples rank tests for within-participant analyses: means, standard deviations, and z-values comparing RDM correlations for each trait to zero.

Trait	Mean (SD)	*z*	*p*
Attractiveness	.40 (.32)	4.03	.01
Dominance	.12 (.24)	1.92	.055
Likeability	.14 (.27)	2.37	.02
Trustworthiness	-.01 (.26)	-.90	.37

Figure 11

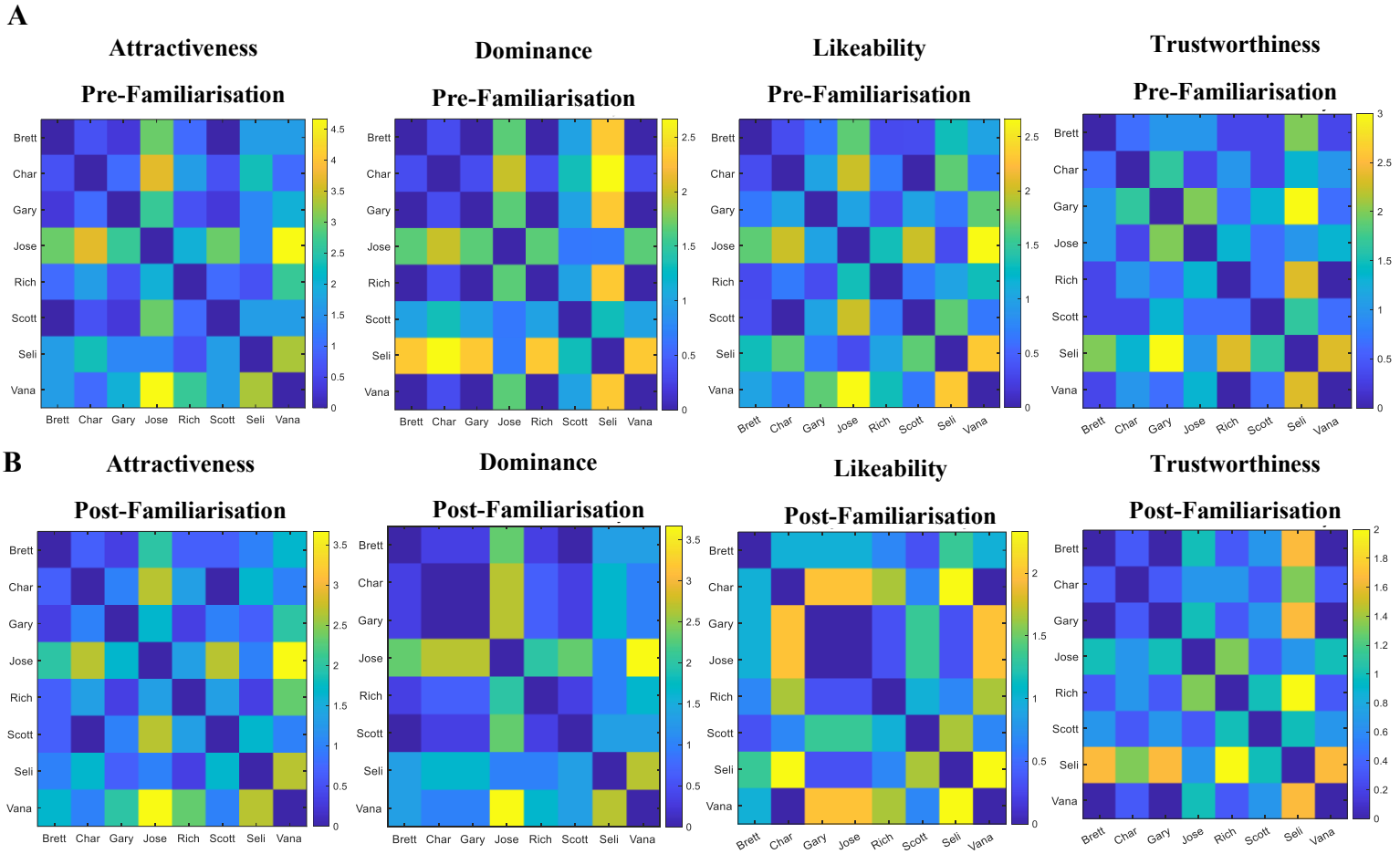
Analysis B: Mean correlation between pre- and post-familiarisation RDMs within participants.



Note. Within-participant correlations: mean correlations across sessions for all traits (attractiveness, dominance, likeability and trustworthiness). Red dots indicate the mean correlation between RDMS across all participants; grey circles indicate individual mean correlations across sessions for each participant.

Figure 12

Analysis B: RDMs depicting the Euclidean distance between ratings of each of the face identities, for each trait both pre- and post-familiarisation, for Participant 1.



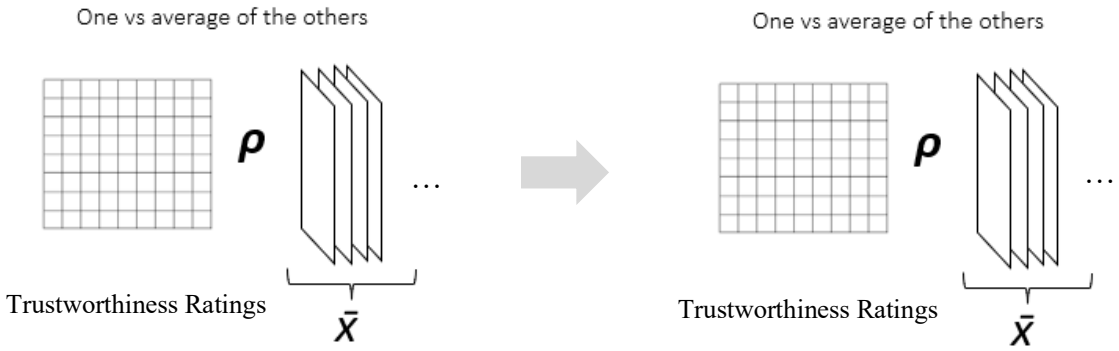
Note. **A:** Representational dissimilarities matrices of identities for each of the traits (attractiveness, dominance, likeability and trustworthiness) pre-familiarisation. **B:** RDMs for identities for the same traits post-familiarisation. All RDMs depicted here from Participant 1.

3.3.3 Analysis C: *comparing inter-subject similarity of representations pre- and post-familiarisation*. Our third analysis investigated the similarity between participants' representations for each trait and session (see **Figure 13**). Pre-familiarity RDMs for each participant were correlated with the average of everyone else's pre-familiarity RDMs, and the same was carried out for post-familiarity RDMs. Mean correlations for each trait and session are shown in **Figure 14** and **Table 3**. Correlations were subsequently compared to 0 using Wilcoxon rank tests. Correlations significantly higher than zero were interpreted as showing high agreement between participants' RDMs. Results showed that correlations across participants were indeed significantly higher than 0 for attractiveness, dominance and likeability, both pre-familiarisation and post-familiarisation. For trustworthiness, however, correlations were not significantly different from 0 pre-familiarisation but only post-familiarisation.

We then tested whether this agreement (or lack thereof) changed significantly among participants post-familiarisation. For each trait, we used Wilcoxon signed-rank tests to compare the agreement across sessions. A significant difference would indicate the agreement changed after participants watched the show. In other words, whether participants agreed or disagreed in the way they represented the faces *prior* to watching the show would not have changed *after* watching the show. Results (see **Table 3**) indicated there was only a significant difference in the agreement for trustworthiness across sessions, indicating an increase of agreement across participants post-familiarisation.

Figure 13

Analysis C: Example diagram depicting the agreement in trait RDMs between participants separately for pre- and post-familiarisation sessions.



Pre-familiarisation

Post-familiarisation

Note. **Analysis C**: 8 x 8 RDMs were correlated between each participant and the average of all the other participants, thus correlating RDMs in the third dimension. Between-participant

correlations were carried out for each of the 4 traits, for each participant, and for each session. Correlations were then compared across sessions.

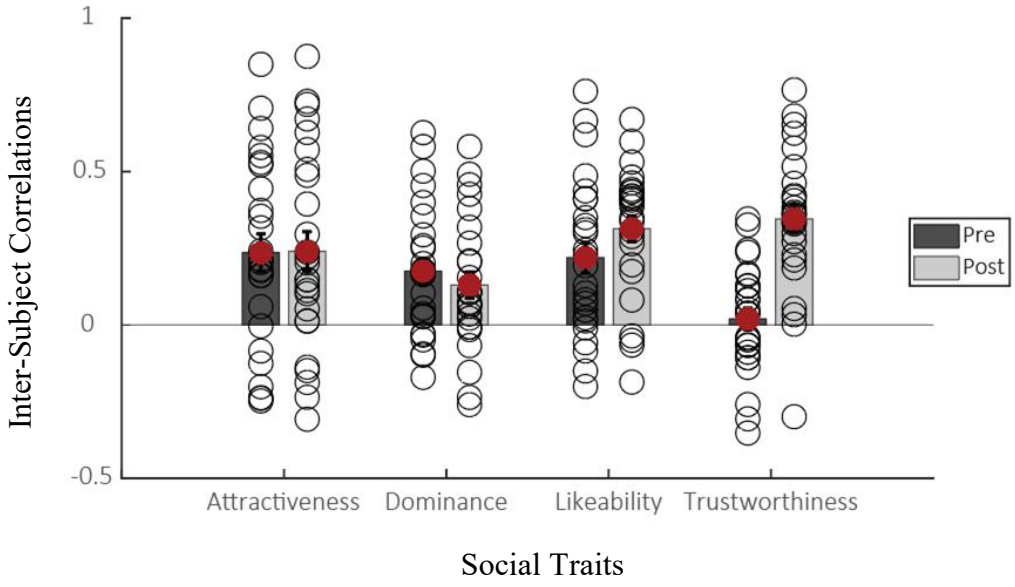
Table 3

Analysis C: means, standard deviations, and z-values for one-samples and paired samples rank tests comparing inter-subject RDMs per trait across sessions.

Trait					Wilcoxon signed-rank tests			
					One-samples rank tests		Paired-samples rank tests	
	Pre		Post		Pre	Post		
	Mean (SD)	*p*	Mean (SD)	*p*	*z*-value	*z*-value	*z*-value	*p*-value
Attractiveness	.24 (.31)	.003	.24 (.33)	.003	3.01	2.96	-.22	.83
Dominance	.18 (.21)	.0007	.13 (.22)	.005	3.39	2.81	.34	.73
Likeability	.22 (.25)	.0005	.31 (.21	.0003	3.49	4.15	-1.59	.11
Trustworthiness	.02 (.18)	.53	.35 (.23)	.00002	.62	4.23	-4.25	.00002

Figure 14

Analysis C: Inter-subject correlations for each trait both pre- and post-familiarisation.



Note. **Analysis C:** bar chart for average correlations between each participant's representations for each trait (attractiveness, dominance, likeability, and trustworthiness) and the averaged of all other participants, separately for pre- and post-familiarity sessions superimposed dot plots of average and individual participants' correlations for each trait and session. Error bars show standard deviations across participants. Red dots depict the mean correlation and empty circles show the correlations for individual participants.

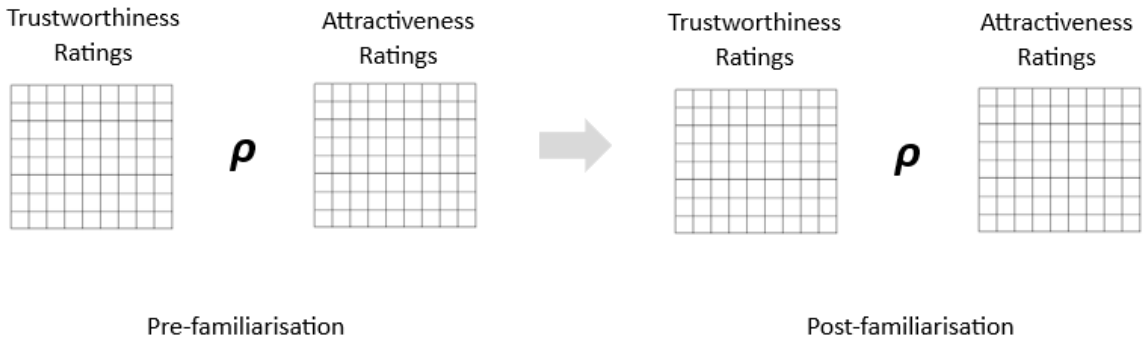
3.3.4 Analysis D: *comparing relations between traits pre- and post-familiarity.* The final analysis investigated the structure or organisation of trait space (i.e., the relationship between traits themselves) and the change in this structure post-familiarisation (see **Figure 15**). In other words, we investigated whether the inter-correlations between the traits changed with familiarisation. For each participant, and separately for pre- and post-familiarity trait RDMs, we correlated each trait RDM with the RDMs of each of the other traits. Mean correlations for each pair of trait

RDMs are shown in **Table 4** and **Figure 16**. We subsequently tested whether these correlations were significantly higher than 0, which would suggest overlapping representations for a pair of traits. Results showed that the highest correlations were between likeability and trustworthiness RDMs, both pre- and post-familiarisation. Correlations between attractiveness and likeability RDMs were also significantly above zero for both pre- and post-familiarisation, as well as for attractiveness and trustworthiness across sessions. Dominance only became significantly associated with likeability and trustworthiness post-familiarisation.

Wilcoxon signed-rank tests were used to test whether there was a significant difference in each of the relationships, or pairwise similarity, of these trait structures (i.e., did the trait relationships change post-familiarisation?). There was only a significant difference in the correlation between likeability and trustworthiness RDMs, indicating a significant increase in the magnitude of the correlation post-familiarisation (see **Table 4**).

Figure 15

Analysis D: Example diagram depicting the similarity between trait RDMs, separately for pre- and post-familiarisation sessions.



Note. **Analysis D:** individual RDMs for each trait were correlated with each of the other traits, for each participant. Inter-trait correlations were then compared across sessions to assess whether the similarities in the face-trait organisation changed.

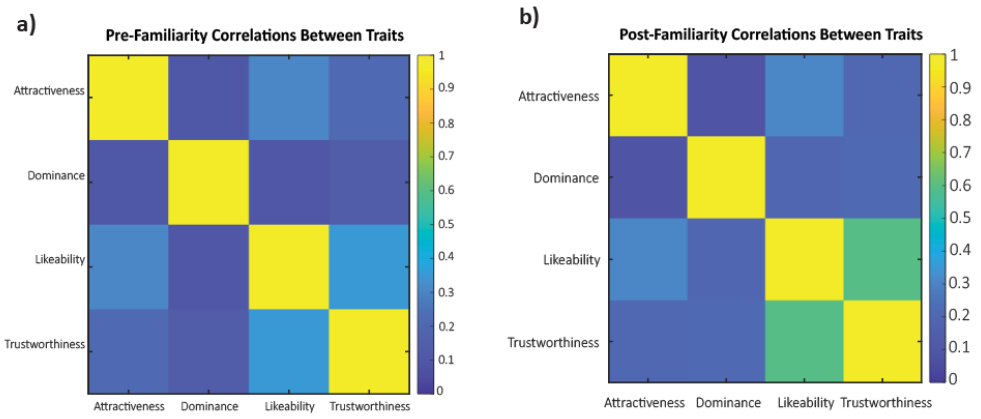
Table 4

Analysis D: Means, standard deviations, and z-values for one-samples and paired-samples rank tests, for each pair of trait RDMs, for each session, and across sessions.

Trait	Means (SD)				Wilcoxon signed-rank tests			
					One-samples rank tests		Paired-samples rank tests	
	Pre		Post		Pre	Post		
		p-values		*p*-values	*z*-values	*z*-values	*z*-values	*p*-values
Attractiveness x Dominance	.10 (.24)	.052	.08 (.25)	.17	1.94	1.38	.83	.41
Attractiveness x Likeability	.31 (.26)	.00002	.31 (.32)	.0004	4.10	3.57	-.06	.95
Attractiveness x Trustworthiness	.21 (.35)	.009	.21 (.34)	.008	2.60	2.65	.27	.79
Dominance x Likeability	.10 (.27)	.14	.20 (.24)	.001	1.46	3.24	-1.10	.27
Dominance x Trustworthiness	.13 (.31)	.10	.21 (.28)	.002	1.64	3.06	-1.05	.29
Likeability x Trustworthiness	.35 (.35)	.0002	.60 (.26)	.000008	3.70	4.46	-2.55	.01

Figure 16

Analysis D: Mean RDMs depicting pairwise correlations between Attractiveness, Dominance, Likeability and Trustworthiness, both pre- and post-familiarisation (across participants).



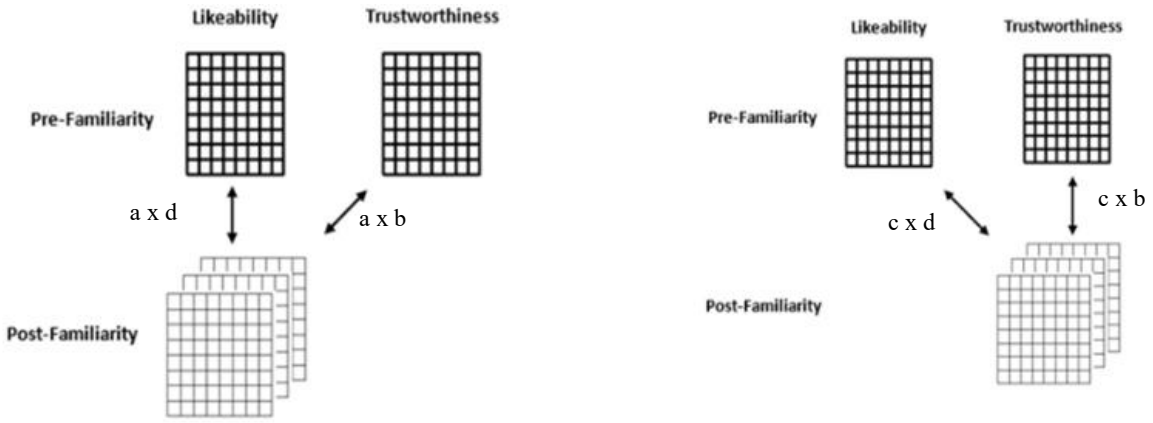
Note. Pre- and post-familiarity RDMs depict the representational structure of the traits, where the traits distances are depicted in each entry rather than face identities. ***a)*** *pre-familiarity RDM depicting the pairwise distance between the traits pre-familiarity, and* ***b)*** *post-familiarity RDM.*

3.3.5 Additional analyses. With the increase of pairwise correlations between likeability and trustworthiness post familiarity, we wanted to investigate whether representations became more influenced by one trait or another. To explore the directionality in the relations between likeability and trustworthiness RDMs, *individual* post-familiarity RDMs were compared with *averaged* pre-familiarity RDMs (i.e., within- and across traits in a criss-cross manner). Specifically, **a)** individual post-familiarity RDMs for likeability were first correlated with **b)** averaged pre-familiarity trustworthiness RDMs, and the same was carried out for **c)** individual post-familiarity trustworthiness RDMs and **d)** averaged pre-familiarity likeability RDMs (see **Figure 17**). Correlations within-trait were also analysed for completeness. Mean correlations are presented in **Table 5 and Figure 18**. One-samples tests were carried out on these correlations.

Results showed significant correlations between **a)** and **d)**, and **c)** and **d).** Wilcoxon tests showed a significant difference between **c x d** and **c x b**, but there is a significant difference between **a x d** and **a x b** (see **Table 5**). Results indicate that the representational structure of likeability post-familiarisation does not significantly relate to that of trustworthiness pre-familiarisation in the way that it relates to its own representations pre-familiarisation. This may indicate that trait trustworthiness may become more like trait likeability post familiarity. In other words, a person may go on to trust someone they initially liked, possibly for the same reasons.

Figure 17

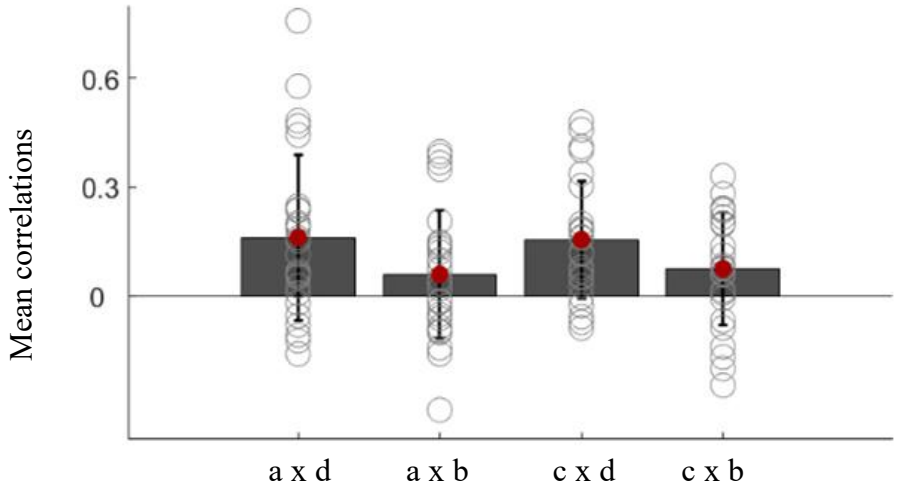
Analysis D: Additional analyses for likeability and trustworthiness: diagram depicting correlations between averaged pre-familiarisation RDMs with individual post-familiarisation RDMs in a criss-cross manner.



Note. A diagram depicting **individual post-familiarisation** RDMs (either **(a)** likeability or **(c)** trustworthiness) correlated with **averaged pre-familiarisation** RDMs (both **(d)** likeability and **(b)** trustworthiness).

Figure 18

Analysis D: Additional analyses for likeability and trustworthiness: correlations between likeability and trustworthiness RDMs (and correlations within the traits).



Note. Additional analyses show correlations between each individual post-familiarity likeability and trustworthiness RDMs with average pre-familiarity likeability and trustworthiness RDMs, both within traits and across traits. Correlations were computed between participants.

Table 5

Analysis D: Means, standard deviations, one-samples and paired samples Wilcoxon tests for correlations between likeability and trustworthiness RDMs.

Trait correlations		Wilcoxon signed-rank tests			
	Means (SD)	One-samples rank tests		Paired samples rank tests	
		z-values	*p*-values	*z*-values	*p*-values
a x d	.29 (.37)	3.14	.0017	-	
a x b	.12 (.29)	1.87	.062	2.99	.0028
c x d	.18 (.28)	2.83	.0046	-	
c x b	.04 (.31)	.50	.62	2.35	.0188

Note. correlations between each individual post-familiarity likeability and trustworthiness RDMs with average pre-familiarity likeability and trustworthiness RDMs, both within traits and across traits. Correlations were computed between participants.

3.4 Discussion

The current study investigated the nature of social trait representations of faces and how they might change as people become familiarised with them. We were particularly interested in how these changes occur for the *same* faces and had 4 aims for the present study. First, we were interested in whether there would be an overall increase in judgements of trustworthiness, dominance, and likeability (and not attractiveness); we hypothesised that with additional personal information during familiarisation, the participants would rate the identities higher on most traits overall. We then assessed the representational organisation of faces based on the traits. Thus, our second aim was to ascertain representations of unfamiliar face traits and measure

how these changed with familiarisation for individual participants. We also assessed this structure between participants to understand the similarity in these representations based on individual traits, and then how this similarity might have changed over the two weeks. While research shows that familiar face representations for famous faces are likely to be more idiosyncratic between people (Mileva et al., 2019; Kim et al., 2022), we expected the representations obtained in the present study to be similar both before and even more so after familiarisation. This is because participants would have gotten to know the same faces and would have learned the same personal and social information. Finally, we assessed the representational structure between the traits themselves (i.e., trustworthiness and attractiveness representations) for each participant. We hypothesised that associations between attractiveness and trustworthiness, and attractiveness and likeability, would remain stable across sessions (Mileva et al., 2019). We also hypothesised that trustworthiness and likeability representations would be associated pre-familiarisation (Oosterhof & Todorov, 2008). We had no other predictions about the other trait relationships.

Analysis A*: overall change in face-trait ratings*

First, we tested whether there would be an overall change in people's trait judgements after becoming familiarised with a set of faces. The results showed that only the fundamental dimensions (i.e., dominance and trustworthiness; Oosterhof & Todorov, 2008) ratings significantly increased overall. We hypothesised that, overall, people would rate the faces higher for likeability also, as additional personal information may have influenced their judgements overall in a positive direction, simply due to increased exposure (Zajonc, 1968; Lavan et al., 2021). Findings show that trustworthiness and dominance ratings in particular increase with familiarisation not only for the auditory modality (Lavan et al., 2021) but also for the face modality. This may suggest that any initial perception of these fundamental traits become further consolidated after just two weeks of familiarising with a person. This is in line with results from

Ambrus et al. (2021), showing that face representations acquired through para-social familiarisation may be related to increased neural activity in face-selective brain regions.

Also in line with our expectations, attractiveness ratings did not significantly change overall. The present findings show that with repeated visual exposure to face identities, despite exposure to semantic, emotional, personal and autobiographical information, attractiveness ratings remain stable. Peskin and Newell (2004) showed that with repeated exposure to faces, attractiveness ratings increased. In Peskin and Newell (2004), participants were repeatedly exposed to faces within a single session, whereas in the present study, participants were exposed to faces over a comparatively longer period of two to three weeks. According to the premise of the mere exposure effect by Zajonc (1968), the type of exposure in Peskin and Newell's (2004) may relate to increases in perceived attractiveness due to familiarity temporarily. This effect was likely characterised by developing a preference in the short term, or possibly even a diminished repulsion for faces they had been repeatedly exposed to in a relatively short space of time, typically in minutes or seconds. This effect is likely related to a reduction in initial negatively valanced curiosity associated with seeing a new object or person, where we may initially assume the object or person to be more dangerous due to uncertainty. In other words, the effect seen in Peskin and Newell's (2004) study may be due to some sort of adaptation or habituation (Bornstein, 1989), and a recognition that the stimulus is safe and not harmful. Findings from the current study may provide evidence for a more stable, averaged, consolidated perception of trait attractiveness. This stability may arise as a consequence of the potential tapering off of the mere-exposure effect very early on which would not have been detected in the present study, and which no longer influences attractiveness perception in the longer term.

Our findings also show that likeability ratings did not change overall, which goes against our expectations. It is possible that participants are either certain about their (dis)like for the identities, or that 2 weeks is not sufficient to develop or diminish their liking for the identities. Importantly, it could be the case that likeability requires more personal relations with the

identities (i.e., two-way interaction, or in-person interaction), which was not tested in the present study. Nevertheless, results show that 2 weeks of familiarisation may not lead to a change in the overall likeability of a face.

Analysis B: within-participant changes in the representational structure of face-traits

Our second analysis investigated changes in the representation of face traits with familiarity. We were interested in whether the pairwise association of faces for each trait (i.e., how similarly two face identities are perceived based on trustworthiness), would shift or be re-organised after becoming familiar with these faces. This would shed light on the way we may utilise information extracted from the face of one identity to perceive the face of another identity based on that trait, particularly post-familiarisation. Representational similarity analysis allowed for comparing this organisation of face traits across sessions (i.e., pre- and post-familiarisation) to explore whether there was a significant change. It was expected that trait RDMs would significantly change for all traits except attractiveness (Tsantani et al., 2021). Results showed that there was a significant change for trustworthiness and dominance but not likeability and attractiveness RDMs, reflecting the pattern of results from the first analysis. Previous research shows that people may rely on a different set of traits to make judgements about unfamiliar people than for those they are already familiar with (Kim et al., 2022; Mileva et al., 2019), and our findings add to this, demonstrating how these representations might *change* during a process and whether or not they do for specific traits. It appears that participants increasingly relied on novel social information that became accessible while watching the show, with changing social dynamics between people and a richer social context provided. This indicates that trustworthiness and dominance judgements may constantly be shaped by new information, as opposed to attractiveness and likeability, which may primarily rely on more stable visual features. This finding is also in line with the demonstrative model developed Kramer et al. (2018), suggesting that both bottom-up and top-down processes likely co-operate during familiar face perception as they utilise distinct types of person-related information.

Analysis C: between-participant agreement in the representational structure of face-traits

For our third analysis, we were interested in the similarity of the face-trait structure among the participants, and whether this agreement changed with familiarisation. We believe this analysis is novel and may provide insights into the essence of shared face perception more generally. Inter-judge agreement was analysed in a slightly different way to that of previous studies (Oosterhof and Todorov, 2008; Hönekopp, 2006). Instead of looking at an agreement in means ratings, we analysed the relationships in the *representational structure* of traits, which reveals more about the subtle nuances in face perception, specifically relating to the way faces are rated *in relation to each other*, rather than on average, for individual faces. Firstly, we expected that there would be some initial agreement among participants for all trait RDMs (but highest agreement for attractiveness, Oosterhof & Todorov, 2008) as participants had access to the same visual information. We also expected this agreement to significantly increase post-familiarisation for all traits except attractiveness (based on the same rationale as the previous analyses; Tsantani et al., 2021). We expected an increase in agreement due to the encoding of the *same* social and personal information across participants.

Results showed that representations of all face traits (except trustworthiness) were similar across participants before familiarisation. Findings for agreement pre-familiarisation, for traits attractiveness and dominance, are somewhat in line with results from Oosterhof and Todorov (2008) and Hönekopp (2006), but not the agreement for trustworthiness. After familiarisation, agreement remained stable for all traits except trustworthiness, where there was a highly significant increase, surpassing that of the other traits, including attractiveness. A few things may be deduced from these results: first, we appear to initially find the same people attractive and dominant, and likely rely on similar information to make these judgements, especially for those we begin to trust. This shows that there is some consistency in the way that we 'cluster' people and form impressions. This agreement does not seem to significantly change with familiarisation for most traits, even familiarisation with the same people, and with access to the same social

information. Due to the similarity in the representational structure across people, we seem to find the same people attractive, dominant and likeable. However, it could be interpreted that people initially rely on different features or information to that of the other traits in order to determine who is (un)trustworthy. It could be that people are relying on the same social dynamics and behaviours to judge trustworthiness, but only after becoming familiar with the same people. Might this reveal some kernel of truth in our judgements? It could be that people are accurately using behaviour and other social cues to make objective decisions about the identities' intentions, which might be a more accurate way to make judgements (John et al., 2010). Findings may also help explain the basis for thought processes relating to herding behaviour and other social situations where people might rely on similar information to determine trustworthiness, such as sexual faithfulness in potential partners (Rhodes et al., 2013). Moreover, the nature of the show required that candidates display traits of dominance (i.e., leadership and competence during team competitions) and even trustworthiness (i.e., being entrusted with task responsibilities, not 'throwing other candidates under the bus' when on the losing team) for instance, which may contribute to the variance found in the present results. Thus, naturally, nuances in the changes to the representational structure of traits may manifest differently when watching shows of another nature.

Analysis D: representational structure between traits and changes post-familiarisation

Our final analysis compared between representations of the traits themselves. We assessed the relationships between each pair of traits, both for first impressions and also for familiarised face judgements. The aim was to explore which traits may be used to make inferences about another trait (i.e., does attractiveness help us determine who to trust?). In other words, do trait representations consist of similar face information? Naturally, we also investigated whether these change over time. We expected similarities in attractiveness representations to be similar to that of likeability and trustworthiness (Oosterhof and Todorov, 2008; Mileva et al., 2019), which may decrease with familiarisation. We reasoned that as perceptions shift to rely on higher-

level information with increased familiarisation (Schwartz & Yovel, 2016; Kramer et al., 2018), the potential Halo Effect (HE) (Thorndike, 1920) would diminish. It was also expected that likeability and trustworthiness representations would initially be similar (Oosterhof & Todorov, 2008), but we had no expectations for whether and how this would change after 2 weeks. There were no expectations for the other trait relationships, although Oosterhof and Todorov (2008) show moderate correlations between ratings of likeability and dominance, attractiveness and dominance, and, to a lesser extent, trustworthiness and dominance. Thus, it is possible that the previous findings of univariate ratings may also extend to the multivariate representational structure of these same traits. It may also indicate that the HE based on positive representations including trustworthiness and likeability may increase with familiarisation; people may make other positive inferences about someone they have gotten to know because they have come to trust and/or like them, and in the context of the present study, the HE may strengthen. Importantly, however, the HE may be strong *because* people do not have rely on other explicit pieces of information about an unfamiliar person. In the context of knowing or becoming familiar with someone, if one has reason to trust or like, or assign positive qualities to another, this could be due to the actual acquisition of relevant information rather than mere social trait associations at the visual level. Thus, it is debatable as to whether the HE increases with familiarisation, and conversations about the effect may require that we specify characterisations of it (i.e., is the Halo Effect restricted to unfamiliar face perception? Is it inherently unfounded?).

Results showed that associations of various trait pairs were significantly associated with each other pre-familiarisation, including attractiveness and trustworthiness (Mileva et al., 2019), but only the association between that of likeability and trustworthiness significantly increased with familiarisation. Findings overall highlight the presence of shared mechanisms between the traits, indicating that people likely rely on information relating to one trait to make judgements based on another trait. It could also be the case that social roles matter: will we find similar results with familiarisation as people become romantic partners? Liking and trusting a friend or a celebrity

would involve distinct representations to each other as well as that of someone of a romantic interest. The present study is not able to address this possibility and is, thus, a limitation.

We further explored the directionality between trustworthiness and likeability and found that trustworthiness representations post-familiarisation become similar to likeability representations pre-familiarisation. Likeability representations, however, remained stable. It could be said that trustworthiness practically *becomes* likeability itself while likeability is a trait in its own right, and the information we associate with those we learn to trust becomes more reliant on that which makes someone likeable in the beginning. It should be noted, however, that the traits used in the present study may still very much be linked to those beyond the ones tested in the current study.

Limitations and future directions

Due to the nature of the present work, results may primarily be generalised only to the context of para-social familiarisation, although we may be able to extrapolate them to the wider population. Essentially, we should consider the possibility that results may vary depending on the type of stimuli and information available to the participants, some of which were absent from the present study. For example, it is likely that in-person familiarisation would have a significant impact on changes to the representational structure of traits (Kovács, 2020). This might be mediated by the presence of conceptual (Schwartz & Yovel, 2016), emotional and episodic information (Ambrus et al., 2021), as well as two-way interaction rather than only one. Ambrus, et al. (2021) found increased cortical activity associated with in-person familiarity compared to para-social familiarisation (and minimal perceptual familiarisation). Future studies could explore changes to the representations of face traits with an in-person paradigm.

The time frame for familiarisation is also likely to influence the results: a longer time period may have allowed for the further development of particular trait representations. There is no solid evidence or theory to suggest that this would be the case, but it is possible. While the time frame is longer than that of some previous studies, and while robust visual familiarity representations

may have started to take place (Popova and Wiese, 2023a), sufficient familiarisation peaks somewhere between 2 months and 14 months of familiarisation (Popova & Wiese, 2022), possibly by the 8-month mark (Popova & Wiese, 2023b). However, this was only tested for in-person familiarisation (Popova & Wiese, 2022). In any case, this is a stage where neither increasing the frequency, recency, nor closeness of in-person contact, nor gaining more personal, semantic information, would significantly increase familiarity representations (Popova & Wiese, 2022). Consequently, two weeks may have been insufficient for capturing further development in trait representations (Burton et al., 2005). Rather, it could be said that these 'post-familiarity' representations are somewhere lower on a familiarity continuum, rather than at its maximum or optimum (Karimi-Rouzbahani et al., 2021; Li et al., 2022). These representations may be some relating to a type of familiarity that goes beyond mere perceptual exposure and first impressions, but not representations that reflect familiarisation in its fullness.

Future studies could also further dissect the variability in the structure of traits. Our results show that while there was no significant change in some of the representational patterns between participants, for dominance, there was a trend towards idiosyncrasy, highlighting the developmental trajectory towards more individualised perceptions. This trait in particular is related to many other traits which might implicitly be used as a proxy for dominance (e.g., competence, aggression, meanness; Oosterhof & Todorov, 2008), all of which may be linked to different types of social information. Future studies could directly explore what exactly comes to mind when asked to rate a face on trait dominance. This will allow for controlling the variability in the judgements that may be related to differences in inferring its definition or trait associations.

3.5 Conclusion

To conclude, the current study presents an assessment of the changes in the representational structure of 4 face-traits, including attractiveness, dominance, likeability and trustworthiness. A gap in the social trait perception literature remained around the potential change during the

process of familiarisation, which by extension, would be marked by differences in the perception of the *same* faces. The present study aimed to address this in multiple ways. First, we explored overall change in ratings and found that there was only an increase for the fundamental dimensions (trustworthiness and dominance; Oosterhof & Todorov, 2008). This may indicate an overall shift in the perception of the face identities due to repeated exposure. Then, we assessed initial face trait representations and its subsequent change post-familiarisation. Changes were present within participants, revealing differences in face-trait structures again for the fundamental dimensions only (Oosterhof & Todorov, 2008). Results also showed that the initial inter-rater *agreement* of most traits remained stable except for trustworthiness, where similarities for this trait structure were initially negligible, but then increased substantially post-familiarisation. This was a particularly novel finding, providing new insights indicating that people may increasingly rely on the same social information to make decisions about who to trust, in only two weeks of familiarisation. This information seems to go beyond just the visual and physical, involving the integration of top-down semantic and person-related information (Bruce & Young, 1986; Schwartz & Yovel, 2016; Kramer et al., 2018). Findings are in line with that of Kim et al. (2022), namely that we may appear to rely on distinct information to make judgements about unfamiliar and familiar people, and the present results demonstrate how these might change in the interim, and for which traits. We also learn that some of the trait structures are related to each other, and thus, similar social information might be used to make judgements about another trait (Mileva et al., 2019). This was particularly the case for inter-correlations between attractiveness and likeability, attractiveness and trustworthiness, and especially likeability and trustworthiness. After further analysis, results showed that trustworthiness as a trait may '*transform*' into likeability, or that the trust we develop for someone turns into liking. Future research could extend the familiarisation time for a minimum of 2 months (Popova & Wiese, 2022, 2023b) and could employ in-person paradigms (Kovács, 2020; Ambrus et al.,

2021), as well as use an extended set of social traits (Ambrus et al., 2021; Popova & Wiese, 2022). Studies could also further explore the interim variability in the trait structures.

Chapter 4: Investigating how brain representations of faces change with familiarisation

Abstract

Studies have explored the involvement of face-selective regions in familiar face processing over the past 20 years. Much of the research has revolved around understanding the response properties of regions of interest (ROIs) that respond more to faces than other visual stimuli, but less is known about how face-related representations *change* with familiarisation. The current study sought to investigate the representations of faces in three core face-selective regions (the FFA, OFA and pSTS), and how these might change with familiarisation of the *same* face identities. Participants watched a television series over the course of 2-3 weeks in order to become familiarised with a set of face identities. First, we investigated whether the face representations in the brain might change for each participant individually, then assessed the agreement across participants in these representations, both before and after face familiarisation. We assessed whether the agreement in these representations did or did not change with familiarisation, either by maintaining stability or becoming more or less idiosyncratic. Finally, we assessed the similarity in representations *between* the ROIs themselves for all participants (i.e., between representations in the OFA and FFA), and then whether the similarities changed with familiarisation. 29 Participants watched The Apprentice UK over 2 weeks, and underwent functional Magnetic Resonance Imaging (fMRI) before and after watching the show to obtain face-related brain activity patterns. Results showed that (1) for individual participants, face representations for each ROI substantially changed with familiarisation, (2) representations of faces were significantly more aligned across participants in the OFA and pSTS separately, pre-familiarisation. However, there was only a significant change in agreement after familiarisation for the OFA. Finally, the similarity in the representational structure of faces was significant between the FFA and OFA, both pre- and post-familiarisation, but this was not the case for other ROI pairs, and there was only a trend towards significance between the OFA and pSTS post-familiarisation. Results reveal insights into the way the organisation or 'grouping' of faces change within the core visual system within just 2-3 weeks of para-social familiarisation. Further

research could explore the content of these representations and the directionality of these relationships, shedding light on the respective stages of processing and specified functions for each ROI during the process of face familiarisation. The present study may inform and challenge previous models of familiar face perception (Gobbini & Haxby, 2007; Duchaine & Yovel, 2015). Future research is warranted to attempt to assign some of the explicit face information to these brain representations for both unfamiliar and familiar faces, and how these might change during the process of familiarisation.

4.1 Introduction

Face-selective regions and their computations

The ability to recognise a familiar face is important for social interactions. Our brains seem to have developed a complex network of face-selective brain regions that reliably respond more to faces than to other visual stimuli, including the occipital face area (OFA; Gautier et al., 2000), the fusiform face area (FFA; Kanwisher et al., 1997), and the posterior superior temporal sulcus (pSTS; Engell & Haxby, 2007). The OFA is generally implicated in processing more physical face properties including the T-shaped arrangement of internal face features (i.e., eyes, nose and mouth), face shape, and luminance (Gauthier et al., 2000; Tsantani et al., 2021). Analyses exploring the representational similarity of neural patterns in OFA when presented with faces were also significantly associated with patterns of behavioural ratings based on attractiveness and gender, indicating the region is also likely involved in encoding visual properties that might be related to these characteristics in a face (Tsantani et al., 2021). Regarding the FFA, studies suggest that it is involved in both lower- and higher-level, perceptual information related to processing face identity (Verosky et al., 2013; Tsantani et al., 2021). The pSTS is most likely implicated in image-based as well as abstract social information extracted from faces including intention, particularly when the perceived intention of target is not matched with goal-directed outcomes of that same target (Shultz et al., 2011). The nature of an interaction between entities (human or non-human; Shultz & McCarthy, 2012) may be facilitated by the co-ordination of processing averted eye-gaze and facial expressions (Campbell et al., 1990; Engell & Haxby, 2007) in the pSTS, suggesting this region is largely involved in the interpretation of social interactions more generally (Isik et al., 2017), however, there is some evidence to suggest that threat detection is a prominent function of the pSTS (Candidi et al., 2011).

Models of familiar face perception

It could be expected that the core face-selective regions would be implicated in the perception of familiar faces in a distinct manner from that of unfamiliar faces. Models of familiar face perception detail the potential connectivity between relevant face-selective regions in two visual pathways: the dorsal pathway, primarily dedicated to processing information from dynamic face stimuli, and the ventral pathway, responsible for decoding visual-based information (Haxby et al., 2000; Duchaine & Yovel, 2015). The dorsal and ventral are thus two separate but interacting systems, each containing regions constituting what is called a core system (comprised of the OFA, FFA and pSTS) and an extended system (i.e., the anterior temporal lobe (ATL) intraparietal sulcus (IPS) and amygdala), each system responsible for decoding visual and non-visual information respectively (Haxby et al., 2000). It has more recently been proposed that there is an additional third visual pathway, containing the pSTS (Pitcher & Ungerleider, 2021). For the purpose of this thesis, however, I am focusing on the three regions the core system as defined by Haxby and colleagues (2000), and therefore will focus on the OFA, FFA, and pSTS.

Previous research has rarely distinguished between the processing between unfamiliar and familiar faces, both of which should not be conflated in face perception research (Ramon & Gobbini, 2018). Gobbini and Haxby (2007) made revisions to their previous model (Haxby et al., 2000) to account for both unfamiliar and familiar face processing, with a slight addition stating that the decoding of semantic information related to familiar faces occurs in the ATL in the extended system, which is connected to and receives input from the FFA (for a review, see Duchaine & Yovel, 2015). It is likely that the co-ordination of these brain regions may facilitate the process of accurate recall of these pre-stored person-related representations so that we can determine how to interact with others accordingly. As a result, brain representations pertaining to higher-level person-related information may become idiosyncratic, making it more difficult to ascertain general, shared representations associated with familiarity. Additionally, Kovács (2020) proposed a model outlining that the three core face regions are implicated in both unfamiliar and familiar face processing (in line with previous findings, i.e., Rossion et al., 2001),

but that there are no significant *identity* representations associated with these ROIs for familiar faces. This indicates that these regions might significantly change the way in which they store information about faces of the people we know.

It has been roughly two decades since Gobbini & Haxby (2007) proposed their framework for familiar face perception, yet questions remain as to the differences between unfamiliar and familiar face perception. Recent empirical evidence, however, provides insight into the possible functions of face-selective regions and their potential connectivity in facilitating familiar face perception. Focusing empirical research on the role of face-selective areas will allow for obtaining the more visual-based representations that are most likely shared across people (Natu & O'Toole, 2015).

Research carried out by Natu and O'Toole (2015) investigated changes in neural activity in face-selective FFA and OFA as a function of increased familiarisation. Initially, for the behavioural segment of the study, participants viewed unfamiliar faces, of which they were subsequently familiarised to specified levels of low familiarity (characterised by 1 exposure to 10 faces, each from 8 viewpoints), medium familiarity (characterised by the same, except with 2 exposures), and high familiarity (with 3 exposures). Then, participants were scanned using fMRI while viewing whole body images of the familiarised faces. Activity in voxels pertaining to localised FFA and OFA, which were analysed together, each with a 3mm radius, were measured across all conditions. Results showed that for activity relating to the FFA-OFA, there was a significant increase in the pattern and magnitude of activity between the unfamiliar and highly familiar face conditions, but no differences with medium familiarity. There was, however, a non-significant trend showing an increment in correlation between FFA-OFA and condition type. Namely, activity in these regions of interest (ROIs) increased with familiarisation, implying that increased exposure to faces likely also strengthens visual-based familiarity representations, though more subtle distinctions in the type of familiarisation would likely need more time or longer exposure.

A similar aim was investigated by Bayer et al. (2021). Neural activity related to familiar faces, as well as their timings, were directly sourced from individual core face-selective regions within the occipito-temporal cortex. The authors used electroencephalography (EEG) in tandem with fMRI to ascertain timings of activity in individual ROIs related to unfamiliar and familiar face-related activity. In the MRI scanner, participants were presented with images of unfamiliar faces, faces of their close friends, and faces of their romantic partners. Whole-brain activity patterns were measured during the passive viewing of face stimuli. Similarities in the representational dissimilarity matrices (RDMs) for the timings of event-related neural activity were compared to the RDMs of the spatial locations eliciting this event-related activity, using representational similarity analysis (RSA). Stimulus-related activity in each voxel (i.e., specified areas of 3-dimensional brain space) were obtained for each participant using fMRI, as well as searchlight analysis, where significant event-related neural activity across the whole brain were correlated with EEG RDMs (of brain activity timings) in order to identify the areas that elicited event-related activity at the specified times. RDMs were computed based on pairwise correlations between each voxel, for each participant. RDMs of the timings recorded using EEG were computed for all amplitudes across participants and electrodes, where each pairwise entry in the RDM depicted the Euclidean distance between the faces of all categories (i.e., unfamiliar and familiar including close friends and romantic partners). EEG and fMRI RDMs were subsequently correlated to assess the similarity in the representational structure between the timing of activity (measured with EEG) and brain spatial representations of faces (measured with fMRI). Results showed that there were significant correlations between fMRI RDMs and EEG RDMs at multiple successive time points, suggesting sustained activity for faces of both close friends and romantic partners in relevant regions. Notably, correlations were found for actual or nearby core face-selective ROIs, including the fusiform gyrus (FG), superior temporal sulcus (STS) and middle temporal gyrus (MTG), beginning just before 204ms and lasting up to approximately 800ms. Face images of close friends and romantic partners significantly elicit

peak activity 108ms, 204ms, 428ms and 660ms (removing variance related to romantic love from analysis by only retaining the amplitudes common to both groups). It has been interpreted that the remaining variance relates to familiarity representations, implying that core visual areas related to face processing play a direct role in processing familiar faces. The activity in these regions appear to be more pronounced (in line with Natu & O'Toole, 2015), suggesting that person-related representations in these regions may become increasingly pronounced, particularly in the FFA (Duchaine & Yovel, 2015), which may require top-down input from regions implicated in higher-level person identity information relating to familiar people (Dobs et al., 2019; Fan et al., 2023).

In a similar study, Ambrus et al. (2021) compared the neural amplitude in the broader occipito-temporal region between unfamiliar and familiar faces, with a key focus on differing types of familiarity as well as duration and intensity of familiarisation. Participants' brain activity was recorded using EEG while they were presented with unfamiliar faces, and subsequently familiarised with sets of faces by either visual familiarisation (induced by mere visual repetition), para-social familiarisation (i.e., watching a TV series over a two-week period), or in-person familiarisation for 1 hour per day (over 3 days in total). The authors then computed RDMs consisting of pairwise decoding accuracies extracted from brain activity responses to faces, using linear discriminant analysis classifiers. These RDMs were averaged across participants for each of the 6 clusters. RDMs were then compared between perceptual familiarity, para-social familiarity, and personal familiarity condition. Results from EEG recordings showed that, when presented with the respective familiarised faces, there was significant deactivation of neural activity in the broader occipito-temporal regions across participants, occurring around 250ms after face presentation onset for participants who were either para-socially familiarised or familiarised with people in-person. This timepoint is known as the negative 250 potential (i.e., N250), the point at which the brain begins to significantly distinguish between unfamiliar and familiar faces (Tanaka et al., 2006; Dobs et al., 2019). Additionally, those who were familiarised

with people in-person showed even stronger deactivation in comparison to those who were familiarised para-socially, suggesting that increasing the intensity of familiarisation also increases the amplitude in neural activity in the broader posterior area of the brain, containing the three core face-selective ROIs. Findings reveal that the dynamic development of familiarity representations can emerge in as quickly as 3 hours, as observed in ROIs close to or within the face-selective regions in the core visual network.

To further ascertain the variance in neural activity patterns associated with visual familiarity versus that which is associated with higher-level personal information, Visconti di Castello et al. (2017, 2021) compared activity patterns associated with each separately, thus, quantifying their relative contributions to familiar face processing. First, participant brain responses to unfamiliar faces were recorded using fMRI, as well as responses to visually familiar faces (acquired through visual repetition of these unfamiliar faces), and personally familiar faces, faces of whom the participants had become friends with over an average of 3.2 years. Participants then watched a movie whilst they underwent fMRI scanning to acquire shared brain activity patterns. The authors employed searchlight analysis to identify event-related activity across the brain, which was obtained while participants watched the same movie. Canonical covariance analysis (CCA) was then used to assess the similarity in the neural response patterns across familiarity conditions, weighting the relative contributions of patterns associated with, e.g., person knowledge obtained while watching the movie as well as patterns elicited by, e.g., personally-familiar faces. Results showed that activity associated with decoding identity for visually familiar faces was significantly associated with activity patterns in core visual regions including bilateral OFA and FFA, while patterns associated with identity for personally familiar faces related to activity in both core visual and extended regions, including bilateral pSTS in the core network and right temporo-parietal junction (TPJ), medial prefrontal cortex (mPFC) among others. The findings indicate that the visual aspects of face familiarisation are mostly decoded in core visual areas, while person knowledge and other relevant non-visual information are mostly

decoded in the extended network, thus supporting previous models of face perception (Hoffman et al., 2000; Gobbini & Haxby, 2007). These results are in line with previous research (Ambrus et al., 2021; Bayer et al., 2021) and add to our knowledge of the potential functions of these core regions in familiar face perception.

Together, these findings show that core face-selective regions are likely directly involved in familiar face processing as studies consistently show that these regions elicit increased neural activity. This may mark the potential consolidation of person-related information about an increasingly acquainted person. It is currently understood that ROIs including the FFA and OFA are implicated in visual-based functions of face processing, for both unfamiliar and familiar faces. However, mixed findings remain as to their direct involvement in familiar face processing. Bayer et al. (2021) showed differences in neural activity in face ROIs between groups of already familiarised individuals, while Ambrus et al. (2021) showed significant changes for increasingly familiarised identities in terms of overall amplitude. Natu and O'Toole (2015) demonstrated that combined activity in FFA and OFA change with increased visual exposure to faces without accounting for variance associated with the ROIs individually, and Visconti di Castello et al. (2017; 2021) demonstrated that activity patterns change in relevant regions over short-term familiarisation of only 50 mins. There remains a research gap regarding changes to the structure of representations of the same faces and over a longer time period, long enough to ascertain person-related information whilst maintaining much of the visual influence of familiar face perception, as this aspect of familiar face processing is more likely to be shared across individuals (Natu & O'Toole, 2015).

The Present Study

The present research aims to address a gap in the literature relating to potential change in the representational structure of brain responses to faces. Specifically, we aim to extend research by measuring the structure or organisation of unfamiliar face representations and compare this with

that of the same faces after familiarisation. We specifically want to explore brain activity in response to faces in each face-selective ROI and how these might change with familiarisation. Face familiarisation was restricted to para-social familiarisation (Kovács, 2020) where participants watched a TV show and underwent brain scans both prior to and after watching the show. We reasoned that participants would learn relevant social information about the people in the show and that the neural responses that mark this change might be captured as we measure brain activity in response to the newly familiarised faces. Focusing on changes in the visual ROIs only might naturally allow for measuring variance that might be more common across participants (Natu & O'Toole, 2015), reducing the variance associated with more idiosyncratic face identity perception based on higher-level informational content such as emotional, personal, and other non-visual information. Specifically, we are interested in the visual-based representations of faces, and how these might change with familiarity.

To assess the brain representations of faces in face-responsive FFA, OFA and pSTS, the present study will use representational similarity analysis (RSA; Kriegeskorte et al., 2008a). First, brain activity patterns in response to faces were recorded using fMRI, while participants viewed videos of the face identities and had to detect the faces and voices of famous people. Participants underwent fMRI scanning both pre- and post-familiarisation with the faces. In between scanning sessions, participants watched all 12 episodes of a television series called the Apprentice UK (series 11). The brain activity patterns in response to each identity were then compared (using Euclidean distance) to compute dissimilarity matrices (RDMs). RSA was then used to compare the RDMs between pre- and post-familiarisation sessions. We aimed to investigate how the representational structure of faces changed in three different ways:

1. Assess whether there is a significant change in the representational structure of face-identities (face-identity RDMs) between pre-familiarity and post-familiarity phases. Like all analyses, we focused on face core ROIs (FFA, OFA and pSTS). We expect the structure of RDMs to change less across familiarisation sessions for the OFA and, thus,

for OFA RDMs to correlate highest. OFA is often implicated in low-level image processing (Gobbini & Haxby, 2007; Gauthier et al., 2000), and we thus expect that RDMs in this ROI would remain more stable as face features would be identical across sessions. As for FFA and pSTS, implicated in both visual and abstract processing (Campbell et al., 1990; Engell & Haxby, 2007; Gobbini & Haxby, 2007; Candidi et al., 2011; Tsantani et al., 2021; Yan et al., 2023), we expect that the RDMs would significantly change post-familiarisation. These changes could be due to potential top-down feedback that may enhance representations in the FFA (Kramer et al., 2018; Dobs et al., 2019; Fan et al., 2023). Similarly, the pSTS may be implicated in both unfamiliar and familiar face processing in distinct ways (Gobbini & Haxby, 2007; Natu & O'Toole, 2011). Thus, RDMs for both of these regions are not expected to be significantly correlated across sessions.

2. Assess whether there is a significant change in inter-subject agreement of face-identity representations (face-identity RDMs) between pre-familiarity and post-familiarity phases. We will investigate how similar the face-identity RDMs are across participants, both pre- and post-familiarisation. We believe this particular analysis is novel and would shed light on whether learning the same information about faces affects the representations of those faces. We expect that pre-familiarised OFA RDMs will be significantly correlated between participants, whilst FFA and pSTS RDMs will also be less similar across participants. Correlations for post-familiarised OFA RDMs will not significantly change, as we expect that the nature of representations in this region is not affected by the learning. Conversely, we predict that FFA and pSTS RDMs will become significantly more correlated between participants as participants will have access to the same semantic information, which may be reflected in the similarity of face representations in these ROIs.

3. Assess whether there is a change in the representational structure of RDMs of different brain regions between pre- and post-familiarity sessions. We will assess the correlations between face-identity RDMs for each pair of ROIs, to assess whether they are representing the same information. We will then assess whether the correlations between pairs of ROIs change with familiarity. We only have expectations for representations to be similar between the OFA and FFA RDMs for unfamiliar faces, and thus, be significantly correlated pre-familiarisation (Gauthier et al., 2000; Tsantani et al., 2021). We expect them to represent similar lower-level image-based properties, which is a primary source of information before familiarisation (Gobbini & Haxby, 2007). FFA and pSTS RDMs are then expected to become more significantly correlated post-familiarisation due to their shared implication in visual and non-visual content including face features (Campbell et al., 1990; Engell & Haxby, 2007; Tsantani et al., 2021), perceived dominance (Tsantani et al., 2021) and perceived threat (in pSTS; Candidi et al., 2011), which may contain similar social information (Oosterhof & Todorov, 2008). Similarities in RDMs between core face-responsive regions may suggest shared cognitive mechanisms and, by extension, shared content. In other words, ROIs with similar brain representations may be implicated in processing similar information.

4.2 Methods

4.2.1 Participants

Thirty-one participants were recruited from Royal Holloway, University of London and from Brunel University London. Participants were required to be between 18 and 30 years of age and to have lived in the UK for at least 10 years. This ensured that that participants were fluent in English. All data was collected by Nadine Lavan and Olga Korolkova.

Participants' age ranged between 18 to 27 (*male* = 5, *female* = 26; *mean age* = 21.35, *SD* = 2.68; *age range* = 18-27). Data for 2 of the participants were removed due to incomplete data. Thus, data for the final sample consisted of 29 participants (*male* = 4, *female* = 25; *mean age* = 21.10, *SD* = 2.53; *age range* = 18-27).

Ethics was approved by the Research Committee at Brunel University London (see **Appendix A**). All participants provided written consent before participating (see **Appendix B**).

4.2.2 Design

The present study has a repeated-measures correlational design. Brain RDMs for each ROI were correlated across pre-familiarisation and post-familiarisation sessions.

Participants took part in four testing sessions: two pre-familiarisation and two post-familiarisation. There was an MRI session involving structural and functional MRI pre-familiarisation and another post-familiarisation. There was one behavioural testing pre-familiarisation and another post-familiarisation — part of the results of the behavioural testing will be reported in Chapter 3. As part of the familiarisation, participants watched the British television series 'The Apprentice UK' (Series 11), which is fully explained in Procedure. For the fMRI segment, participants were scanned whilst viewing and listening to the face and voice stimuli of 8 identities from this TV series. The sections below describe the stimuli and procedure for the fMRI session.

4.2.3 Materials

Face stimuli. The stimuli are the same as those presented in Chapter 3. Three individual face videos (i.e., 3 tokens) of each face identity were taken of 8 people from the British series 'The Apprentice', which is about people who complete tasks in teams with the aim to become the next apprentice for multi-millionaire Lord Alan Sugar. This resulted in 24 face videos in total. These identities were the 8 final candidates of the season, which increased the chances of participants

becoming more familiarised with the faces as they remained on the show for the longest. The names of the face identities were as follows: Brett Butler-Smythe, Charleine Wain, Joseph Valente, Gary Poulton, Richard Woods, Scott Saunders, Vana Koutsomitis, and Selina Waterman-Smith. Video segments were taken such that the background did not reveal any information about the identity. Videos were also silent, non-speaking, and contained slight movements of the head, primarily with a front-facing view. **Figure 19** shows images from one of the videos used for identity (give name here). Each video lasted for 1.2 seconds.

Figure 19

Sample video stimulus from the fMRI experiment.

Redacted

Note. Example of 1.2 second video stimulus presentation during scanning in both pre- and post-familiarity conditions.

Voice stimuli. Voice recordings of all 24 identities were also played for 1.2 seconds each during scanning in both sessions.

4.2.4 Procedure

First session MRI session: pre-familiarisation

Participants underwent structural and functional magnetic resonance imaging (fMRI).

Main fMRI experimental runs. Participants were presented with intermixed face and voice stimuli using Psychophysics Toolbox (version 3; Brainard, 1997; Pelli, 1997) for MATLAB. In each run, participants viewed videos of the eight identities (all 24 videos) and listened to voice recordings of all identities (24 voice stimuli). Each stimulus (48 stimuli in total) was repeated four times (192 experimental trials). Participants performed a task in which they had to detect a face or a voice from a famous person by pressing a button. There were 12 task trials, 6 with famous faces and 6 with famous voices. In each trial, the stimulus was presented for 1.2 seconds, followed fixation for a total trial duration of 3.5 seconds. Finally, there were 41 null events in which the fixation cross was presented for duration of the trial. All trials were presented in a randomised order in each run. There were three runs in total.

Participants also received a structural scan and did functional localisers (see section *4.2.7.2 Functional localisers to estimate ROIs*).

Interim session: watching the TV series

Participants were provided with a USB stick containing instructions for the study, links to all 12 episodes of the show, and questionnaires about the identities from the show (see **Appendix C**). Participants then watched the whole of The Apprentice UK series at home over 2 weeks. This was our familiarity task as we reasoned that participants would become familiar with the characters in the show during this time, particularly the final 8 identities who would have naturally been on screen longer than those who left the show earlier (for context, one person leaves the show after each episode).

To ensure that participants were watching the show, they had to complete trivia questions after each episode. These questions were presented using Qualtrics (www.qualtrics.com). Example of multiple-choice questions include “Which market did the contestants go to buy most of their ingredients?”, to which participants would select the correct answer. Participants also provided

interim ratings for each of the 12 episodes for how familiar they were with each identity (i.e., "How familiar are you with… Gary?") and for trait questions for 3 traits, including trustworthiness, dominance and likeability (i.e., "How dominant is… Richard?"). Participants rated the identities on familiarity and social traits using a 7-point scale. Results for this task showing the intermediate effects of familiarisation are reported in Chapter 3.

Second MRI session: post-familiarisation

After approximately two weeks, participants were invited back into the lab for the second MRI session. The mean interval between the two sessions was 20.76 days (*SD* = 1.81). Participants repeated the main fMRI experimental runs, again with three runs. The order of trials was randomised for each run. A structural scan and the functional localiser runs were also repeated.

Participants were then debriefed and compensated for their time at the end of the final behavioural rating session.

4.2.5 MRI Data Acquisition

Participants underwent (f)MRI brain scans at the Combined Universities Brain Imaging Centre (CUBIC) at Royal Holloway, University of London. Participants' brains were scanned using a 3.0-Tesla Tim Trio MRI scanner (Siemens, Erlangen), with a 32-channel head coil. In each of two scanning sessions (pre- and post-familiarisation), whole-brain T1-weighted anatomical scans were acquired using magnetisation-prepared rapid acquisition gradient echo (MPRAGE) [1.0 x 1.0 in-plane resolution; slice thickness, 1.0 mm; 176 axial interleaved slices; PAT, Factor 2; PAT mode, GRAPPA (Generalised Autocalibrating Partially Parallel Acquisitions); repetition time (TR), 1900 ms; echo time (TE), 3.03 ms; flip angle, 11°; matrix, 256 x 256; field of view (FOV), 256 mm].

In all functional runs, T2-weighted whole-brain functional scans were acquired using echo-planar imaging (EPI) [3.0 x 3.0 in-plane resolution; slice thickness, 3.0 mm; PAT, Factor 2; PAT

mode, GRAPPA (Generalised Autocalibrating Partially Parallel Acquisitions); 34 sequential (descending) slices; repetition time (TR), 2000 ms; echo time (TE), 30 ms; flip angle, 78; matrix, 64 x 64; field of view (FOV), 192mm]. Slices covered all parts of the brain for most participants, except for the most dorsal portion of the parietal cortex. In each experimental run, 263 brain volumes were obtained, and in each run of the face, voice, and person localiser runs we obtained 227 brain volumes.

4.2.6 fMRI Data Pre-processing

Statistical Parametric Mapping (SPM12) was used to pre-process the obtained brain data (Wellcome Department of Imaging Science, London, UK; RRID:SCR_007037; http://www.fil.ion.ucl.ac.uk/spm) operating in MatLab (version R2013b; MathWorks; RRID:SCR_001622). Each scanning session was pre-processed separately, and all runs within each session (i.e., the main experiment or the localiser runs) were pre-processed together. To allow for T1-equilibration effects, the first three EPI images in each run (i.e., dummy scans) were removed. Images of brains were slice-time corrected based on the middle slice in each volume, and were then realigned in order to correct for head movement based on the first image. The structural image in native space was then co-registered with the mean realigned functional image, and was segmented into grey matter, white matter, and cerebrospinal fluid. Smoothing was not performed on the images obtained from the experimental runs. From the localiser runs, however, functional images were smoothed with a 4 mm Gaussian kernel (full width at half maximum). After carrying out pre-processing of the images in each session, the structural image from the second session were realigned (co-registered) to the structural image from the first session. Transformation was then applied to all functional images from the second session. Thus, all functional images obtained from each participant were in the same space. Mass univariate time-series models were computed for each participant and from both the main experimental runs and the functional localiser runs. The blood-oxygen level-dependent (BOLD) response

post-stimulus onset were modelled using regressors and were convolved with a canonical haemodynamic response function (HRF). A high-pass filter was also used with a cut-off of 128 s, and an autoregressive AR (1) model was used to account for serial correlations. Six head motion parameters that were computed during the realignment stage were included as regressors of no interest.

4.2.7 Data analysis

4.2.7.1 Main experimental runs: first-level data analysis

For each scanning session and experimental run, models were defined separately. The 8 different face identities in each modality were entered as separate regressors in the model (i.e., 8 regressors per modality). Each regressor included the three different stimuli for each identity, as well as the repetitions for each stimulus. Task trials were included as regressors of no interest. As part of the crossvalidation procedure used in the main analyses described below, separate models were estimated for each partition of each crossvalidation fold, thus resulting in parameter estimates and residual time courses for every possible independent partition. For partitions with two runs, data was concatenated before estimating the model. In the analyses described below, we used the beta estimates computed at each voxel of each ROI for each of the 16 experimental conditions (8 face-identities and 8 voice-identities). Here, we are only reporting the results for the beta estimates of the face conditions. The parameter estimates for each face-identity were extracted for each of the core face-selective ROIs.

4.2.7.2 Functional localisers to estimate ROIs

The three core face-selective regions were defined using a dynamic face localizer, which presented famous and non-famous faces, along with a control condition consisting of objects and scenes. Stimuli were silent videos of non-speaking faces which were moving, and silent videos

of objects and scenes, presented in an event-related design. Participants completed between one and two runs of the localizer across the two scanning sessions. The localizer presented different stimuli in each of two runs.

Functional ROIs were defined using the Group-Constrained Subject-Specific method (Fedorenko et al., 2010; Julian et al., 2012), which has the advantage of being reproducible and reducing experimenter bias by providing an objective means of defining ROI boundaries. Subject-specific ROIs were defined by intersecting localizer contrast images specific to each participant with group-level masks, for each ROI obtained from an independent dataset. In the present study, group masks of face-selective regions (right FFA [rFFA], right OFA [rOFA], and right pSTS [rpSTS]) were obtained from a separate group of participants who completed the same localizer (for details, see Tsantani et al., 2019). We focused on face-selective regions from the right hemisphere because they have been shown to be more consistent and larger compared to the left hemisphere (e.g., Rossion et al., 2012; Rossion, 2014; Bukowski et al., 2013). The masks are publicly available at https://doi.org/10.17633/rd.brunel.6429200.v1.

Contrast images were defined for each participant. Face selectivity was defined by contrasting brain activation related to faces to that of non-face stimuli using *t* tests. We then intersected these subject-specific contrasts with the group masks and extracted all significantly activated voxels at $p < 0.001$ (uncorrected) that fell within the boundaries of each mask. In cases where the resulting ROI included fewer than 30 voxels, the threshold was lowered to $p < 0.01$ or $p < 0.05$. ROIs that included fewer than 30 voxels at the lowest threshold were not included, and this occurred for the rFFA in 2 participants and for the rOFA in 1 participant. For full details of size and location of all ROIs, see Tsantani et al. (2019).

4.2.7.3 Representational Similarity Analysis: second-level analysis

Representational similarity analysis (RSA; Kriegeskorte et al., 2008a) was used to compare the structure of brain representations of faces before familiarisation with the structure of brain representations after familiarisation. Representational dissimilarity matrices (RDMs) were created using brain activation patterns (vectors of beta estimates) for each face identity from each face-selective ROI (i.e., the FFA, OFA and pSTS). Pairwise distances between the brain activation patterns for each pair of identities were computed into an 8 x 8 RDM. The dissimilarity measure was the linear discriminant contrast (LDC; Walther et al., 2016; Tsantani et al., 2021), in which zero means the patterns cannot be distinguished and values above zero mean that they can be distinguished. We computed an RDM for each participant, ROI, and session (i.e., pre- and post-familiarisation). All analysis was computed using MATLAB, version R2021b. We then used RSA to compare these RDMs in three different analyses.

4.2.7.4 Bayesian Analyses

Bayesian analyses, using Bayesian Information Criteria (Wagenmakers, 2007) were conducted to assess the probability of whether ROI RDM correlations across sessions were more likely to differ 1. from zero (H_1; i.e., correlations are higher) as opposed to being dissimilar (H_0; i.e., correlations are lower) given the data, and 2. the probability of ROI RDM correlations significantly differing from each other, within and between sessions (H_1; i.e., correlations are higher) as opposed to not differing from each other (H_0; i.e., correlations are lower).

BF10 shows the probability that the alternative hypothesis is more likely than the null with either weak/anecdotal evidence (characterised with values between 1-3), moderate evidence (between 3-10), strong evidence (between 10-30), very strong evidence (between 30-100) and very strong evidence (> 100) according to Jeffrey's (1961) scale. BF01 represents the statistical inversion of BF10 (i.e., 1 / BF10) which denotes the value for the null hypothesis. Values for the posterior of the median are measured according to Cohen’s (1992) conventions which are considered either small (~.2), medium (~.5), or large (~.8).

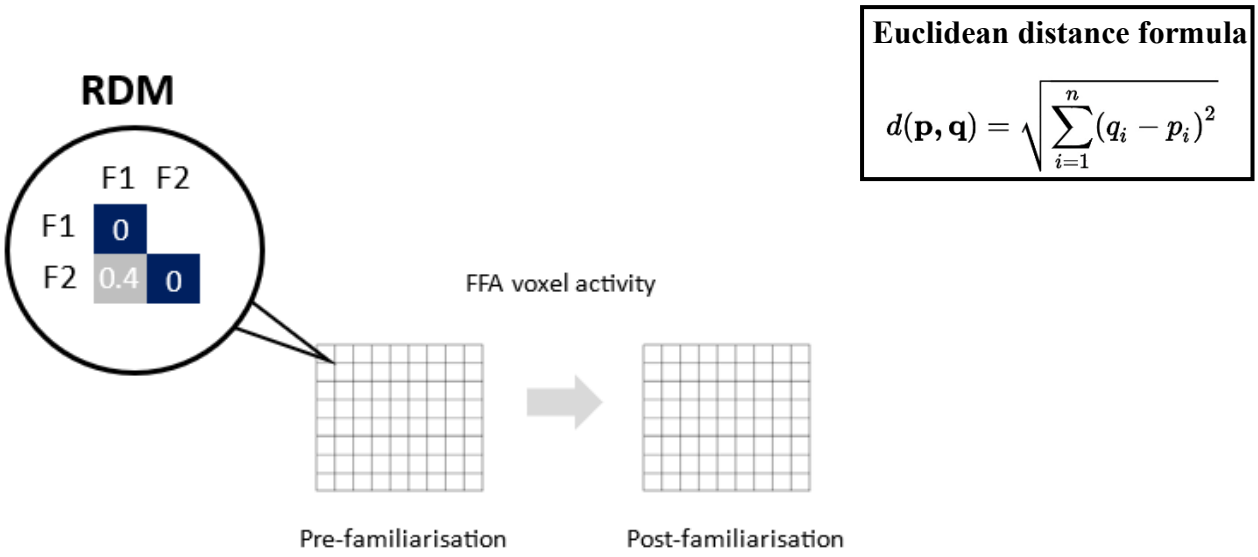
Bayesian analyses were conducted using JASP, version 0.95.1.

4.3 Results

4.3.1 Analysis A: The first analysis investigated whether there was a significant change in the representational structure of face-identities (face-identity RDMs) between pre-familiarity and post-familiarity phases (see **Figure 20**). RSA was used to assess the similarity in the representational structure of ROI RDMs pre- and post-familiarity. Face-identity RDMs were computed for each participant, ROI, and session. RDMs were then correlated (Pearson) across sessions for each participant, to evaluate the similarity of representations across sessions. **Figure 21** shows the correlations of RDMs for each participant, and also the mean across participants for each ROI. All means are close to zero. RDMs for Participant 1 are shown in **Figure 22** for visualisation. One-samples rank tests showed that correlations between pre- and post-familiarity RDMs were not significantly above zero for any of the ROIs, namely, the FFA (M = -.03, SD = .25, p = 0.45; *Bayesian*: *BF10* = .277, $\hat{r}$ = 1.001, *median* = -.130, 95% CI [-.532, .263]), OFA (M = .02, SD = .26, p = 0.76; *Bayesian*: *BF10* = .230, $\hat{r}$ = 1.000, *median* = .062, 95% CI [-.312, .459]), and pSTS (M = .01, SD = .21, p = 0.61; *Bayesian*: BF10 = .247, $\hat{r}$ = 1.000, *median* = .064, 95% CI [-.340, .464]; all data had excellent integration) (see **Table 6**). These findings suggest substantial changes in the face representations of each ROI as participants became familiar with a set of face identities over a two-week period.

Figure 20

Example diagram depicting the comparison between ROI RDMs pre- and post-familiarisation.



Note. **Analysis A**: 8 x 8 RDMs were computed consisting of pairwise distances between the faces based on voxel activity patterns (VAPs) for each ROI and session, using Linear Discriminant Contrast (LDC). RDMs were compared between pre- and post-familiarisation sessions within each participant and for each ROI.

Table 6

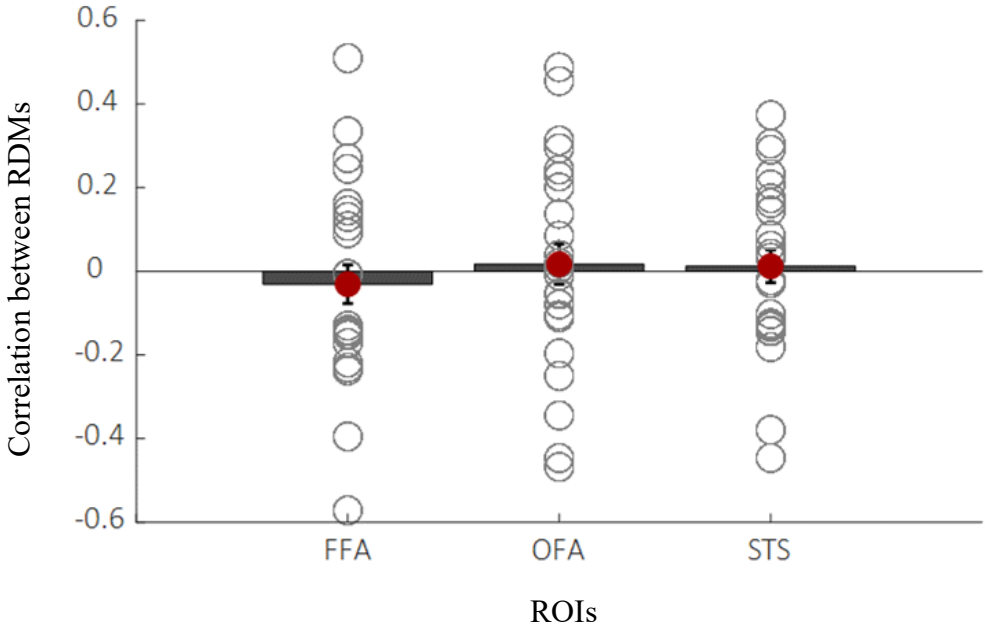
Analysis A: One-samples rank test: z-scores for RDM correlations within participants for each ROI across sessions, compared to zero.

ROI	Means (SD)	*z*-values	*p*-values
FFA	-.03 (.25)	-0.76	.45
OFA	.02 (.26)	0.30	.76
pSTS	.01 (.61)	0.52	.61

Note. RDMs for FFA, OFA and pSTS were correlated across sessions within participants and then compared to zero using one-samples rank tests to assess whether there was a significant correlation for each of the ROIs.

Figure 21

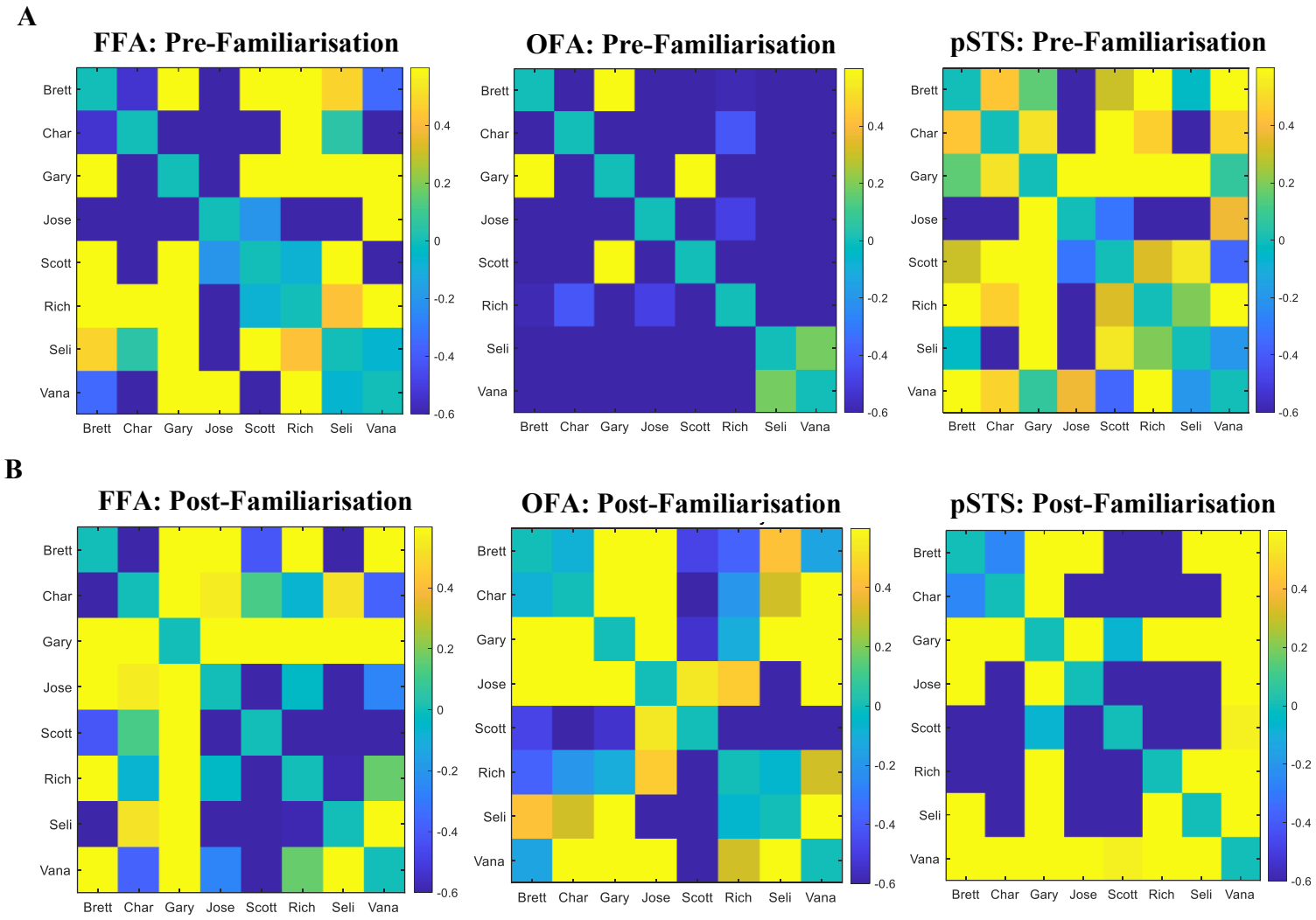
Analysis A: within-subjects: correlations between RDMs pre- and post-familiarisation.



Note. Red dots show mean correlations across participants for the FFA, OFA and pSTS. Grey circles show correlations for individual participants. Error bars show standard errors of the mean.

Figure 22

Analysis A: RDMs depicting the distance between faces for the FFA, OFA and pSTS, both pre- and post-familiarisation, for Participant 1.



Note. From left to right, images of 8 x 8 RDMs for the FFA, OFA and pSTS, both **A.** pre-familiarisation and **B.** post-familiarisation for Participant 1. Each cell of the RDMs depict the distance between the brain activity patterns for each pair of identities. Distances were computed using Linear Discriminant Contrast (LDC).

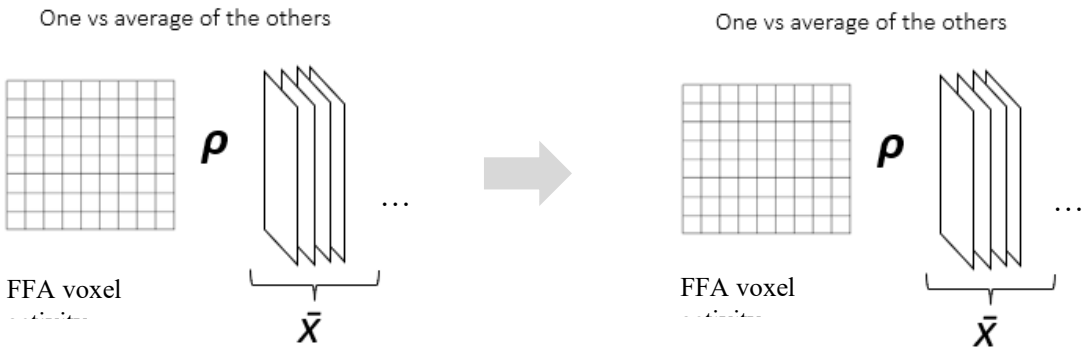
4.3.2 Analysis B: The second analysis investigated whether there is a significant change in inter-subject agreement of face-identity representations (face-identity RDMs) between pre-familiarity and post-familiarity phases (see **Figure 23**). We first investigated the similarity of face-identity RDMs between participants, both prior to and after familiarisation. Individual pre-familiarisation RDMs were correlated (Pearson) with the mean of all other participants' RDMs to assess the reliability of representations across participants. The same was carried out for post-familiarisation RDMs. Inter-subject correlations between RDMs were significantly above zero (tested with Wilcoxon sign-rank tests and Bayesian analyses) for pre-familiarisation OFA RDMs (M = .12, SD = .29, p = .048; *Bayesian: BF10* = 1.952, $\hat{r}$ = 1.008, *median* = .424, 95% CI [.013, .828]) and pSTS RDMs (M = .13, SD = .26, p = .03; *Bayesian: BF10* = 1.657, $\hat{r}$ = 1.021, *median* = .402, 95% CI [-.021, .853]) while all other inter-subject correlations were not significant (all p > .25) (see **Table 7**). Results reveal that the face brain representations were similar across participants in the OFA and pSTS pre-familiarity, but not the FFA. Results show that post-familiarisation, there are no inter-subject correlations above zero for any of the ROI RDMs, suggesting that representations in all ROIs are more idiosyncratic post-familiarity, and is more likely to be the case according to Bayesian analyses (*FFA: BF10* = .446, $\hat{r}$ = 1.005, *median* = .240, 95% CI [-.158, .639]; *OFA*: *BF10* = .375, $\hat{r}$ = 1.002, *median* = -.204, 95% CI [-.600, .178]; *pSTS*: *BF10* = .299, $\hat{r}$ = 1.000, *median* = .151, 95% CI [-.245, .544]), with integration of data ranging from excellent to moderately good.

Wilcoxon sign-rank tests were used to assess whether inter-subject correlations changed across sessions. Results showed that there was only a significant decrease in the inter-subject agreement between participants for the OFA RDM (z = 2.04, p = .042; *Bayesian*: *BF10* = 2.242, $\hat{r}$ = 1.004, *median* = .438, 95% CI [.037, .862]), suggesting that representations became more idiosyncratic in the OFA post-familiarisation. There was no evidence of a significant change for FFA and pSTS (all p > .32) (see **Table 7** and **Figure 24**). Given that mean correlations were low in both

sessions, results indicate high idiosyncrasy in both sessions, which did not significantly change within the two-week period.

Figure 23

Example diagram depicting ROI RDM correlations between subjects, for each session, then comparing pre- and post-familiarisation.



Pre-familiarisation

Post-familiarisation

Note. **Analysis B**: individual participants' RDMs were correlated with the average of all other RDMs. This was carried out for all 3 ROIs and both sessions. RDM correlations were then compared across sessions.

Table 7

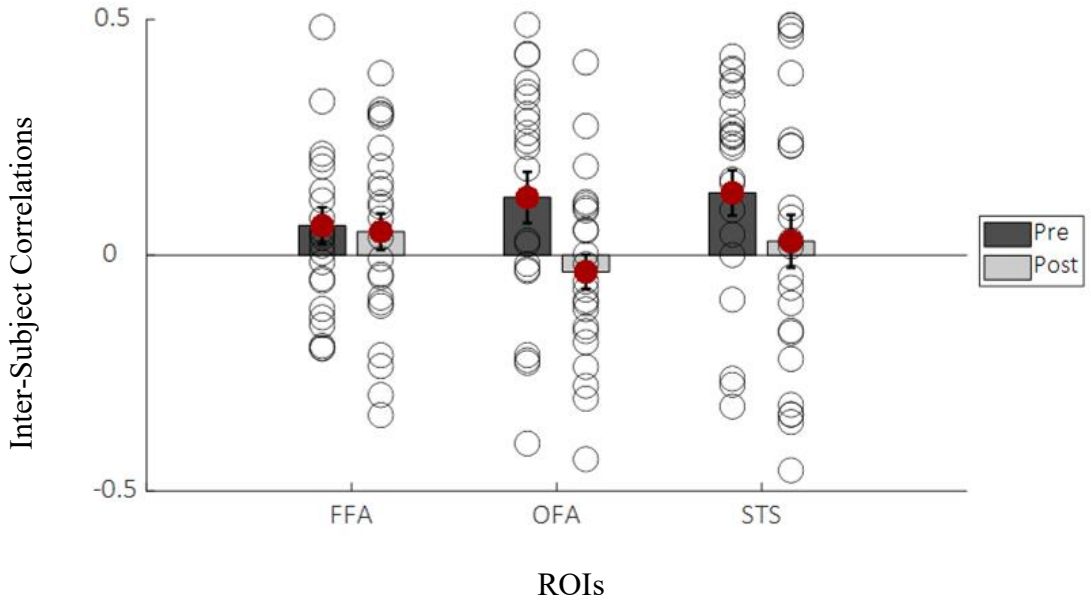
Analysis B: Means and SDs (and medians and ranges) for inter-subject correlations of face-identity RDMs, separately for pre- and post-familiarisation.

ROI	**Pre-Familiarisation**			**Post-Familiarisation**		
	Mean (SD)	*z*-values	*p*-values	Mean (SD)	*z*-values	*p*-values
FFA	0.06 (0.21)	1.06	.29	0.05 (0.21)	1.16	.25
OFA	0.12 (0.29)	1.98	.048	-0.04 (0.19)	-1.00	.32
pSTS	0.13 (0.26)	2.13	.03	0.03 (0.30)	0.49	.63

Note. One-Samples Rank tests were used to compare these correlations to zero and the table shows z-values and p-values. Wilcoxon Signed-Rank tests were also used to compare inter-subject correlations from pre- to post-familiarisation.

Figure 24

Analysis B: Inter-subject Correlations for each ROI both Pre- and Post-Familiarisation.



Note. Red dots indicate mean inter-subject correlations across participants for the FFA, OFA and pSTS both pre- and post-familiarisation. Grey circles indicate individual participant correlations with the average of all the other participants. Error bars show standard errors of the mean.

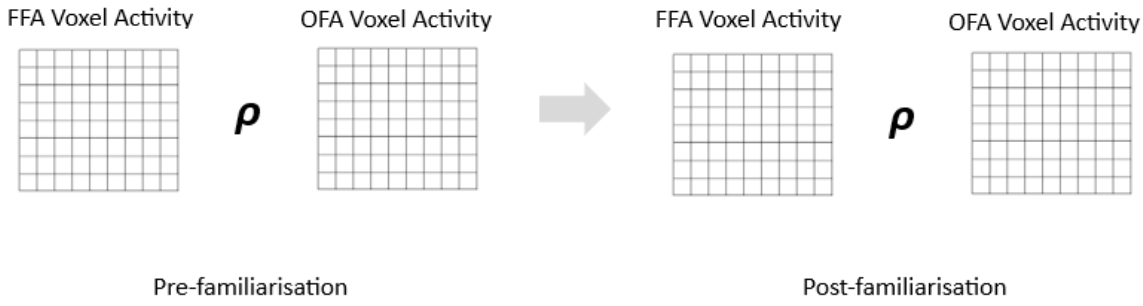
4.3.3 Analysis C: The third analysis investigated whether there is a change in the representational structure of RDMs (face-identity RDMs) of different brain regions between pre- and post-familiarity sessions (see **Figure 25**). In other words, we assessed the correlations (Pearson) of the representations between the three core face-selective RDMs, separately for pre- and post-familiarisation. For each participant, RDMs for each ROI were correlated with each other ROI (i.e., OFA RDMs were correlated with FFA RDMs, then with pSTS RDMs, and so on). This was carried out separately for pre- and post-familiarised brain RDMs. **Table 8, Figure 26** and **Figure 27** show the mean correlations for each pair of ROIs. One-samples rank tests were then used to assess whether each set of correlations was significantly different from zero. There were correlations significantly higher than zero between OFA and FFA RDMs (M = .24, SD = .22, p = .0002; *Bayesian*: *BF10* = 2,072.96, $\hat{r}$ = 1.187, *median* = 1.226, 95% CI [.640, 1.874]) pre-familiarisation and post-familiarisation (M = .16, SD = .21, p = .004; *Bayesian*: *BF10* = 34.12, $\hat{r}$ = 1.006, *median* = .771, 95% CI [.282, 1.243]; ranging from questionable to very good integration). The correlations between OFA and pSTS post-familiarisation only trended towards significance (M = .14, SD = .30; z = 1.86, p = .06; *Bayesian*: *BF10* = .228, $\hat{r}$ = 1.000, *median* = .040, 95% CI [-.354, .430]). All other pairwise correlations between ROIs were also not significant (all p > .11) (see **Table 8**, **Figure 26** and **Figure 27**). These results suggest that representations of faces in the OFA and FFA were similar to each other, both pre- and post-familiarisation.

Wilcoxon signed-rank tests were used to statistically examine whether these inter-ROI correlations changed with familiarisation, and results show there was no significant change for

any of the correlations, but again, only a trend towards a change in the RDM relation between the OFA and pSTS (z = -1.86, p = .06) (see **Table 8**).

Figure 25

Example diagram depicting the similarity between ROI RDMs, separately for pre- and post-familiarisation sessions.



Note. **Analysis C:** individual RDMs for each ROI were correlated with each of the other ROIs, for each participant. Inter-ROI correlations were then compared across sessions to assess whether the similarities in the organisation of ROI activity changed.

Table 8

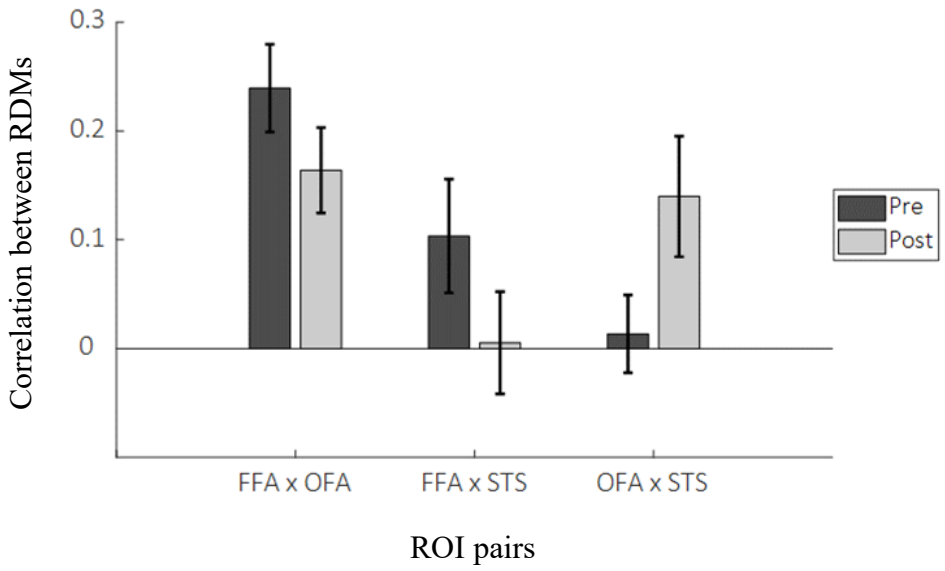
Analysis C: Means and SDs for correlations of face-identity RDM across ROIs, separately for pre- and post-familiarisation.

ROI					Wilcoxon signed-rank tests			
					One-samples rank tests		Paired-samples rank test	
	Pre		Post		Pre	Post		
	Mean (SD)	*p*-values	Mean (SD)	*p*-values	*z*-values	*z*-values	*z*-values	*p*-values
FFA x OFA	.24 (.22)	.0002	.16 (.21)	.004	3.68	2.89	1.40	.16
FFA x pSTS	.10 (.28)	.11	.01 (.25)	.88	1.58	.15	1.22	.22
OFA x pSTS	.01 (.19)	.95	.14 (.30)	.06	-.06	1.86	-1.86	.06

Note. One-Samples and paired-samples rank tests were used to compare correlations for each session to zero and then across sessions.

Figure 26

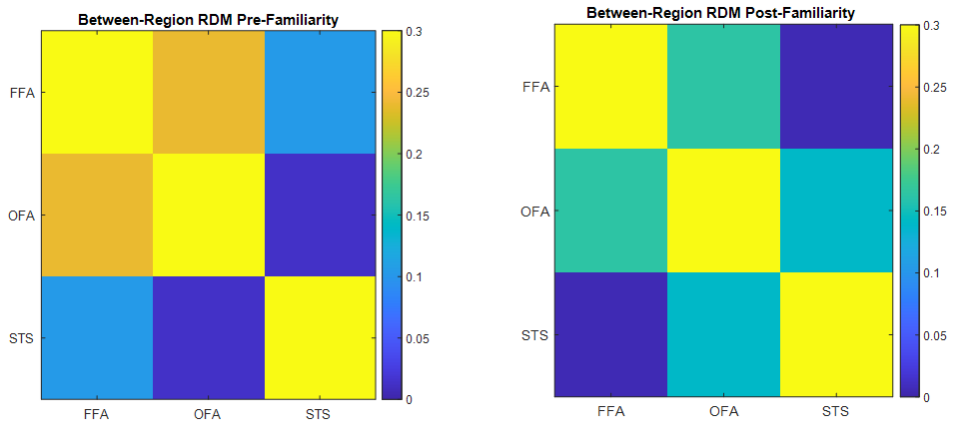
Analysis C: Inter-ROI correlations both pre- and post-familiarisation.



Note. Bars represent means of pairwise correlations between FFA and OFA RDMs, FFA and pSTS RDMs, and OFA and pSTS RDMs. Error bars show the standard error of the mean.

Figure 27

Analysis C: Mean RDMs depicting pairwise correlations between the FFA, OFA and pSTS, both Pre- and Post-Familiarisation (across participants).



Note. Inter-trait RDMs depicting average pairwise correlations between ROI RDMs across participants.

4.4 Discussion

The present study investigated the brain representations of faces of people we do not know in core visual face-selective regions (i.e., the FFA, OFA and pSTS), and whether these representations changed after having become familiarised with them. Participants became familiar with a set of faces over a two-week period by watching The Apprentice UK for about two weeks. We used fMRI to investigate their brain representations of the faces of people in the show and recorded their brain activity response patterns to each face both before and after watching the show. We then built face-identity RDMs for each core face-selective region and assessed how these representations changed from pre- to post-familiarisation. For the first analysis (Analysis A), we investigated whether representations changed within each participant

from pre- to post-familiarity and found that correlations between pre-familiarisation RDMs with post-familiarisation RDMs were low for all ROIs, suggesting that representations may have changed substantially at the individual level. For the second analysis (Analysis B), we investigated the consistency of the brain representations across participants, both prior to and after getting to know the face identities from the show. We were interested in whether and how any initial similarity would change with familiarisation. We found that representations of faces in the OFA and pSTS (but not FFA) were similar across participants pre-familiarity, but not post-familiarity, suggesting changes in participants brain representations that became more idiosyncratic after familiarisation. Lastly, for Analysis C, the relationships between representations in each pair of ROIs was assessed, both before and after familiarisation. We found that the FFA and OFA represented similar information about the faces, and this did not change with familiarisation. We now focus on each of these analyses in more detail.

Analysis A: within-participant changes in the representational structure of ROIs

The first analysis aimed to shed light on the potential way that the organisation of face-identity representations in the FFA, OFA and pSTS may change with increased familiarisation. We predicted that the representations would change across sessions for all ROIs, but less so for the OFA, given that the physical properties of the target faces would be identical across sessions. Results showed that correlations between pre- and post-familiarisation RDMs for all ROIs were not significantly above zero, suggesting substantial changes across all ROIs. This is in line with expectations, except for our predictions about the OFA. One possible interpretation could be that the 'grouping' of faces in all three ROIs occur according to a particular characteristic or feature, which may change as person-related representations become more refined. In other words, these ROIs may begin to group these faces based on different features with increased familiarisation.

These findings are in line with previous research, suggesting that neural activity in the occipitotemporal cortex changes with familiarisation (Natu & O'Toole, 2015; Ambrus et al.,

2021; Visconti di Oleggio et al., 2017, 2021) and could be explained by the presence of more person-related information now available to the observer. This information may enhance representations in lower-level face regions via feedback processes from higher-level regions (Dobs et al., 2019; Ambrus et al., 2021). This is also in line with findings suggesting that the FFA and pSTS, in particular, may be implicated in multiple processes, both visual and non-visual (Kanwisher et al., 1997; Deen et al., 2015; Tsantani et al., 2021), and where the pSTS may particularly play a dual role in both core and extended face processing functions (Gobbini & Haxby, 2007).

While findings provide evidence for the implication of core visual ROIs in familiar face processing, it is not yet clear from the present results as to the specified characteristics or features by which these particular faces are grouped. From previous research, we can deduce that the OFA would initially play a more general role in processing these faces simply as 'faces' (as can be seen from the less varied pattern in the pre-familiarity OFA RDM), which appears to change post-familiarisation (where the pattern of the same post-familiarisation appears to be more varied, see **Figure 23**) (Gauthier et al., 2000, Tsantani et al., 2021). It can also be deduced that the FFA and pSTS would distinguish these faces based on both higher and lower-level information (Kanwisher et al., 1997; Gobbini & Haxby, 2007; Tsantani et al., 2021). Nevertheless, further research is required to specify the content of these groupings to further inform previous models of familiar face perception (Gobbini & Haxby, 2007).

Analysis B: between-participant changes in the representational structure of ROIs

Our second analysis examined the consistency in the brain representations elicited by faces in core regions across participants. As far as we know, this analysis is novel and could shed light on the extent of similarity of brain representations across individuals. It would also shed light on whether and how this similarity may change as participants become acquainted with the same identities. It was expected that the brain representations in the core ROIs would be correlated

across participants pre-familiarisation, mostly for the OFA, but that these would become even more similar after familiarisation, particularly for the FFA and pSTS, while similarities in the OFA would not significantly change. Again, our predictions were not entirely supported. For the RDMs in OFA and pSTS, pre-familiarity RDMs were significantly positively correlated across individuals, showing similar representations across participants. But this was no longer the case for post-familiarity RDMs, suggesting a trend towards idiosyncrasy rather than increased similarity of brain representations.

These findings do not support our hypotheses. As for the OFA specifically, it was expected that inter-subject correlations would be highest among participants pre-familiarisation, but that this would change least post-familiarisation. Inter-subject correlations of pre-familiarisation RDMs were indeed highest, yet had become the lowest post-familiarisation, indicating complete dissimilarity in brain representations of faces among the participants. These findings suggest that the representations of faces in the OFA became highly idiosyncratic. These changes could reflect the enrichment and refinement of face-related representations as a potential result of input of person-related information from more dorsal brain areas (Duchaine & Yovel, 2015; Dobs et al., 2019). This is in line with a recent study by Eick et al. (2020), who found the OFA to play a causal role in processing semantic information associated with faces, as opposed to many previous studies who found the OFA to consistently play a role in image-based processing (Gauthier et al., 2000; Kietzmann et al., 2015; Tsantani et al., 2021; for a review, see Pitcher et al., 2011). This is also in line with a study involving a patient with lesions to the rOFA, who was unable to recognise familiar faces (Rossion et al., 2003), also pointing towards possible feedback processes involved in familiar face perception. In either case, OFA representations seem to become more enhanced with familiarisation (Natu & O'Toole, 2015; Ambrus et al., 2021; Bayer et al., 2021, Visconti di Oleggio et al., 2017, 2021, for a review, see Pitcher & Ungerleider, 2021), and in different ways across participants.

Analysis C: changes in the representational structure between ROIs

The final analysis explored the similarity of face representations across the core face ROIs, both prior to and after para-familiarisation. We had two specific hypotheses about these correlations: first, that the representations in OFA and FFA would be significantly correlated pre-familiarisation, but that this correlation would significantly decrease post-familiarisation, while correlations between the FFA and pSTS increase post-familiarisation. Results supported our first hypothesis, showing a significant positive correlation between the brain representations of faces in the OFA and FFA pre-familiarisation, which remained significant after two weeks. However, FFA and pSTS associations did not become more positively associated, and thus there was no support for our second prediction.

These findings are in line with previous models of face perception and how face-selective regions are connected to each other. We think that regions that could be connected directly could have more similar representations, though these connections could change with familiarity. Current models pose that the OFA and FFA are connected via feedforward processes in both unfamiliar (Haxby et al., 2000) and familiar (Gobbini & Haxby, 2007) face processing via the ventral pathway. Additionally, Duchaine and Yovel (20215) suggested that through one of many pathways, the OFA and pSTS are directly connected, potentially via back propagation or recurrent processing of socially-relevant information. Our findings could have implications for these face perception models as they reveal the nature of the information potentially stored in these regions. Our results suggest that there are shared representations between the OFA and the FFA, which may indicate connectivity between these ROIs where the OFA feeds visual information into the FFA (Haxby et al., 2000; Gauthier et al., 2000). The findings also indicate that familiarity representations in the OFA and pSTS may also be shared and, thus, may indicate connectivity (Duchaine & Yovel, 2015). Particularly relating to the OFA and pSTS, shared representations may relate to changes in viewpoint and/or dynamic processing, as the present study used videos of face identities, which naturally showed moving faces and changes in eye

gaze. Previous studies and models suggest these functions are implicated in either or both of these regions (Haxby & Gobbini, 2000; Gobbini & Haxby, 2007; Kietzmann et al., 2015; Deen et al., 2015; Tsantani et al., 2021).

Regarding familiar face processing, as mentioned, the present findings suggest potential connectivity between the OFA and pSTS, where the pSTS feeds pre-stored information into the OFA via feedback or recurrent processing. Importantly, however, when assessing the *changes* in these inter-regional relationships, there appeared to be no significant change, but only a trend towards significance for the OFA and pSTS. Given that results for inter-regional representational similarity within each session (i.e., individually for pre-familiarisation and post-familiarisation RDMs) are significant, the associations are still meaningful. It can be deduced that there is likely to be significant changes in the connectivity between regions over time, but that it is not detectable within two weeks, which may still be a very early stage of familiarisation (considering a study that suggests 2-14 months induces significant familiarity representations; Popova & Wiese, 2022). Non-significant findings here may also reflect the requirement for a more intense type of familiarisation (i.e., in-person familiarisation; Ambrus et al., 2021; Bayer et al., 2021). Nevertheless, the present results are in line with the proposal of a multiple visual pathways by Duchaine & Yovel, (2015), which the authors proposed is implicated in familiar face processing. In short, our findings may provide evidence for regional connectivity (Kriegeskorte et al., 2008) within the core visual system and, thus, may inform previous face perception models in both unfamiliar and familiar face perception.

Limitations and future research

In line with a previous point, while a two-week period is longer than much familiarity research, detectable changes in the neural structure of core face ROIs may occur with a longer timeframe, such as 2 months (Popova & Wiese, 2022). In order to examine changes based on an *optimal* time period, it may be warranted to measure familiarisation over 8 months (Popova & Wiese,

2023b). Popova and Wiese (2023b) investigated both visual and non-visual based familiarity representations in the brain using EEG and found that both types of familiarisation were optimal around this period. Future research could replicate the present study with consideration for the optimal time period to assess the robustness of visual-related neural reorganisation in these face-selective regions.

Additionally, the present results are restricted to familiarity representations attained solely through para-social audio-visual means rather than representations obtained with in-person interaction, which appear to be more enriched (Ambrus et al., 2021; Kovács, 2020). More personal forms of familiarisation should be investigated directly to examine whether patterns of results are comparable to that of the present study. Would the representational reorganisation of faces occur in a similar way to that of the present study when accounting for representations of people developed through in-person meetings? And would we find a similar pattern of results for different types of familiar individuals? (i.e., *becoming* friends or romantic partners?) (Bayer et al., 2021).

4.5 Conclusion

The present empirical chapter investigated multiple ways in which face representations in core face-selective regions might change with familiarisation with a set of face identities. We were interested in whether the brain reorganises the way it stores information extracted from faces after learning about them through watching a TV show over a two- to three-week period. First, we assessed changes within-participants to see if representations remained stable, or if they indeed changed, and for each ROI (i.e., the right FFA, OFA and pSTS). Results showed that the representations of the faces had changed substantially within a two-week period, indicating that the brain likely changes the way it stores information within this short period of time. While we cannot make strong conclusions about the content of this information, it can be deduced that it is likely related to social, semantic and person-related information, as this was present and

increasing throughout the show. We were also interested in how similar these representations were between participants, both prior to and after getting to know the same people. Results showed that representations in the OFA and pSTS were more similar across participants for unfamiliar faces, but became less similar with familiarisation, though only to a significant extent for the OFA. Results indicate that the way these ROIs organise information that is extracted from the face significantly change with familiarisation, particularly in the OFA. The OFA most likely becomes involved in distinguishing between faces based on different information for different people. Similar conclusions may be drawn for the pSTS, though there is only a trend suggesting this possibility. Finally, we were interested in the similarity between the ways the ROIs themselves organise face information (i.e., do the FFA and OFA share similar face representations?). Findings showed that FFA and OFA representations were significantly associated both prior to and after familiarisation, indicating the two ROIs appear to represent similar face information and that this does not change with increased person knowledge. It could be the case that the FFA and OFA remain implicated in processing the physical properties of the faces to a similar extent, and that this does not change even after having acquired higher-level perceptual information about others. While there were no other significant correlations, there was a non-significant trend towards an association between the OFA and pSTS post-familiarisation, indicating the potential for increased similarity in the informational content of the representations in these regions. It may indicate that two weeks is not sufficient for the OFA and pSTS to become implicated in the same face-related information, however, as results were non-significant, these conclusions are not concrete. The present research potentially informs and challenges models of face perception models (Gobbini & Haxby, 2007; Duchaine & Yovel, 2015; Kovács, 2020). Future research is warranted to attempt to map some of the explicit face information to the brain representations of unfamiliar and familiar faces, and how these might change in the process.

Chapter 5: Dyadic Person Similarity Predicts Similarity in Face Judgements

Peer Reviewed Contribution

Chapter 5 has been published in the Quarterly Journal of Experimental Psychology.

Williams, R., & Garrido, L. (2026). Dyadic Person Similarity Predicts Similarity in Face Judgements. *Quarterly Journal of Experimental Psychology*.
https://doi.org/10.1177/17470218261420

Dyadic Person Similarity Predicts Similarity in Face Judgements

Rochelle Williams[1], Lúcia Garrido*[1]

[1] Department of Psychology, City St George's, University of London

Abstract:

Perceivers consistently extract information from faces to judge whether others are attractive, trustworthy, or dominant. However, there is also substantial variability among perceivers when making these face-judgements. Here, we investigated whether dyadic similarities in participants' personalities are related to similarities in their face-judgements. 307 participants based in the UK rated 24 faces on six traits. Participants also rated themselves on social-traits and completed a personality questionnaire. We computed dissimilarities between pairs of participants for face-judgements, self-perceived-traits, and personality-traits, resulting in three separate dissimilarity matrices. Using representational similarity analysis, we showed that both the self-perceived-trait and personality matrices were significantly correlated with the face-judgement matrix. Importantly, these associations were stable when controlling for age, gender, ethnicity, and location. These findings show that people who are more similar to each other also perceive others in a similar manner, and could form the basis for how we gravitate towards others and build friendships.

Keywords: Face perception; Social perception; Intersubject similarity

Data availability statement:
The hypotheses, methods, and plan for main analyses were pre-registered before starting data collection: https://doi.org/10.17605/OSF.IO/ZR92G.
Primary data and scripts for main analyses are publicly available: https://doi.org/10.25383/city.26519890.

Acknowledgements:
This work was supported by a PhD studentship from City, University of London to RW and a Leverhulme Trust Research Fellowship to LG (RF-2024-326).

Abstract

Perceivers consistently extract information from faces to judge whether others are attractive, trustworthy, or dominant. However, there is also substantial variability among perceivers when making these face-judgements. Here, we investigated whether dyadic similarities in participants' personalities are related to similarities in their face-judgements. 307 participants based in the UK rated 24 faces on six traits. Participants also rated themselves on social-traits and completed a personality questionnaire. We computed dissimilarities between pairs of participants for face-judgements, self-perceived-traits, and personality-traits, resulting in three separate dissimilarity matrices. Using representational similarity analysis, we showed that both the self-perceived-trait and personality matrices were significantly correlated with the face-judgement matrix. Importantly, these associations were stable when controlling for age, gender, ethnicity, and location. These findings show that people who are more similar to each other also perceive others in a similar manner, and could form the basis for how we gravitate towards others and build friendships.

5.1 Introduction

Over the past two decades, psychologists have explored the importance of face-trait perception and the potential roles they play in our social interactions. People tend to judge very quickly whether someone has the intention and the ability to help or harm us, traits that psychologists call trustworthiness and dominance (Oosterhof & Todorov, 2008; Sutherland et al., 2013). These traits seem to be fundamental social dimensions across the world (Jones et al., 2021) as they underlie many other judgements that we make about others, particularly on first encounter. Since social interactions likely involve discerning how approachable and safe a person is, one potential characteristic about these traits is that they may be perceived in relation to the observers themselves.

Since the 1960s, researchers have investigated this relationship between the way we perceive ourselves and the way we perceive others. Dornbusch et al. (1965) found that there is more overlap in a perceiver's use of terms to describe other people than there is when others describe a single target person. This points to the possibility that there is likely a framework that the observer is using consistently to make judgements beyond the information extracted from the face. Markus, Smith and Moreland (1985) found that people are more likely to attribute certain person characteristics to others if they attribute the same traits to themselves. The observed person is also perceived as more similar and more likeable by the observer. Albright, Kenny, and Malloy (1988) found significant positive correlations between people's self-perceptions and their judgements of others based on specific characteristics such as attractiveness, formal dress sense, neatness of dressing appearance and age. Thus, it seems likely that there is a common tendency to judge others using mental frameworks based on our own traits and dispositions. While we extract social information from the face, it is also apparent that our own dispositions play an important role in our judgements of others.

Considering this, it is important to explore the *extent* to which our own dispositions, relative to face characteristics, play a role in the way we perceive a face. Many studies have focused on the target-face characteristics that influence our social judgements. For instance, smiling faces are consistently perceived as more trustworthy and high facial masculinity is consistently perceived as more dominant (e.g. Oosterhof and Todorov, 2008; Sutherland et al., 2013; Vernon et al., 2014). However, past research has also shown that there yet remains substantial variability among perceivers when making social judgements about the same faces. Specifically, while *inter-rater* agreement (i.e. correlations between pairs of perceivers) of social judgements of faces is typically low to moderate (about 0.3 to 0.4; Hönekopp, 2006; Oosterhof and Todorov, 2008), *intra-rater* agreement (i.e. test-retest reliability of the same perceiver) is considerably higher (about 0.7; Hönekopp, 2006), indicating the presence of important and consistent individual differences among perceivers (see also Germine et al. 2015; Sutherland et al., 2020).

Hönekopp (2006) showed that both shared taste and private taste contributed about equally to face attractiveness judgements. More recently, Hehman et al. (2017) estimated the effects of perceiver and target variance for a large range of social traits, including intelligence, competence, attractiveness, friendliness, trustworthiness, dominance, among others. Results from multilevel models quantified the variance that was attributable to the perceiver and to the target, and the authors found that the variance in ratings for all traits was due to *both* the perceiver and to the target. In other words, raters consistently relied on factors related to themselves and on factors related to the target face characteristics, and the extent of the relative contributions of perceiver and target depended on the trait.

Given that much variance in face judgements is related to the perceiver, this variance could be related to our individual differences. Face judgements are indeed influenced by the particular

groups or subgroups to which the perceivers belong, and are thus affected by ethnicity, gender (Hönekopp, 2006), and cultural background (Sofer et al., 2017). Face judgements are also influenced by individual differences in other dispositions such as anxiety levels (Willis et al., 2013) and personality traits (Mattarozzi et al., 2015). For example, Mattarozzi et al. (2015) found that those higher in trait agreeableness perceived faces as more trustworthy.

A recent study investigated the relationship between the structure of personality and face perception across the world (Oh, Martin and Freeman, 2022). Specifically, the authors investigated whether the personality structure of people living in one region relate to how they perceive social traits in the faces of others. Self-report personality trait ratings and face trait judgements were gathered from large existing datasets from 42 countries. Oh et al. (2022) computed average distances between trait pairs, separately for personality and face judgements. These resulted in representational dissimilarity matrices (RDMs) of the structure of both personality and face judgements for each world region. Results showed that, for most world regions, the two RDMs were significantly associated, indicating that the structure of personality was related to the structure of face perception. In other words, where there is a co-occurrence of traits in one's personality, this co-occurrence is also reflected in the social judgements made about the face. Furthermore, two countries that were more similar in their personality structure also showed higher similarity of the structure of face judgements.

It is possible that, even within a single world region, people who are more similar to each other in their own dispositions and traits also show more similar social trait judgements. Support for this comes from recent neuroimaging studies that have shown that pairs of people with similar personality traits also have similar brain activation patterns. Finn et al. (2020) investigated the relationship between pairwise brain similarity and personality similarity, using inter-subject representational similarity analysis (RSA; Kriegeskorte, Mur & Bandettini, 2008). The authors

analysed functional magnetic resonance imaging (fMRI) data of participants who viewed naturalistic video-clips, and personality ratings from the same participants. Results showed that inter-subject similarities in personality between participants were associated with inter-subject similarities in brain activity in ventral temporal cortex, frontal cortex, and cerebellum. Matz et al. (2022) found comparable results using both fMRI and electroencephalography (EEG). These results suggest that an alignment in brain activity can indeed be predicted by an alignment in personality.

Successful social interactions between people partly rely on a consistent understanding of our social world, so the alignment between people with similar dispositions may lead to successful social interactions, such as the initial stages of friendship development, affiliation, motivations towards community formation, and relevant in-group behaviours (Robbins & Krueger, 2005). Selfhout et al. (2010) showed that people choose others as friends who are similar to themselves in levels of agreeableness, extraversion and openness. Parkinson, Kleinbaum and Wheately (2018) have also demonstrated that friendship and social distance between people can be predicted by similarities in brain activity. In terms of trait perception, Bronstad and Russell (2007) showed that pairs of people in close relationships have higher inter-rater agreement when rating face attractiveness compared to pairs of strangers. However, it was not possible from this study to distinguish whether these shared perceptions were a result of spending time together and having shared experiences, or instead had the preferences themselves contributed to the forming of the relationships. In any case, it seems possible that the motivation to gravitate towards those who are like us and become friends may be related to an alignment in perspectives.

Taken together, the studies reviewed here suggest that there is an important link between how similar we are to others and its effect on any shared perspectives we might have. However, there remains the question as to the specific content of these aligned perspectives, especially relating

to social perception. Specifically, do we perceive people in a similar way if we have more similar personalities? The present study aims to investigate whether individuals who are more similar in their personality and self-perceived social trait dispositions also perceive face traits in a more similar manner. We used inter-subject RSA (Kriegeskorte et al., 2008; Finn et al., 2020) to analyse the association between person similarity and face judgement similarity. We hypothesised that pairs of people who are more similar in their own (self-perceived) personality and social traits also make judgements about faces in a more similar way. More specifically, we predicted that intersubject person similarity (based on self-perceived social traits and on personality traits) is significantly positively correlated with face judgement similarity (based on ratings of faces).

5.2 Methods

5.2.1 Participants

We recruited 320 participants in total (*M*age = 37.88; *SD* = 11.54; 182 identified as women, 133 as men, and 4 as other). We pre-registered our study, specifying sample size, inclusion, and exclusion criteria (https://doi.org/10.17605/OSF.IO/ZR92G). Participants were recruited through online participant pool Testable Minds (https://minds.testable.org) and received compensation for their time. Our goal was to test 300 participants. We stipulated that, if after removing participants because of the exclusion criteria we ended up with fewer than 300 participants, we would test 10 additional participants at a time until we obtained a final sample of at least 300 participants.

As inclusion criteria, all participants had to be aged 18 or older, have English as first language, and had to be living in the UK. Given that Oh et al. (2022) found an effect of country on the perception of social traits, here we wanted to avoid country as a source of variability and therefore focused on people in the UK only.

We had three pre-registered exclusion criteria to ensure data quality. First, there were four "catch" trials and participants who failed two or more catch trials were excluded. The catch trials are described in the Procedure. Second, we excluded participants who provided the same rating in 10 consecutive trials of the Face Judgements Task. Third, we excluded participants who responded too quickly (responses of 200ms or less) in 10% or more of the trials of the Face Judgements Task. The exclusion criteria were applied before analysis, and 13 participants were excluded. Two participants were excluded because of the first and second criteria and eleven participants were excluded because of the second criterion only. No participants were excluded because of the third criterion.

The final sample consisted of 307 individuals (*M*age = 37.97; *SD* = 11.55; 171 women, 132 men, 4 other). 268 participants were right-handed, 29 were left-handed, and 10 were ambidextrous. 238 participants described their ethnicity as White, 12 as Mixed, 27 as Asian, 25 as Black, and 5 as 'Other' (we used these ethnicity terms based on options from the 2021 Census; Office for National Statistics, 2021). All participants lived in the UK and reported their geographical location (see **Supplementary Information 1**).

The study was approved by the Psychology Research Ethics Committee at City, University of London.

Design and variables

We used a correlation design with two main variables: person similarity and face judgement similarity. Person similarity corresponds to (1) the similarity between participants in terms of their personality traits, and (2) the similarity between participants in terms of self-perceived

social traits. Face judgement similarity corresponds to the similarity between participants in terms of their judgements of social traits in faces.

5.2.2 Materials

Personality Traits Questionnaire. We used the International Personality Item Pool (IPIP-NEO-60) personality questionnaire (Maples-Keller et al., 2019). The IPIP-NEO-60 has shown to have strong internal consistency and validity when compared against questionnaires such as the NEO PI–R (Costa & McCrae, 1992; Maples-Keller et al., 2019). It consists of 60 statements, such as "Love to daydream.". Participants rated how well each statement described them using a 5-point scale ranging from 1 (Very inaccurate) to 5 (Very accurate). There are 12 items that correspond to each of the 5 Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism).

Questionnaire of Self-Perceived Social Traits. We created a questionnaire to measure self-perceived social traits. We used the 13 traits used by Oosterhof and Todorov (2008): Trustworthiness, Dominance, Attractiveness, Sociability, Emotional Stability, Aggressiveness, Intelligence, Care, Meanness, Confidence, Weirdness, Unhappiness and Responsibility. Participants were asked to 'Please rate how well each word describes YOU'. Each word was presented next to 9-point scale from 1 (Does not describe me) to 9 (Describes me).

Face Judgements Task. Stimuli consisted of 24 neutral face stimuli from the Chicago Face Database (https://www.chicagofaces.org; Ma, Correll, & Wittenbrink, 2015; Ma, Kantner, & Wittenbrink, 2020). Six faces per ethnicity were selected (three men and three women). Faces from four ethnicities were selected: White, Black, Latino and East Asian. The pictures had white backgrounds and were presented in colour with a 600 x 450-pixel size.

5.2.3 Procedure

The study was created on Testable (https://www.testable.org; Rezlescu et al., 2020). Participants completed the study online on a desktop or laptop computer. Participants read the information sheet and provided consent to take part in the study. Participants then completed the Face Judgements Task, followed by the Questionnaire of Self-Perceived Social Traits and finally the Personality Traits Questionnaire.

In the Face Judgements Task, participants were presented with the face pictures and were asked to judge each face on six social traits. The traits were: Trustworthiness, Dominance, Attractiveness, Sociability, Aggressiveness and Intelligence. These traits are known to be associated with the main dimensions in which we judge the faces of others (Oosterhof & Todorov, 2008; Sutherland et al., 2013) and were shown to explain a substantial part of the variance in social judgements ratings of participants in the UK (Jones et al., 2021). In each trial, participants were presented with one face paired with one trait and were asked to rate the face, e.g., 'How TRUSTWORTHY is this person?' on a 9-point scale from 1 (Not at all) to 9 (Extremely). Each of the 24 faces was paired with each of the traits resulting in 144 total trials. The order of trials was randomised for each participant.

A total of 4 attention quality check trials or "catch" trials were included in the study and were related to the Face Judgements Task. For the first attention check, after reading the instructions of the Face Judgements Task but before starting the task, participants were presented with a statement which read "Please confirm what you will be doing in the present study". Participants selected out of 3 simple options: "rating places", "rating faces" and "rating objects". The remaining attention check trials were presented at the end of the Face Judgements Task: these tested whether participants remembered which faces they had seen during the task. We reasoned that if participants had been paying attention, they would be able to distinguish faces that were

presented during the task (each repeated 6 times) from novel faces. There were three trials. In each trial, three faces were presented, where two of the faces had been shown in the Face Judgements Task and one face was novel. Participants had to select the novel face. The three faces within each trial were of the same ethnicity and gender (e.g., 3 white males, 3 black women, 3 Latino males).

Participants then completed the Questionnaire of Self-Perceived Social Traits and finally they completed the Personality Traits Questionnaire. Upon finishing the study, participants were prompted to read the study debrief and click 'FINISH' to save their results.

5.2.4 Data Analysis

We used inter-subject RSA (Kriegeskorte et al., 2008; Finn et al., 2020) to analyse the association between person similarity and face judgement similarity. Representational dissimilarity matrices (RDMs) were computed where each entry in a matrix represents the pairwise dissimilarity between two participants on each of our measures. Three main RDMs were computed: (1) RDM Personality Traits: each participant's scores on the Personality Traits Questionnaire were compared with each of the other participants' scores in a pairwise way (i.e., each matrix entry contains the dissimilarity value between pairs of vectors, with each vector containing 60 items), (2) RDM Self-Perceived Social Traits: each participant's scores on the Questionnaire of Self-Perceived Social Traits were compared with each of the other participants' scores in a pairwise way (i.e., each matrix entry contains the dissimilarity value between pairs of vectors, with each vector containing 13 items), and (3) RDM Face Judgements: each participant's ratings on the Face Judgements Task were compared with each of the other participant's scores in a pairwise way (i.e., each matrix entry contains the dissimilarity value between pairs of vectors, with each vector containing 144 items). We used Euclidean distance

as the main measure of dissimilarity in the RDMs. We first z-scored the vectors for each participant before computing Euclidean distances.

For the main analyses, we first correlated RDM Personality Traits with RDM Face Judgements and second, we correlated RDM Self-Perceived Social Traits with RDM Face Judgements. We used Spearman rank order correlations of the lower diagonals of each matrix. Because the entries in each matrix are not independent from each other (i.e., the same participant appears paired with each of the other participants), we cannot use parametric statistics which rely on independence assumptions (Selfhout et al., 2009; Chen et al., 2016). We thus tested whether each of these two correlations were significantly higher than zero by using random permutation tests. To do this, we conducted 1,000 iterations. For each iteration, we re-computed the Person Similarity RDM by first randomly permuting the order of participants and then re-computing the RDM. Then, we computed the correlation between the permuted Person Similarity RDM and the original RDM Face Judgements to obtain a correlation for that iteration. This allowed us to estimate the null distribution with 1,000 permuted correlations. To be considered significant, the 'actual' correlation needed to be higher than 95% of the 'permuted' correlations ($p < .05$).

We also computed confidence intervals of the correlations by bootstrapping (random resampling with replacement). For each of 1,000 resamples, we randomly selected 307 participants (resampling with replacement), re-computed all RDMs, and conducted all correlation analyses again. We provide the 95% confidence intervals of the correlations based on the 2.5^{th} and 97.5^{th} percentiles of the bootstrapped correlations.

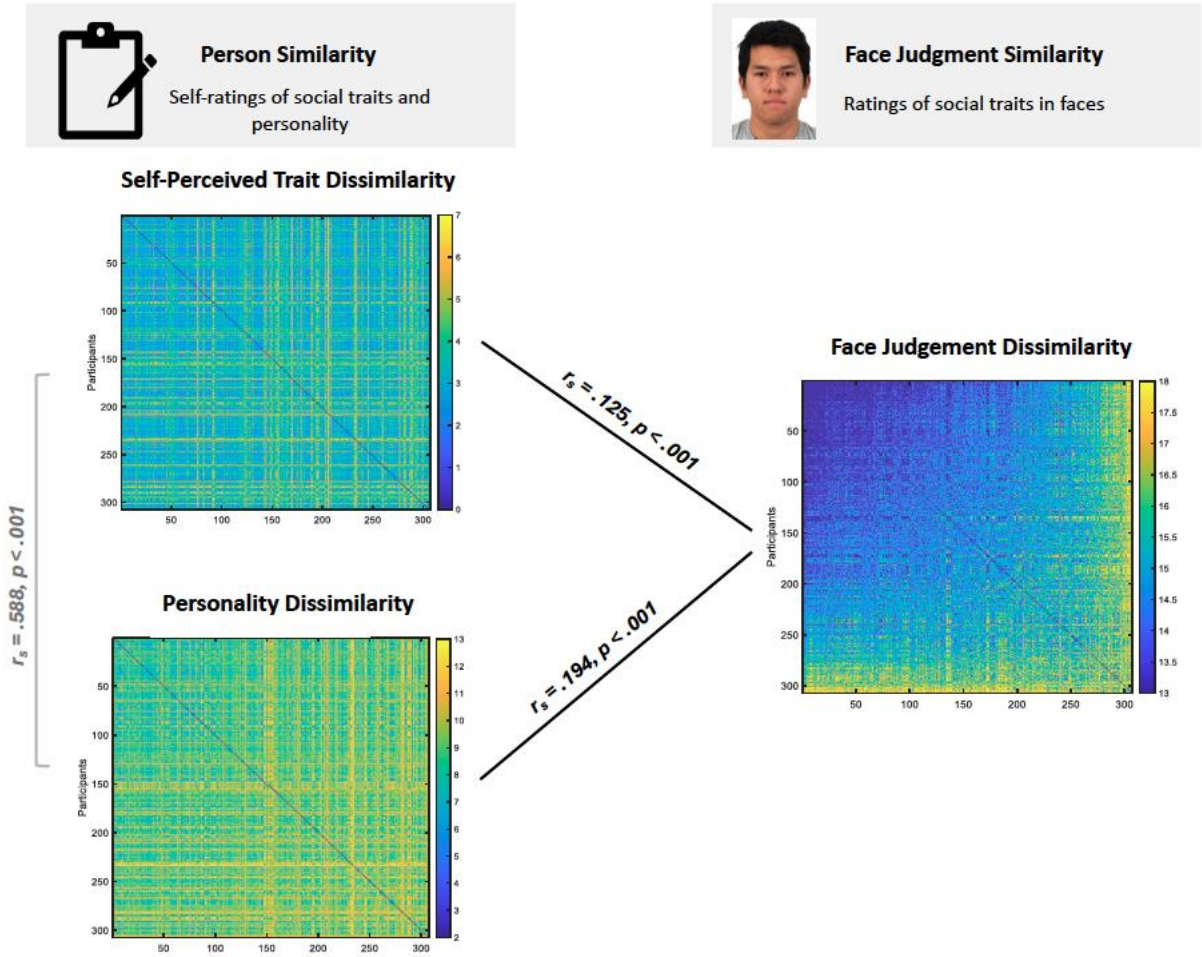
All analyses (except mixed effects modelling) were conducted in Matlab (version R2021b; Mathworks, www.mathworks.com). All anonymised data and code for main analyses are openly available: https://doi.org/10.25383/city.26519890.

5.3 Results

In this study we aimed to investigate whether person similarity is related to face judgement similarity. In other words, do people who are more similar to each other in terms of their dispositions also make more similar face judgements? In the Face Judgements Task, participants rated 24 faces, each face on six social traits (trustworthiness, dominance, attractiveness, sociability, aggressiveness, and intelligence). To measure their own dispositions and personality traits, participants completed a Questionnaire of Self-Perceived Social Traits, and a Personality Traits Questionnaire. For each of the three variables, we computed pairwise similarity for all pairs of participants. Briefly, we computed vectors of all z-scored ratings for each participant. We then computed Euclidean distances between vectors of all pairs of participants (for the same variable) to obtain pairwise distances (dissimilarities) between participants. This resulted in one intersubject representational dissimilarity matrix (RDM) for each of the three variables: RDM Face Judgements, RDM Self-Perceived Social Traits, and RDM Personality Traits (**Figure 26**). We then investigated whether intersubject person similarity could predict intersubject face judgment similarity.

Figure 28

Intersubject RDMs depicting the Pairwise Dissimilarity between all Participants' Self-Perceived Social Trait Ratings, Personality Ratings, and Face Judgements and their Spearman correlations.



Note. Intersubject RDMs for each variable and Spearman correlations between RDMs: RDM Self-Perceived Social Traits, RDM Personality Traits, RDM Face Judgements. Colour bars show Euclidean distances with blue indicating higher similarity (lower distance) and yellow indicating lower similarity (higher distance). RDM Face Judgements is ranked for visualisation purposes only, depicting the most similar pairs of participants from the top left of the matrix (i.e., darker blue entries) to the least similar pairs of participants at the bottom right of the matrix (i.e.,

yellower entries). The other RDMs are ranked based on this same RDM. Spearman correlations are shown between RDMs, with *p*-values estimated with random permutation testing.

Face CFD-AM-225-102-N From the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015).

5.3.1 Person similarity is associated with Face judgement similarity

Correlation between Self-Perceived Trait Similarity and Face Judgement Similarity. The first analysis looked at whether similar people (in terms of their self-ratings based on social traits) also make similar judgements of social traits in faces. To do this, RDM Self-Perceived Social Traits was correlated with RDM Face Judgements using Spearman rank-order correlation (after vectorising the lower diagonal of each matrix), resulting in a positive correlation of r_s = .125 (95% bootstrapped confidence interval = [.069 .198]) (**Figure 26**). We carried out non-parametric permutation tests to estimate a null distribution and test whether this correlation was significantly higher than zero. For each of 1,000 iterations, the RDM Self-Perceived Social Traits was re-computed after randomly shuffling (permuting) the order of participants, and then we re-computed the correlation between the original RDM Face Judgements and the permuted RDM Self-Perceived Social Traits. To be considered significant, the actual correlation had to be higher than 95% of the permuted correlations. Results showed that the actual correlation was higher than all 1,000 permuted correlations, and therefore was significantly higher than zero ($p < .001$, permutation tests). In other words, the more similar pairs of people were in their ratings of their own social traits, the more similar their ratings were of other people, which supports our hypothesis.

Correlation between Personality Similarity and Face Judgement Similarity. The same procedure was carried out RDM Personality Traits was correlated with RDM Face Judgements. There was a positive correlation of r_s = .194 (95% bootstrapped CI = [.139 .262]), which was significantly higher than zero (p <.001) (**Figure 28**). These results revealed that the more similar people are

in their personality structure, the more likely they are to rate faces in a similar way, which supports our hypothesis. For completeness, we also correlated the RDM Self-Perceived Social Traits with the RDM Personality Traits (see **Supplementary Information 2**).

Split-half reliability of Face Judgement Similarity. We also computed split-half reliability of the RDM Face Judgements as way of estimating the highest possible correlation we could expect with face judgement similarity (i.e., as a way of estimating the noise ceiling). To compute split-half reliability, we conducted 1,000 random splits of the Face Judgements Task data, and re-computed the RDM Face Judgements for each split, and then correlated the RDMs across splits. More specifically, for each iteration, we randomly divided the data on the Face Judgements Task for each participant in two halves (with 72 trials in each half). We then computed an RDM Face Judgements for Split 1 and an RDM Face Judgements for Split 2. For each iteration, we correlated the two RDMs using Spearman correlations. The resulting correlations across 1,000 splits varied between .474 and .605, with a mean of r_s = .558. We interpret this value as the maximum correlation we could expect of any variable or model with the RDM Face Judgements. When considering this noise ceiling, the person similarity variables seem to explain a substantial proportion of the variance in face judgement similarity (22.40% for RDM Self-Perceived Social Traits and 34.77% for RDM Personality Traits), though there is still considerable unexplained variance in face judgement similarity.

5.3.2 Person similarity is associated with Face judgement similarity after controlling for demographic variables

We next investigated whether the positive correlations between Person RDMs and Face Judgements RDMs could be accounted by similarities in demographic variables. In other words, we investigated whether the correlations between person similarity and face judgement similarity could be accounted by similarities in the demographic variables age, gender, ethnicity, and

participant geographical location. We thus carried out multiple regression analyses with RDM Face Judgements as the outcome variable and person similarity as predictor variables, controlling for age, gender, ethnicity, and participant location. As before, and because the entries in each RDM are not independent, we conducted permutation tests to test whether parameter estimates from the regression analyses were significantly different from zero. P-values for the beta-values were estimated using random permutations (1,000 iterations) to estimate null distributions of the parameter estimate for each variable. We additionally conducted linear mixed effects modelling with cross-random effects as suggested by Chen et al (2017) and results are comparable to the ones presented below (see **Supplementary Information 3**).

RDMs were computed for each demographic variable (RDM Age, RDM Gender, RDM Ethnicity, and RDM Region). Pairwise distances between each pair of participants were computed based on each demographic variable separately, thus producing a dissimilarity matrix for each variable (i.e., for RDM Age, pairwise distances in age were computed between all participants). RDM Age used Euclidean distance as the dissimilarity measure for RDM Age. For RDM Gender and RDM Ethnicity (which were based on categorical variables), we classified each pair of participants as same ('0') or different ('1'). For RDM Region, the actual geographical distances between latitude and longitude co-ordinates for the centre of each UK region were computed using Haversine distance. Co-ordinates for the midpoint of the 12 UK regions were obtained using GeoMapApp (www.geomapapp.org; Ryan et al., 2009).

Multiple regression predicting Face Judgement Similarity with Self-Perceived Trait Similarity and demographic variables as predictors. We first looked at the association between RDM Self-Perceived Social Traits and RDM Face Judgements. As expected, simple regression analysis showed that similarity of self-perceived traits could significantly predict face judgements similarity ($\beta = .123$, $p < .001$, permutation tests), explaining 1.78% of the variance ($R^2 = .0178$).

When all demographic variables were included as predictors along with the RDM Self-Perceived Social Traits, the total model explained 6.13% of the variance in face judgement similarity (R^2 = .0613). Looking at the parameter estimates for each predictor variable (**Table 9**), similarity in all demographic variables were associated with face judgement similarity (especially ethnicity), but there is unique substantial variance explained by person similarity (β = .138, p < .001).

Table 9

Regression analyses using Self-Perceived Trait Similarity to predict Face Judgement Similarity.

Face Judgement Similarity	Simple Regression β, p	Multiple Regression β, p
Self-Perceived Similarity	.123, p < .001	.138, p < .001
Age		.065, p = .007
Gender		.033, p = .013
Ethnicity		.184, p = .001
Region		.082, p = .036

Note. Simple regression results including only Self-Perceived Trait Similarity as predictor. Multiple regression results including Self-Perceived Trait Similarity and demographic variables as predictors: β = parameter estimates; *P*-values estimated by random permutations.

Multiple regression predicting Face Judgement Similarity with Personality Similarity and demographic variables as predictors. We next looked at the association between RDM Personality Traits and RDM Face Judgements. Simple regression showed that similarity in personality traits could predict face similarity (β = .199, p < .001, R^2 = .0397). When all demographic variables were included as predictors along with the RDM Personality Traits, the model explained 7.8% of the variance in face judgement similarity (R^2 = .078). Looking at the

parameter estimates for each predictor variable (**Table 10**), similarity in all demographic variables were associated with face judgement similarity (specially ethnicity), but there is unique substantial variance explained by personality similarity (β = .190, p < .001).

Table 10

Regression analyses using Personality Similarity to predict Face Judgement Similarity.

	Simple Regression β, p	Multiple Regression β, p
Personality Similarity	.199, p < .001	.190, p < .001
Age		.059, p = .017
Gender		.031, p = .014
Ethnicity		.170, p < .001
Region		.085, p = .032

Note. Simple regression results including only Personality Similarity as predictor. Multiple regression results including Personality Similarity and demographic variables as predictors. β = Parameter estimates; *P*-values estimated by random permutations.

5.3.3 Results are stable with at least 120 participants and across traits

We next conducted exploratory analyses to investigate the stability of the above results. We first investigated the stability of the RDM correlations across multiple sample sizes, and found that correlations between RDMs became stable with around 120 participants (see **Supplementary Information 4**). This analysis also demonstrated that the results are not just dependent on our large sample size, nor are they dependent on only a few participants.

We next investigated whether the association of person similarity with face judgement similarity is stronger for some face judgements than others. Results showed that Person similarity was

significantly associated with Face judgement similarity for all traits separately, though all correlations were lower than when considering all traits together (see **Supplementary Information 5**). We also conducted exploratory analyses investigating the extent to which individual differences in specific social and personality traits contribute to average face judgements (see **Supplementary Information 6**). These results suggested that our main findings of similarity in face judgements being predicted by person similarity cannot be fully explained by associations of any specific Person traits with specific face judgements, and instead seem to reflect patterns of responses across traits. Finally, we computed inter-rater agreement of face judgements to investigate agreement of face ratings across participants (see **Supplementary Information 7**). Results showed low to moderate inter-rater agreement, which is comparable to results from previous studies (e.g. Oosterhof & Todorov, 2008).

5.4 Discussion

In the present study, we investigated whether people who are similar in their personality and social dispositions also make more similar social judgements about others. We measured pairwise similarity between participants' traits and dispositions (i.e., person similarity) using their responses from self-report questionnaires on social traits and the Big 5 personality traits. We separately looked at similarity between participants' face judgements using their ratings of unfamiliar faces based on six social traits (trustworthiness, dominance, attractiveness, sociability, aggressiveness, and intelligence). We found that there was a small but reliable correlation between pairwise person similarity and pairwise face judgement similarity. Thus, two people who perceive themselves in a similar way, in terms of their social trait dispositions as well as their personality traits, also make more similar social judgements about others' faces. This effect is present even after controlling for age, gender, ethnicity, and geographical location (though, these variables, especially ethnicity, also contribute to a significant proportion of

independent variance found in face judgement similarity). The results suggest that pairs of people who are similar in their own traits and dispositions *perceive* others in a similar manner.

The results from this study provide support for our hypothesis and are consistent with results from previous studies (Oh et al., 2022; Matz et al., 2022; Finn et al., 2020). In particular, Oh et al. (2022) demonstrated that the personality structure within a given world region is related to the structure of face perception within that same region. In our study, we extended these findings to show that these associations are not dependent on world region only, and can be seen within the same country. Thus, within the UK only, pairs of people who have similar personalities also perceive others' faces in a more similar way. Therefore, individuals' own personalities and dispositions (within the same culture and speaking the same native language) are also related to how they perceive faces. It would be interesting for future studies to replicate this effect within different countries and world regions.

Finn et al. (2020) and Matz et al. (2022) demonstrate that those with similar personalities also had more similar brain activity. Here, we extended these findings to show that similarity in personality and social traits is also related to their social perception, specifically in participants' first impressions of faces. It would be interesting to investigate in future studies whether similarity in dispositions predicts similarity in other behaviours as well, and specifically if the association observed here depends more strictly on using social stimuli such as faces.

In the present study, we used self-report measures to capture person similarity. Future studies could explore alternative ways of measuring person similarity. It could be interesting to measure similarity in social behaviours, such as volunteering (e.g., at homeless shelters) (Matz et al., 2022) or perceived loneliness (Baek et al., 2023). Similarities in upbringing, environments,

culture (Sofer et al., 2017), mood, and even cognitive domains may also contribute to the variance in face judgement similarity.

One important limitation is that we did not control for the possibility of participants knowing each other. Given that our data was collected online, it was impossible for us to obtain such information. In a country of more than 60 million people, it seems unlikely that many participants knew each other. Future studies could record IP addresses or latitude/longitude for more accurate determination of location, and even recruit a more equal number of participants from each region. It could also be interesting for future studies to directly manipulate whether participants knew each other or not.

It could be that the similarities between participants' dispositions are actually relevant to how people form friendships and relationships. Selfhout et al. (2010) showed that friends choose others who are similar to themselves in levels of agreeableness, extraversion and openness. If people with similar dispositions also perceive others in a similar way, it seems likely that they will trust and approach the same people. It would be interesting to test this directly in small-scale communities while they are forming (Parkinson et al., 2018).

To conclude, we investigated the relationship between dyadic person similarity and the similarity in face judgements. We showed that pairs of people who are similar in their dispositions and personalities also make more similar judgements about others' faces. Additionally, those who were the same gender and ethnicity, and those who were nearer in terms of age and geographical location also made more similar judgements, but these effects were independent from the one accounted for by person similarity. Future studies could investigate whether the association between person similarity and face judgement similarity is important for how people build friendships.

References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgements at zero acquaintance. *Journal of personality and social psychology*, *55(3)*, 387.

Baek, E., Hyon, R., López, K., Du, M., Porter, M., & Parkinson, C. (2023). Lonely individuals process the world in idiosyncratic ways. *Psychological Science, 34(6)*, 683-695.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1)*, 1-48.

Bronstad, M., & Russell, R. (2007). Beauty is in the 'we' of the beholder: Greater agreement on facial attractiveness among close relations. *Perception*, *36*, 1674-1681.

Chen, G., Shin, Y., Taylor, P., Glen, D., Reynolds, R., Israel, R., & Cox, R. (2016). Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *Neuroimage, 142*, 248-259.

Chen, G., Taylor, P., Shin, Y., Reynolds, R., & Cox, R. (2017). Untangling the relatedness among correlations, part II: Inter-subject correlation group analysis through linear mixed-effects modeling. *Neuroimage, 147*, 825-840.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO Personality Inventory-Revised (NEO-PI-R)*. Psychological Assessment Resources.

Dornbusch, S. M., Hastorf, A. H., Richardson, S. A., Muzzy, R. E., & Vreeland, R. S. (1965). The perceiver and the perceived: Their relative influence on the categories of interpersonal cognition. *Journal of Personality and Social Psychology*, *1(5)*, 434.

Finn, E., Gleren, E., Khpjandu, A., Nielson, D., Molfese, P., Handwerker, D., Bandettini, P. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *Neuroimage*, *215*, 116828.

Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A., Smoller, J. W., Kwok, H., Anthony, S. E., Nakayama, K., Rhodes, G., & Wilmer, J. B. (2015). Individual Aesthetic Preferences for Faces Are Shaped Mostly by Environments, Not Genes. *Current Biology, 25(20)*, 2684–2689.

Hehman, E., Sutherland, C., Flake, J., & Slepian, M (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology, 113(4)*, 513-529.

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgements of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32(2)*, 199-209.

Jones et al. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour*, *5*, 159-169.

Kriegeskorte, N., Mur, M., Ruff, D., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*, 1126–1141.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82(13)*, 1-26.

Ma, Correll, & Wittenbrink (2015). The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. *Behavior Research Methods, 47*, 1122-1135.

Ma, Kantner, & Wittenbrink, (2021). Chicago Face Database: Multiracial Expansion. *Behavior Research Methods, 53,* 1289-1300.

Maples-Keller, J. L., Williamson, R. L., Sleep, C. E., Carter, N. T., Campbell, W. K., & Miller, J. D. (2019). Using item response theory to develop a 60-item representation of the NEO PI–R using the International Personality Item Pool: Development of the IPIP–NEO–60. *Journal of personality assessment*, *101(1)*, 4-15.

Markus, H., Smith, J., & Moreland, R. L. (1985). Role of the self-concept in the perception of others. *Journal of Personality and social Psychology*, *49(6)*, 1494.

Mattarozzi, K., Todorov, A., Marzocchi, M., Vicari, A., & Russo, P. M. (2015). Effects of gender and personality on first impression. *PloS One*, *10*(9), e0135529.

Matz, S. C., Hyon, R., Baek, E. C., Parkinson, C., & Cerf, M. (2022). Personality similarity predicts synchronous neural responses in fMRI and EEG data. *Scientific Reports*, *12*(1), 14325.

Office for National Statistics. (2021). *England.* https://explore-local-statistics.beta.ons.gov.uk/areas/E92000001-england.

Office for National Statistics. (2021). Ethnic group variable: Census 2021. https://www.ons.gov.uk/census/census2021dictionary/variablesbytopic/ethnicgroupnationalidentitylanguageandreligionvariablescensus2021/ethnicgroup.

Oh, D., Martin, J. D., & Freeman, J. B. (2022). Personality across world regions predicts variability in the structure of face impressions. *Psychological Science*, *33*(8), 1240-1256.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1), 332.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. Progress in Brain Research, 253, 243-262.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*(1), 32-47.

Ryan, W. B. F., Carbotte, S. M., J, Coplan., O'Hara, S., Melkonian, A., Arko, R., Weissel, R.A., Ferrini, V., Goodwillie, A., Nitsche, F., Bonczkowski, J., and Zemsky, R. (2009). Global Multi-Resolution Topography (GMRT) synthesis data set, *Geochemistry*. *Geophysics*. *Geosystems*., 10, Q03014, doi:10.1029/2008GC002332.

Selfhout, M., Burk, W., Branje, S., Denissen, J., Van Aken, M., & Meeus, W. (2010). Emerging late adolescent friendship networks and Big Five personality traits: A social network approach. *Journal of Personality*, *78*(2), 509-538.

Selfhout, M., Denissen, J., Branje, S., & Meeus, W. (2010). In the eye of the beholder: Perceived, actual, and peer-rated similarity in personality, communication, and friendship intensity during the acquaintanceship process. *Journal of Personality and Social Psychology*, *96*(6), 1152-1165.

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D., & Todorov, A. (2017) For your local eyes only: Culture-specific face typicality influences perceptions of trustworthiness. *Perception, 46(8)*, 914-928.

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 of the United States of America*, 117(19), 10218–10224.

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105-118.

Sutherland, C. A. M., Rhodes, G., Burton, N. S., & Young, A. W. (2020). Do facial first impressions reflect a shared social reality? *British Journal of Psychology*, *111*(2), 215–232.

Vernon, R., Sutherland, C., Young, A., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences, 111(32)*, E3353-E3361.

Willis, M. L., Dodd, H. F., & Palermo, R. (2013). The relationship between anxiety and the social judgements of approachability and trustworthiness. *PLoS One*, *8*(10), e76825.

SUPPLEMENTARY INFORMATION

Dyadic Person Similarity Predicts Similarity in Face Judgements

Supplementary Information 1

Participants' regions of residence in the United Kingdom

All participants lived in the UK and reported their geographical location (see **Table S1**; we used UK regions based on Census 2021 data published from the Office of National Statistics: https://explore-local-statistics.beta.ons.gov.uk/areas/K02000001).

Table S1

Participants' regions of residence in the United Kingdom

	Region	*N Participants*
	North East	7
	North West	35
	Yorkshire and The Humber	23
	East Midlands	17
	West Midlands	29
	East of England	31
	London	56
	South East	37
	South West	22
	Wales	13
	Scotland	32
	Northern Ireland	5
Total	12	307

Note. Number of participants living in each of twelve UK regions (taken from the Census 2021; the Office of 10 National Statistics).

Supplementary Information 2

Correlation between Self-Perceived Trait Similarity and Personality Similarity

For completeness, we correlated the RDM Self-Perceived Social Traits with the RDM Personality Traits. This would inform us about the consistency in self-report person similarity between participants. Results showed that there was a positive and substantial correlation between the two RDMs (r_s = .588; 95% bootstrapped CI = [.537 .641]), which was significantly higher than zero (p < .001, permutation tests). These results suggest that there is high consistency in the two self-report person similarity measures.

Supplementary Information 3

Linear mixed effects modelling with crossed random effects

To address the non-independence issues of intersubject correlation analyses when controlling for covariates, Chen and colleagues (2017) recommended the use of linear mixed effects modelling with crossed random effects to account for repetitions of participants in multiple dyads. We conducted these analyses in R Statistical Software (version 4.4.2; R Core Team, 2024). We followed the procedures suggested by Chen et al. (2017) (doubling the data, correcting for degrees of freedom) and fit linear mixed models using lme4 (version 8 1.1.35.5; Bates et al., 2015) and lmerTest (version 3.1.3; Kuznetsova et al., 2017). We thus used linear mixed effects models to predict face judgement similarity with person similarity (separate models for Self-Perceived Trait Similarity and Personality Similarity) and demographic variables as fixed effects, and participants in each dyad as random effects. Results showed that Self-Perceived Trait Similarity was a significant predictor of Face Judgement Similarity (β = .076, t = 21.840, p < .001). Personality Similarity was also a significant predictor of Face Judgement Similarity (β = .072, t = 23.61, p < .001). These results are in line with the multiple regression results and show that person similarity can predict face judgement similarity, even after controlling for demographic predictors.

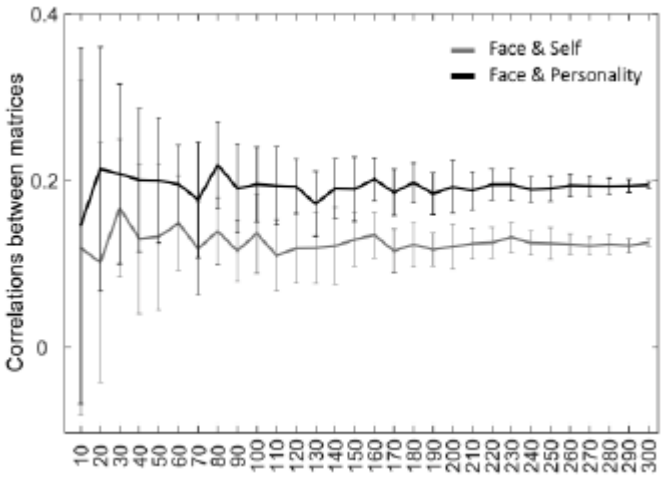
Supplementary Information 4

How many participants are needed for stable correlations?

We investigated the stability of the RDM correlations across multiple sample sizes. Correlations between RDMs were computed at cumulative intervals, where the sample size of participants increased by 10 until reaching 300 (i.e., iterations from 10-300 in steps of 10). For each step, we carried out 20 iterations. In each iteration, the required number of participants were randomly sampled from the total number of participants and computed all the same correlation and simple regression analyses as above. Results were then averaged across the 20 iterations of each sample size. Correlations between RDM Self-Perceived Social Traits with RDM Face Judgements and correlations between RDM Personality Traits with RDM Face Judgements became stable with around 120 participants (Figure S4-1). From this, it is apparent that a sample size between 120 and 160 participants may be sufficient. It is also clear that the results are not just dependent on our large sample size, nor are they dependent on only a few participants.

Figure S4-1

RDM correlations across different sample sizes.



Note. The plot shows the correlation results with different number of participants. For each iteration, a random sample of participants was selected and the correlation analyses were carried out. This process was carried out cumulatively, starting with a sample size of 10 participants until 300 participants. For each step, 20 random samples of the required size were selected, and the main correlation (correlations between face judgement similarity and self-perceived trait similarity, as well as face judgement similarity and personality) were conducted. The line shows mean correlations for each sample size and the error bars show standard deviations.

Supplementary Information 5

Does Person similarity predict similarity in trait judgements of specific traits?

We investigated whether the association of person similarity with face judgement similarity is stronger for some face judgements than others, (e.g., the correlation between similarities in self-perceived traits and personality on face *trustworthiness* judgements, specifically). It could be the case that consistencies in people's dispositions are reflected in certain traits more strongly than in others. In other words, can we use person similarity to predict face judgement similarity for all of the social traits? One problem of doing these correlations separately is that there are fewer trials per condition, so the sizes of these correlations cannot be directly compared to the ones using all traits simultaneously. For RDM Self-Perceived Traits, results revealed that there were small but significant positive correlations with individual face-trait RDMs, namely for face trustworthiness (r_s = .038, p < .001, random permutations), dominance (r_s = .045, p < .001), attractiveness (r_s = .067, p < .001), face sociability (r_s = .061, p < .001), aggressiveness (r_s = .011, p = .005) and face intelligence (r_s = .057, p < .001). For the correlations between RDM personality and the RDMs of each face traits, results revealed that there were also small but positive significant correlations for all traits, namely for trustworthiness (r_s = .067, p < .001), dominance (r_s = .072, p < .001), attractiveness (r_s = .11, p < .001), sociability (r_s = .090, p < .001), aggressiveness (r_s = .036, p < .001), and intelligence (r_s = .058, p < .001). These results showed that social and personality similarity between individuals could also predict similarity of trait ratings for each of the traits individually, though the highest correlations were between personality similarity and similarly of judgements of attractiveness and sociability.

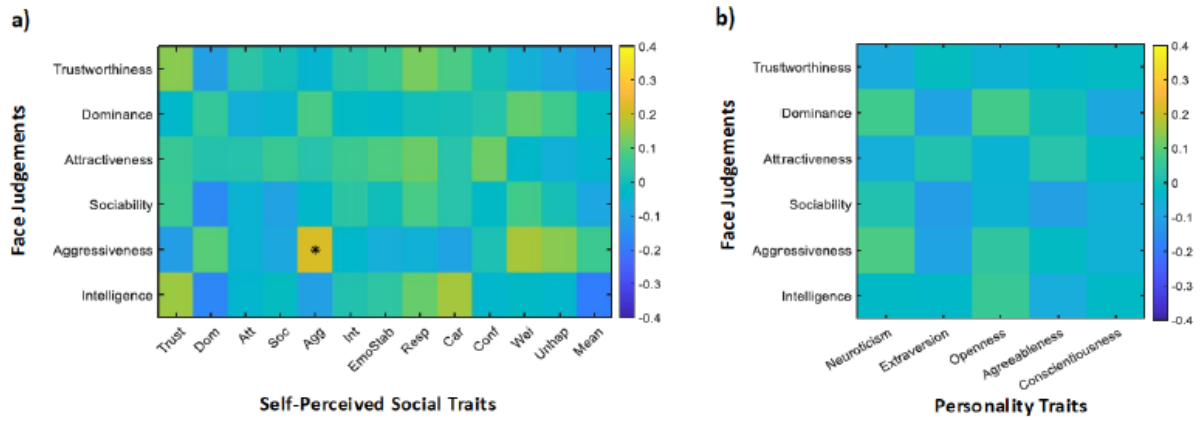
Supplementary Information 6

How do individual differences in person traits predict mean ratings of faces?

We conducted further analyses investigating the extent to which individual differences in specific social and personality traits contribute to the average face judgements. Specifically, we wanted to understand whether scoring higher or lower on each of the personality and self-perceived social traits is correlated with mean of ratings of the face judgements for each trait. For example, if someone rates themselves as more trustworthy, are they more likely to rate others as more trustworthy? Matarozzi et al. (2015) had found that individuals who score higher on Agreeableness tend to rate faces as more trustworthy. Here, we extended these analyses to all traits measured. Specifically, for each participant, we computed: (i) individual scores for each of the 13 self-perceived social traits (based on Questionnaire of Self-Perceived Social Traits), (ii) individual scores for each of the Big Five traits (based on the Personality Questionnaire), and (iii) mean ratings for each social trait on the Face Judgements Task (by computing the mean of all face ratings for each trait for the same participant). We then correlated measures in (i) and (iii), and (ii) and (iii) (Figure S6-1). We corrected for multiple comparisons using False Discovery Rate (FDR) correction. Regarding correlations between self-perceived traits and face judgements (Figure S6-1a), there was only a significant positive correlation between self-perceived aggression and face aggression. With regards to correlations between individual differences in personality traits (Big 5 scores) and mean face judgements (Figure S6-1b), there were no significant correlations after correction for multiple comparisons. These results demonstrate that our main findings of similarity in face judgements being predicted by person similarity cannot be fully explained by associations of any specific Person traits with specific face judgements, and instead seem to reflect patterns of responses across traits.

Figure S6-1

Correlations between individual differences in self-perceived traits (social traits and personality) and mean face judgements



Note. (a) Correlations between individual differences in 13 self-perceived social trait ratings and mean face judgements for each of the 6 social traits. Trust = Trustworthiness; Dom = Dominance; Att = Attractiveness; Soc = Sociability; Agg = Aggressiveness; Int = Intelligence; EmoStab = Emotional Stability; Resp = Responsibility; Car = Caring; Conf = Confidence; Wei = Weird; Unhap = Unhappy. (b) Correlations between individual differences in each of the Big-5 personality traits and mean face judgements for each of the 6 social traits. The asterisk (*) indicates significant correlations after FDR correction for multiple comparisons.

We did not replicate the significant positive correlation between agreeableness and trustworthiness, but our sample was smaller than the one in Mattarozzi et al. (2015) and we controlled for multiple comparisons of all traits. After correction for multiple comparisons, we only found that people who rated themselves higher on trait aggressiveness had also rated others' faces as more aggressive. In any case, our results suggest that individual differences in personality and self-perceived traits are not associated only with specific changes in means of face trait judgements, but instead with changes in *patterns* of judgements across many traits.

Supplementary Information 7

How much do participants agree on the face judgements?

We next computed inter-rater agreement of face judgements to investigate agreement of face ratings across participants. We computed inter-rater agreement in two ways. First, we computed inter-rater agreement by using a 'pairwise approach', in which we computed the correlation of the ratings of each participant with the ratings of each of the other participants. Second, we computed inter-rater agreement by using a 'one *versus* others approach' by computing the correlation of the ratings of each participant with the average of all the other participants. We computed these measures for all trials of the Face Judgements Task and also per trait. Results are shown in Table S7-1 and show low to moderate inter-rater agreement, which is comparable to results from previous studies (e.g. Oosterhof & Todorov, 2008), and again demonstrates that there are substantial individual differences in face trait ratings. Attractiveness was the trait with the highest inter-rater agreement.

Table S7-1

Inter-rater agreement for the Face Judgement Task (all traits and per face trait).

	Mean inter-rater agreement (SD)	
Face Judgement Task	Pairwise	One vs others
All traits (all trials)	.23(.08)	.48(.16)
Trustworthiness	.13(.08)	.36(.22)
Dominance	.16(.09)	.39(.22)
Attractiveness	.28(.10)	.50(.20)
Sociability	.22(.09)	.44(.20)
Aggressiveness	.14(.08)	.39(.23)
Intelligence	.19(.09)	.42(.21)

Note. In the 'pairwise approach', we computed the correlation of the ratings of each participant with the ratings of each of the other participants. In the 'one *versus* others approach', we computed the correlation of the ratings of each participant with the average of all the other participants.

Chapter 6: General discussion

6.1 Thesis aims and summary

The present thesis aimed to explore factors that influence face perception in humans. First, we were interested in how representations of faces change as we become familiar with others, and we were also interested in whether similar people also make similar face judgements of others. We investigated changes in face representations in two ways: first, in Chapter 3, we measured how explicit social trait representations of face judgements change with familiarisation and found that this is indeed the case for some of the traits. In Chapter 4, we explored changes in the representations of faces in the brain and found that representations do appear to change for each of the face-selective ROIs. In Chapter 5, we then assessed individual differences in face trait perception by assessing whether person similarity (i.e., personality and social traits) predicts similarities in face trait judgements and results reveal this to be the case, even at the dyadic level. Below, I summarise the findings for each chapter in more detail and discuss their theoretical and practical implications.

6.1.1 Main findings

6.1.1.1 Chapter 3: Social trait representations of faces change with familiarisation

Chapter 3 of the present thesis investigated how social trait representations change with familiarisation. In particular, we investigated how these perceptions might *change* with increased familiarisation for the same faces, as is characteristic of human interactions across time, using representational similarity analysis (RSA; Kriegeskorte et al., 2008a). That is, how are faces perceived in relation to each other based on, e.g., trustworthiness? And how might this change as we get to know them? We hoped this would shed light on the potential ways in which the mind changes the way it organises faces based on certain information.

Participants watched a television series over the course of 2-3 weeks in order to become perceptually familiarised with the individuals in the show. Both before and after watching the series, participants provided ratings of the faces based on attractiveness, dominance, likeability and trustworthiness. We analysed changes in the representations of these face traits in 4 main ways: a) measure whether there was an overall shift in ratings, b) measure within-participant changes in the representational structure of faces based on these ratings, c) assess the agreement in the representations between participants, both before and after familiarisation, and thus, whether there was a significant change in agreement, and d) assess the similarity in the representations between each of the traits themselves to see whether face representations for one trait overlap with that of another trait, thus potentially sharing the same social information.

Results showed that there was only an overall increase in perceived trustworthiness and dominance, but not attractiveness and likeability. Increases may reflect how tenuous our perceptions are in the early stages of getting to know someone, and that with repeated exposure, we might find people to be more trustworthy and dominant overall. These results are in line with Lavan et al. (2021), showing that with repeated exposure to voices, people are more trusting of them, and perceive them to be more dominant.

Within-participant changes followed the same pattern as the previous analysis, indicating a change in face representations based on trustworthiness and dominance. These findings in particular may suggest that not only is there a general increase in trusting and perceived dominance, but that participants initially relied on a set of characteristics to make judgements about who is trustworthy and dominant, and likely relied on a new set of characteristics to make these judgements, likely abstract in nature (Schwartz & Yovel, 2016). In other words, participants seem to have changed their minds about who is trustworthy and who is dominant as they learned new information about the individuals in the show, indicating that these judgements are quite dynamic within the first two weeks of familiarisation and that we may shift from image-based reliance to a reliance on non-visual person-related information (Bruce & young, 1986;

Schwartz & Yovel, 2016). This may indicate a change in the function of perceived trustworthiness as we go from unfamiliar face perception to familiar face perception. It may become less related to immediate-term survival-based schema (i.e., "Does this person intend to harm or help me?"; Oosterhof & Todorov, 2008) and more related to longer-term, implicit social schema (which may, for instance, be related to reliability honesty, and similar concepts; Rhodes et al., 2013). In light of Analysis D in Chapter 3, perceived trustworthiness may become related to the affective experience of liking the person (i.e., "Do I like who this person is? Would I like to spend time with this person?"), where trustworthiness may become a more prominent tool for social bonding and co-operative behaviours rather than remain an initial automatic instinct.

According to our knowledge, the third analysis is particularly novel. Results for this analysis revealed that there was significant agreement in face representations across participants, for all traits except trustworthiness. In other words, participants seemed to rely on similar visual mechanisms to make judgements about attractiveness, dominance and likeability, but relied on more idiosyncratic characteristics to make judgements about trustworthiness. Post-familiarisation, trustworthiness representations became substantially similar across participants. This may indicate that participants may have come to rely on similar information to make these judgements, which may relate to abstract person and conceptual knowledge (Bruce & young, 1986; Schwartz & Yovel, 2016). Notably, inter-subject correlations for trustworthiness representations surpassed that of all the other social traits, indicating a significant increase in the reliance on similar information to make the judgement. Could this point to some kernel of truth in trustworthiness judgements (Allport, 1954)? It was suggested that the (ontological) essence of a trait can be determined based on consistencies in behaviour, thoughts and feelings (John et al., 2010), patterns of which would have been detected as participants watched the show. As our interpretations of the behaviours of the candidates, possibly in terms of how reliable and consistent they are, honest, and even likeable (in light of Analysis D, Chapter 3), become markedly consistent within a two-week period, it would be useful to garner explicit free-recall

judgements and descriptions about what behaviours and traits contribute to people's changing view of a person's trustworthiness during familiarisation. This would be a useful way to ascertain aspects of what people generally think make a person trustworthy, which may cement some kernel of truth about trustworthiness as a disposition (Allport, 1954).

While dominance representations were significantly associated between participants, there was a non-significant trend across sessions in this agreement towards idiosyncrasy. This could relate to a potential polarity of observer-target judgements, or even variability in any sub-traits of perceived dominance (i.e., perceived anger, masculinity, competence, aggression, meanness; Oosterhof & Todorov, 2008; Sutherland et al., 2013). In other words, people may use these other traits as a proxy for perceived dominance, especially considering that there may be at least two types of dominance (i.e., physical and social dominance; Archer, 1988; Chagnon, 1988; Pérusse, 1993). Thus, explicit reports on how perceivers define this trait might shed light on the potential variance in this trait judgement, which is important as dominance appears to be fundamental to our perception of others (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Future studies could attempt to distinguish between the representational structure of faces based on explicit judgements of physical and social dominance. Moreover, it is important to consider that the nature of the show called for explicit displays of leadership, competence, and the capability of carrying out intentions, which may also lead to forms of (verbal) aggression at times. All facets of which denote forms of dominance and may have directly influenced ratings of the trait over time.

More generally, due to the fact that participants watched the same show, the increase in agreement in ratings was likely induced by this, as was expected. Future studies could explore changes as participants watch different shows and assess any commonality in changes across the representational structure of their traits, so that the general *pattern* of change can be assessed without the variance associated with the sameness of a single show.

The final analysis assessed the representational similarity between each of the traits to understand whether the perception of faces were organised in a similar way across the traits, and then whether this changed with familiarisation. Results showed that some trait representations were initially associated with each other, particularly for attractiveness and likeability, and attractiveness and trustworthiness, in line with previous findings (Mileva et al., 2019) and the premise of The Halo Effect (Thorndike, 1920; Lachman & Bass, 1985; Wilson & Eckel, 2006). However, relations between all social trait representations were significant post-familiarisation, except for attractiveness and dominance (which remained significantly unrelated). Unlike findings from Kim et al. (2022) who found higher intercorrelations for unfamiliar faces, our results indicate that representations or the perceptual clustering of *familiar* faces based on certain social traits may rely on increasingly similar social information to each other. One consideration, however, is that different groups of familiarised faces might have different representational structures. For example, representations of traits might be similar based on the faces from The Apprentice due to the specific traits the participants learned to associate with other traits, which might have produced a more highly intercorrelated set of face representations based on *these* particular traits. For example, perceived dominance and likeability might be significantly associated based on the familiarised face identities from the show as the show was about becoming the next apprentice, where dominance (at least pertaining to competence) may be perceived as a good thing. However, in other contexts, these associations may not exist in the same way if dominance is perceived to be an unlikeable trait (e.g., based on a show where submissiveness is the goal). This may point to the possibility that the representational similarity between certain traits might in fact depend on the social context. Future studies could explore this research aim with other shows to see whether trait associations post-familiarisation would be distinct from that of the present study. Nevertheless, findings from the present analysis indeed indicate that the representational similarity between traits can change with familiarisation, even in as quickly as 2-3 weeks.

The empirical study in Chapter 3 has implications for existing theoretical models of face perception. Our findings suggest that we may rely on distinct face information when perceiving a familiarised face to that of when we were yet unfamiliar with them. For example, face perception models (Oosterhof & Todorov, 2008; Sutherland et al., 2013) have demonstrated that faces that are smiling are more likely to be perceived as trustworthy. However, our results showed that the representations of faces based on trustworthiness judgments made after participants got to know a set of individuals were distinct from that of the faces of the same people represented before familiarisation. One explanation could be that people may have relied on distinct face information in order to make the call on, for instance, trustworthiness (even if the same characteristics were perceived differently among individuals), but that this may have become a less reliable source of information once participants became familiar with patterns of behaviours, expressed thoughts and feelings, all of which may be indicative of a trait (John et al., 2010). This points to the possibility that the dimensionality of perceived trustworthiness in familiar faces might relate to either a different sets of traits (i.e., our analysis suggests likeability), or additional ones not tested it the present study (e.g., responsibility, consistency, integrity, faithfulness, etc.). Thus, as previously addressed, the traits that constitute trustworthiness as a foundational dimension might differ between unfamiliar and familiar faces. This line of reasoning can also be assigned to perceived dominance. However, since there may be much variability in dominance perception, it may be more difficult to ascertain representational commonalities in familiar faces.

These findings also have practical implications. Our results suggest that our perceptions of people, particularly those we initially trust and perceive to be dominant, are likely to change, and in as quickly as 2 weeks. Thus, relying on initial face judgements based on certain traits may not provide observers with the whole picture when deciding how to behave towards them, especially for momentous decisions. It is important to note, however, that participants in the study had the opportunity to observe behaviours and opinions expressed by individuals in the show over a

shorter period of time that would probably have been expressed over an extended period in reality. Thus, the way we represent others may depend more on the type of information learned about a person within this timeframe rather than the timing of familiarisation itself.

6.1.1.2 Chapter 4: brain representations of faces change with familiarisation

Our second study set out to investigate the ways in which brain representations of faces might change with familiarisation. In Chapter 4, we used RSA to assess the representational structure of faces in the three core face-selective regions and how these might change as familiarity increases. While the present analyses do not allow for ascertaining the explicit information or function of which these regions process familiar faces, it may shed light on the nature of these representations and how they might change in various ways. Participants in the present study also took part in the behavioural study in Chapter 3, and the procedure as well as the analyses were similar. Participants watched a TV series over 2-3 weeks. Both before and after watching the series, participants underwent neuroimaging with fMRI whilst they were presented with dynamic videos of the face identities in order to obtain brain activity responses to these faces stimuli. Thus, we aimed to assess the representational structure of faces in the brain, and had three main aims: a) to compare face representations in each of the core regions (FFA, OFA and pSTS) within participants across sessions, to see if the way that these regions organise faces changes, b) measure the representational similarity of faces in the brain between participants, for each session and each region, and compare this similarity across sessions. We were interested in how similarly the brain regions would organise faces across participants and whether this similarity might increase across sessions, given that participants had access to the same faces and same social information. Finally, we measured the representational similarity of faces between each brain region, both for unfamiliar and the familiarised faces, and then compared this similarity across sessions. Our aim was to understand more about how similarly these brain regions organise face information and how this might change over time.

Results for the first analysis show that the representational structure of faces in all three regions were not significantly correlated, which may be indicative of substantial change with familiarisation, and Bayesian analyses show that this is more likely to be the case than otherwise. This may suggest that face representations in core face regions change the way that they organise faces. Previous research shows differential activity in broader areas implicated in familiar face perception, containing the face-selective regions (Ambrus et al., 2021) and even the ROIs more directly (Natu & O'Toole, 2011; Natu & O'Toole, 2015; Bayer et al., 2021; Visconti di Castello et al., 2017, 2021). The present analysis shows that the face-selective ROIs may change the way in which they store or propagate information about faces (to other regions). Research shows that the OFA in particular is implicated in more basic face features (Gauthier et al., 2000; Kietzmann et al., 2015; Tsantani et al., 2021), and given that this region appears to change the way it represents faces with familiarisation, it could be deduced that the OFA either changes the way it clusters faces (i.e., based on a different set of physical features) or even becomes implicated in distinguishing between faces based on an additional type of information (i.e., becomes implicated in both lower- and higher-level abstract information during familiar face perception). Eick et al. (2020) show that the OFA has also been found to be implicated in semantic information when viewing familiar faces. Similar conclusions can be drawn from the results showing that both the FFA and pSTS also changed in terms of the representational structure, thus, findings reveal the possibility that the nature of the information by which these core face-selective ROIs cluster faces appear to change after familiarisation.

The second analysis observed similarities in the representational structure of the core regions across participants and found significant similarity for OFA and pSTS representations separately, pre-familiarisation. This inter-subject agreement only significantly changed with familiarisation, however, for the OFA, which became significantly more idiosyncratic. Even though participants were familiarised with the same faces and had access to the same social and personal information, brain representations appeared to change in separate directions across

participants in this region, highlighting the possibility that the way we interpret the same social and conceptual information, based on the information implicated in these regions, can become increasingly distinct just over a short period of time. This includes the FFA, where initial representations were not significantly similar across participants to begin with, and remained dissimilar across sessions.

A particularly important finding involves the representational changes within the OFA: previous findings consistently show that this region is implicated in processing image-based properties (Gauthier et al., 2000; Kietzmann et al., 2015; Tsantani et al., 2021, for a review, see Pitcher et al., 2011), but our findings are also in line with that of Eick et al. (2020), suggesting the potential for non-visual information to be computed in this region, which, based on our results, may be specifically implicated in familiar face processing. Additionally, Rossion et al. (2003) demonstrated that feedback processes via the OFA involve familiar face representations, as a patient with lesions to the rOFA had major deficits in recognising famous faces. Together, these findings also have implications for models suggesting that these core ROIs are involved in hierarchical processing (Haxby et al., 2000) when, in fact, they may be implicated in more distributed or recurrent-level processes (Eick et al., 2020; Duchaine & Yovel, 2015; Pitcher & Ungerleider, 2021). Duchaine and Yovel (2015) suggest that there are multiple pathways into the visual face processing system, and that one of these pathways may involve back propagation from the pSTS to the OFA. This back propagation might carry relevant abstract identity-related information to the OFA for the refinement and consolidation of face representations, be they visual or non-visual. As these familiar face representations are likely to be idiosyncratic, as per our second analyses and previous findings (Mileva et al., 2019; Kim et al., 2022), at least part of this information might be encoded in the OFA. Put simply, the OFA might primarily encode image-based properties in unfamiliar faces but may later encode both image-based as well as abstract face-related information, information of which may have been communicated from other relevant face regions implicated in processing similar information (i.e., the pSTS). It could be

that at least part of the OFA's role is to match incoming encoded image-based face properties with pre-stored abstract representations about the same face identity.

Importantly, the present study did not directly test connectivity between regions and, thus, results from the present study may only provide a rationale for exploring this in future research. Nevertheless, these findings collectively challenge previous models of the roles of these ROIs in unfamiliar and familiar face perception (Haxby et al., 2000; Gobbini & Haxby, 2007).

The final analysis in Chapter 4 measured the representational similarity between each ROI. Results showed that the FFA and OFA were significantly associated both pre- and post-familiarisation. This is in line with previous findings showing that the OFA and FFA share similar functional properties for unfamiliar faces (Gauthier et al., 2000; Haxby et al., 2000; Tsantani et al., 2021) and our findings suggest that these similarities may not significantly change, at least within 2 weeks of becoming familiar with someone. Relations between other ROI pairs in either of the sessions were non-significant, however, post-familiar representations in the OFA and pSTS trended towards significance. This could suggest that any relations are either due to chance, or that more time was required to find an effect. The latter is likely as correlations changed from around 0 to above .1. It could be the case, therefore, that representations in these regions were becoming similar.

Non-significant associations between most ROI pairs in each session indicate the extent to which they play distinct roles in face perception, even after familiarisation. While there appears to be shared implications in unfamiliar and familiar face processing, particularly between the OFA and FFA, these regions largely organise faces based on distinct characteristics and/or social and personal information. In other words, these ROIs mostly utilise similar social information in distinct ways in order to perceptually encode a face, both before and after familiarisation. Our second analysis revealed that representations particularly between the OFA and pSTS at least trended towards idiosyncrasy across participants, but based on the third analysis, it appears that

the type of information implicated in these regions might eventually become more similar, even if the way their minds perceive the faces might be different. In other words, people may contextualise the same familiar faces distinctly but use the same type of information to do this (i.e., eye gaze patterns, facial expressions, facial viewpoint; Engell & Haxby, 2007; Kietzmann et al., 2015). This may explain why, from our own experiences, we may interpret similar behaviours in different ways. For example, one person may perceive a certain facial expression to be indicative of tiredness, while another might perceive the same expression to indicate indifference or boredom.

In sum, face representations in the brain mostly appear to change as we become familiar with people. Our brains may re-organise the way it contextualises faces based on abstract and conceptual knowledge about the people we are getting to know (Schwartz & Yovel, 2016; Kim et al., 2022; Mileva et al., 2019), and even though we may use the same information to contextualise faces, we may still interpret the information extracted from the face in slightly different ways, which may even include trait information. As previously mentioned, explicit judgements about these faces cannot be interpreted from the representations in the present study. Therefore, future research could attempt to match the social trait representations to the brain representations of faces in the core face ROIs (Tsantani et al., 2021), and the dynamics thereof, as they form during familiarisation. More generally, findings from Chapter 4 may also reinstate the longstanding premise that the core ROIs are selectively involved in face processing. Representational re-organisation of faces indicates a direct involvement in the processing of familiar faces, based on enriched person information extracted from the face. In saying this, it is important to directly compare familiar face representations with that of familiar object representations for completeness.

6.1.1.3 Chapter 5: Dyadic Person Similarity Predicts Similarity in Face Judgements

In Chapter 5, we investigated whether dyadic person similarity predicts similarity in face judgements. Research in face perception naturally involves the exploration of processes and mechanisms that are common among people, but there has also been much investigation into the individual variance in face perception, which can be accounted for by observer-based characteristics including personality (Hönekopp, 2006; Hehman et al., 2017; Mattarrozi et al., 2015; Oh et al., 2022), cognitive traits (Willis et al., 2013), and brain activity patterns (Parkinson et al., 2018; Finn et al., 2020; Matz et al., 2022). Oh et al. (2022) demonstrated that people with more similar personality and social trait dispositions also tend to perceive faces in a similar way. However, this study was conducted at the regional level, and findings show that this relationship varies by country. Thus, our aim was to investigate this relationship within a single world region (i.e., the United Kingdom) and assess the extent to which person similarity predicts face judgement similarity at the dyadic level.

Three-hundred and seven participants rated various faces on 6 social traits including trustworthiness, dominance, attractiveness, responsibility, aggressiveness and intelligence. Participants then rated themselves based on these same traits as well as an additional 7 for completeness (including meanness, sociability, emotional stability, care, unhappiness, weirdness and confidence), and also 60 items of the Big 5 personality traits (i.e., neuroticism, agreeableness, openness, extraversion and conscientiousness; McCrae & Costa, 1985). We used inter-subject RSA (IS-RSA) to assess the similarity between the structure of dispositions and the structure of social trait judgements of faces. We measured this in the form of pairwise similarity between participants based on personality and social trait dispositions, and the respective pairwise similarity in their judgements of faces.

Results showed that there was a significant association between the structures, namely that those who perceived themselves in similar ways based on these two types of dispositions also perceived others in a similar way. These results do not inform us about whether there is an element of truth to either the judgements of their own self nor the judgement of the faces, but

informs us about the perception of the two. This might shed light on the possibility that those who experience their internal world in similar ways may experience their external social world in similar ways. Results also showed that this was the case irrespective of similarities in ethnicity, gender, age, and geographical region of residence. Importantly, similarities in these variables also explained similarities in face trait judgements, but this variance was separate to that of personality and social trait disposition similarities. In other words, people of similar ethnicities, those nearer to each other in geographical space, closer in age, and of the same gender also appear to perceive people more similarly to those they are not similar to in these domains, but these similarities are not significantly related to that of personality nor social trait dispositions. We also broke the analyses down and assessed associations between the structures based on separate traits and only found that those who rated themselves similarly in the dispositions and personality also tended to perceive similar people to be attractive. This may indicate that similarities in personality and social dispositions may dictate social judgements about faces more holistically (i.e., all 6 traits and possibly others not tested in the present study). Instead of assessing personality structure by looking at trait associations (i.e., pairwise associations between traits agreeableness and conscientiousness, as in Oh et al., 2022), we assessed traits as a whole structure (i.e., pairwise similarity between participants based on all personality traits and all social traits). We also assessed pairwise similarity against separate face trait judgements to identify which of the social traits might show higher correlations with similarities in the dispositions. Results showed that the effect of person similarity on similarities in perceived attractiveness was strongest (i.e., correlations between the two were higher than for the other face traits). Therefore, those who are more similar are more likely to perceive the same people as attractive than the people they both perceive to be trustworthy, and so on. Additionally, we did not look at average effects across participants but rather assessed the pairwise similarity between the participants so as to measure the idiosyncrasies in the relationship.

Findings from the present study may reveal that variance in our social trait perceptions (Sutherland et al., 2020) may indeed relate to our personality and social trait dispositions, particularly at the dyadic level. Even though we appear to make judgements about others that pertain to common factors such as whether a person is approachable or potentially harmful, or even a suitable mate (Oosterhof & Todorov, 2008; Sutherland et al., 2013), some of the similarities between our judgements can be attributed to similarities between our own dispositions. Previous research already demonstrates that the way we perceive ourselves directly relates to the way we perceive others (Hönekopp, 2006; Hehman et al., 2017), but what did this look like as a structure (Oh et al., 2022) at the most idiosyncratic level with increased heterogeneity? Our results essentially show that there is an underlying structure to personality and trait dispositions, and while associations between these structures aren't strongly related to that of face trait judgements, they are still meaningfully related and can explain some of the variance that exists in our judgements of others. The other side of the coin is that there is still much individual variance within face perception that is unaccounted for by similarity in the variables used in the present study. Thus, our results indicate that there are likely other variables beyond the ones tested here that may account for similarities in face trait judgements.

Future research could extend these findings to address other types of dispositions, such as similarities in childhood experiences and upbringing, culture (Sofer et al., 2017), and other variables that might shape our perception. For example, increased anxiety has been associated with lower perceived trustworthiness (Mattarozzi et al., 2015), and this may in turn be related to adverse childhood experiences which give rise to anxiety later on in early adulthood (Elmore & Crouch, 2020).

Moreover, our findings may shed light on the basis for how we might gravitate towards each other and form relationships. Part of the reason we may form relationships with others might be due to similarities in the perception of our social world. Parkinson et al. (2018) show that pairs of friends have similar brains and likely perceive the world in similar ways (i.e., based on media

content of various types, including documentaries, comedy, and politics), and Youyou et al. (2017) demonstrate that romantic partners and friends have similar personalities, ranging between correlations of .20-.47 and .12-.31 respectively. Is it then the case that people who are already more similar go on to become friends, or do they become similar as they become friends? Findings from Selfhout et al. (2010) revealed that people actually choose who they'd like to be friends with based on similarities in agreeableness, extraversion and sociability, and very recent work demonstrates that people with similar neural chemistry develop a closer friendship over an 8-month period (Shen et al., 2025). It is even possible that proportions of the variance can be explained by both, though future research could address these questions more directly. Our results also showed that those of the same ethnicity, gender, those nearer to each other in geographical space, and closer in age, also perceive people in similar ways based on the 6 social traits. This reveals the extent to which our environment (Oh et al., 2022), cultural influences (Sofer et al., 2017), and life stage (Lin et al., 2016) may shape our personalities and, to a meaningful extent, our perceptions of others, which may form the basis for these relationships.

6.2. Overall conclusions, limitations and future directions

6.2.1 Overall conclusions

The present thesis set out to investigate various factors that influence the way we perceive faces. We explored the overarching themes of familiarity and how representations of faces change as we get to know them, and were interested in explicit social trait representations including fundamental social dimensions (Oosterhof & Todorov, 2008; Sutherland et al., 2013) that serve important functions for our interactions and the ways in which we behave towards others (Berkowitz & Frodi, 1979; Wilson & Rule, 2015; Todorov et al., 2005). We found that representations of trustworthiness and dominance appear to change within 2-3 weeks of familiarisation, and that trustworthiness representations in particular become more synchronised across those who are getting to know the same people (Chapter 3). Additionally, the later development of trustworthiness representations can be approximated by initial liking, where an

initial instinct on whether a person likes or dislikes a person could be used as a function to later go on to trust them. We also see that representations of the underlying substrates of face perception also appear to change, and our results may provide evidence for the occurrence of interpreting the same actions and behaviours in different ways (Chapter 4). Our finding regarding the idiosyncratic changes to the OFA specifically may also challenge previous models of face perception in the brain (Haxby et al., 2000; Gobbini & Haxby, 2007; Eick et al., 2020). Finally, we found that inter-subject person similarities (measured with personality and self-report trait questionnaires) are related to similarities in face judgements (Chapter 5). In other words, pairs of people who are similar in their personality appear to make similar judgements about other people's faces. This could have implications for relationship formation and other important dyadic situations, namely that we might gravitate towards those we are similar to as we seem to perceive our social world in a similar way (Matz et al., 2022).

Our empirical studies attempted to explore factors that influence face perception in two main domains: a) face familiarisation, including both behavioural reports of face traits as well as representations in face-selective regions in the brain, and b) the individual differences in our face trait judgements. According to our findings across all empirical chapters, these two factors do appear to influence the ways in which we perceive a face, particularly the structure of our face representations.

All empirical studies in the present thesis utilised RSA, which has been a useful method in ascertaining representations at the second-level of analysis (Kriegeskorte et al., 2008a). This allowed for capturing the underlying structure of these face representations and how they might manifest in a given context (i.e., face trait perception, brain representations of faces, similarities in face judgements, how face representations might change over time). Instead of comparing analysis in a univariate manner (i.e., with means or standard deviations based on individual faces), we are able to assess how faces are represented together as a whole construct and how this might change, providing a richer contextual examination of these faces based on traits and

as measured in the brain. Thus, we are able to assess how faces are collectively *organised* in perception rather than how they are merely perceived in isolation from each other.

6.2.2. Limitations

While our findings may have theoretical and practical implications, some limitations of the studies should be considered. Firstly, there is a general consensus that fMRI studies require fewer participants, however, recent studies still advocate for higher numbers (Turner et al., 2018), especially for behavioural testing. A larger sample size may have better represented the population, as the final sample of our study consisted of only 26 participants for our study in Chapter 3 and 29 in Chapter 4. Future studies could use a power analysis and potentially test more participants to examine whether the change in social traits would be reflected in similar ways to the present study.

Additionally, both Chapters 3 and 4 utilised the same paradigm, where participants were familiarised with a set of face identities through para-social or contextual familiarisation (Kovács, 2020; Ambrus et al., 2021). Research shows that there are distinctions in the amplitude of neural activity when comparing para-social familiarisation with in-person familiarisation, where even just 3 hours of the latter (versus 2 weeks of the former) produced more enriched representations as measured by more enhanced neural activity (Kovács, 2020; Ambrus et al., 2021). While there may still be significant neural activity in the brain during para-social familiarisation compared to unfamiliar faces, and faces that have been merely repeatedly presented to people (Ambrus et al., 2021), the nature and/or type of information implicated in the core face regions might differ between para-social and in-person familiarisation. Thus, our findings may be restricted to this type of familiarity and further testing should examine changes in representations with in-person familiarisation paradigms (Kovács, 2020; Ambrus et al., 2021).

When testing familiarity, we also did not statistically measure whether participants actually recognised the face identities at the end of the two weeks of familiarisation. While our aim was

not to assess representations at the 'end' or optimal point of the familiarisation process in and of itself (Popova & Wiese, 2023) as we were particularly interested in how representations evolve during the *process*, it would have been useful to assess this change once all participants were (at least statistically close to) 100% recognition accuracy, for completeness. Upon visual examination, our data (not reported) showed that most people recognised the identities by the end of the two weeks, and trivia questions towards the end of the final week were correct across most participants. In any case, testing for recognition accuracy will further aid in contextualising the results from the present study.

In Chapter 4, we only examined face-selective regions in the core visual network (Haxby et al., 2000; Gobbini & Haxby, 2007; Kovács, 2020). While we aimed to assess the representations that were primarily visual in nature, and while implications might be generalisable to other regions implicated in higher-level face processing (i.e., in the proposed extended system; Gobbini & Haxby, 2007), potential changes in the representational structure of these regions should be examined also. One possibility would be to carry out searchlight analysis to find out which regions show significant changes with familiarisation.

We also do not yet know what information is encoded in the representations of these regions. Previous research informs about the likely functions of the face-selective regions, as previously discussed, but the actual information extracted from the face identities in the show were not ascertained and compared across sessions. Results from Chapter 4 naturally only tell us about the way the re-organisation of face representations occurs with familiarisation, and only broader conclusions about the *nature* of these representations, namely whether they might be based on more visual or more abstract conceptual information (Bruce & Young, 1986; Gobbini & Haxby, 2007; Kovács, 2020; Schwartz & Yovel, 2016). As previously studies demonstrate that core brain regions are implicated in relative image-based and abstract information extraction from faces (Gauthier et al., 2000; , these functions might evolve during familiarisation.

There are also a few limitations to the empirical study in Chapter 5. First, we attempted to examine the personality structure between participants who were strangers, however, we did not control for participants knowing each other. There was no apparent way to do this as random participants were recruited online from all over the U.K., though, future studies could attempt to ensure that participants do not know each other with ethical and innovative means. The present study could have also used a variety of social face traits, personality traits, and social trait dispositions to better capture patterns within the structure of dyadic person similarity and respective face judgement similarity. Presently, conclusions are restricted to the set of stimuli that we used in the study.

Face perception models and studies demonstrate that using a highly varied set of naturalistic face images yields more rich representations associated with faces (Jenkins et al., 2011; Sutherland et al. 2013). While we made attempts to include varied stimuli of face identities, these were still presented in the form of photos and videos across all studies. In line with a point made previously, in-person paradigms where natural variation in face presentation and face dynamics occur might allow for capturing more enriched representations (Kovács, 2020; Ambrus et al., 2021). Testing of the effects of familiarisation should also be carried out over a potentially optimal period of 8 months (Popova & Wiese, 2023). Popova and Wiese (2023) found that in-person familiarisation with others increases visual-based and abstract-level familiarity representations in the brain up to around 8 months, where neither additional contact, increased frequency of contact, nor additional person-related information would amplify this activity. This suggests that 8 months is an optimal time period for becoming familiar with someone. Representations of social traits in faces can be examined and compared across sessions over this time period to see if findings might differ from those of Chapters 3 and 4.

6.2.3 Future directions

Future research could attempt to derive underlying social dimensions for familiar faces (Oosterhof & Todorov, 2008), where free-recall descriptions can be made of faces both before

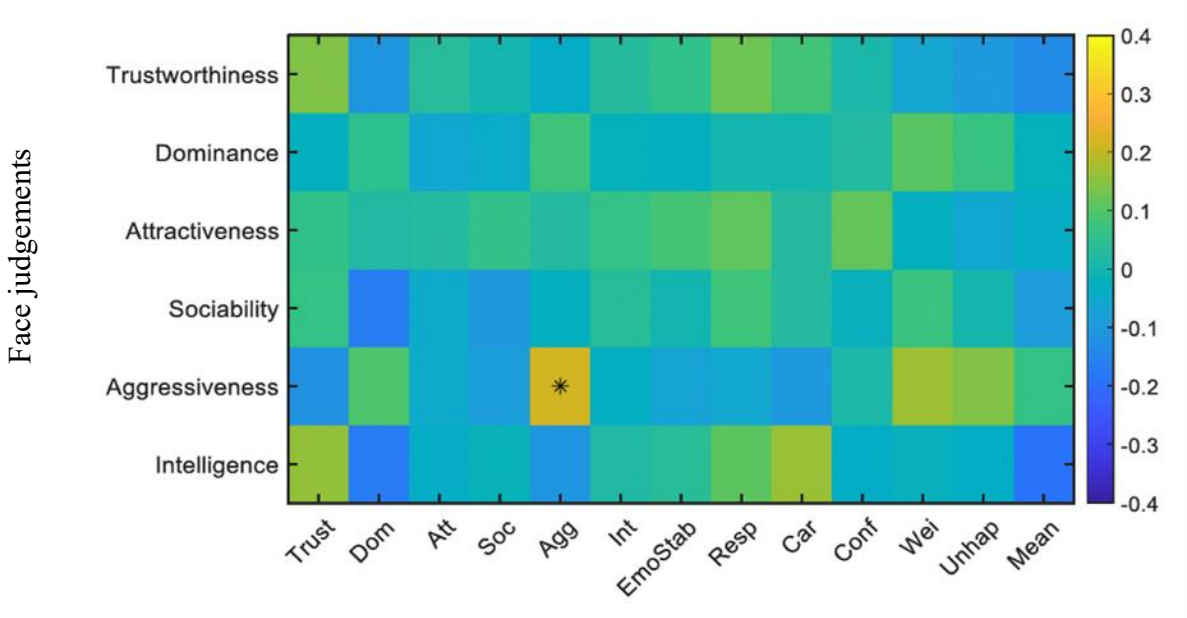
and after familiarisation. It could be the case that descriptions, and the underlying dimensionality thereof, between unfamiliar and familiar faces might differ, as ample evidence suggests the two are representationally distinct and should not be conflated in research (Bruce & Young, 1986; Schwartz & Yovel, 2016; Ramon & Gobbini, 2018; Kramer et al., 2018; Mileva et al., 2019; Kim et al., 2022). The dimensionality of these responses can then be reduced at the individual level to derive independent models of face perception (Sutherland et al., 2020) as there appears to be highly idiosyncratic representations across participants. These individual models can then be compared to search for any consistency in even the *type* or *nature* of information that people use to make these judgements between unfamiliar and familiarised faces, even if the specific descriptive words/terms themselves are unique across participants. For example, our findings in Chapter 3 (Analysis D) suggest that trustworthiness as a trait may *become* likeability, or representations of trustworthiness become highly similar to that of likeability. This suggests that the function of perceived trustworthiness might change as the relationship between people changes with familiarisation. It could be that people may no longer rely on whether a person's face shows that they have good intentions but rather based on whether they like who the person is, or their character, after acquiring new knowledge about their person.

Future studies could also use a much larger sample that examines changes in trustworthiness representations over a longer period of getting to know an even larger set of identities. With this, researchers can a) assess the similarity in representations across people over a longer period to see at what point these become stable, and b) ascertain explicit behavioural reports on which characteristics people have come to rely on to make the judgements about who is or is not trustworthy. It could be the case that more stable characteristics ascertained after a longer period of familiarisation might serve as functions of trust longitudinally, thus, learning what these characteristics might be may better aid our understanding on the functions of trustworthiness and the dynamics thereof.

Additionally, people who perceive others to be dominant may also perceive themselves to be submissive in comparison. Oosterhof and Todorov (2008) stipulate that dominance generally relates to the perception of potential to harm (the observer); this, by nature, is polarising and would require one to be able to 'overthrow' the other in some sense, naturally requiring one to lean toward one end of the dominance dimension while the other leans toward the other end. This might be different from perceived trustworthiness, where the observer and target can both perceive themselves and the other to be trustworthy, potentially even causing them to approach each other for this same reason. However, it could be the case that perceiving dominance in another might provoke a sense of the opposite (i.e., submissiveness) in the observer if they perceive themselves to be weaker or not as dominant as the other. Analyses from Chapter 5 may provide some insight into this possibility (see **Figure 29**). The graph shows that those who perceive themselves to be more dominant also perceive others to be dominant, though this correlation is in the opposite direction of what was predicted above and not significant. In fact, variance relating to physical harm may not polarise as those who perceived themselves to be higher in aggressiveness also perceived others to be higher in aggressiveness. Additionally, from looking at the scatterplot of this correlation to be able to see individual data, (see **Figure 30**), there is no indication of any relation between self-perceived dominance and perceived dominance in others.

Figure 29

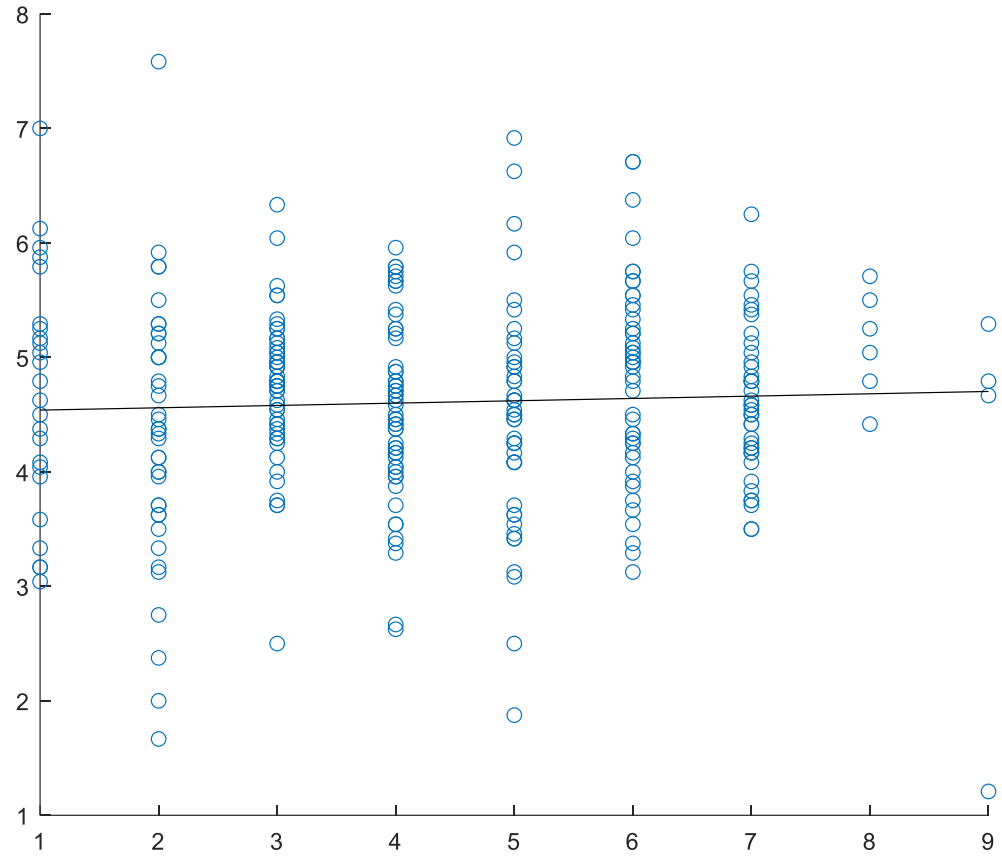
Individual differences in self- and other-judgements: correlations between average self-perceived trait ratings and averaged face trait judgements.



Note. Correlation heatmap showing correlations between each self-perceived social trait and each face trait judgement.

Figure 30

Scatter plot: Correlations between self-perceived dominance and face dominance judgements.



Note. Scatter plot of correlations for each participant between self-perceived dominance (on a Likert scale of 1-9) and perceived dominance in faces (on a scale of 1-7).

Previous research, however, has indicated correlations in self-perceived dominance and others' physical associations of perceived dominance, but only in men (i.e., face width-to-height ratio; fWHR; Mileva et al., 2014). Thus, variance in dominance ratings may relate to gender, and findings from Chapter 5 may require the disentanglement of self-other ratings between men and women.

In future studies, trait representations of familiarised faces could also be compared between faces that participants have become familiarised with in different contexts. For example, multiple sets of participants could watch shows based on various topics (i.e., Business-based, relationship-based, gameshows about trust), and the variability in the relationships between trait representations could be measured and compared between conditions to understand whether the specific pattern of representational similarity found in our study (i.e., Analysis D from Chapter 3) is generalisable across these contexts. In other words, would face familiarisation based on other contexts lead to representational trait similarities akin to that found in the Chapter 3?

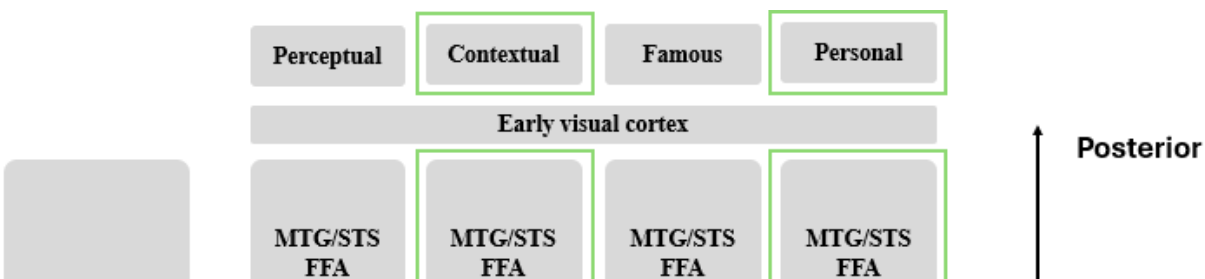
Research extending the investigation of representations of familiar faces for social traits as well as in the brain could be examined with an in-person paradigm to assess whether the organisational structure of more enriched representations might differ to the results found in Chapters 3 and 4. Research shows that para-social familiarisation significantly differs from in-person familiarisation in terms of brain activity (Kovács, 2020; Ambrus et al., 2021, see **Figure 31**), suggesting the potential for other types of information being implicated in these face-selective regions, or even the extent to which they are implicated. It appears to be the case that three additional regions (and thus, the entire face perception network) are implicated in personal familiarity compared to contextual familiarity (i.e., IFG, amygdala, and insula). While we were interested in face representations in core visual areas (which naturally appear to be implicated in both contextual and personal familiarisation, see **Figure 31**), it could be the case that representations in these core regions become more enriched for personal familiarisation due to top-down input from additional anterior regions, resulting in a distinct representational structure of faces.

Moreover, RSA could be used to assess the representational similarity between explicit behavioural measures and these regions, and the subsequent change in these similarities. Thus,

the content of the information implicated in these core regions might be ascertained for both pre- and post-familiarised faces, and the changes between them can be assessed. For example, social trait representations could be correlated with that of core brain representations in both sessions and then compared across sessions to assess the type of information that might (or might not) be implicated in these regions as we get to know someone. It was important to assess the representational structure of the face traits and brain representations separately, as they are domains within their own right and may thus naturally be implicated in face perception in distinct ways. However, a sensible next step would be to compare representations across the domains (Kriegeskorte et al., 2008a; Kriegeskorte et al., 2008b). For example, Tsantani et al. (2021) found that the representational similarity between the OFA and trait attractiveness were significant, while representations in the FFA were similar to that of perceived face similarity and perceived identity. Future research could extend this work to assess how these representational similarities might change with familiarisation.

Figure 31

Familiar face perception model by Kovács (2020)



Note. Neuroscientific familiar face perception model by Kovács (2020) depicting distinctions between perceptual, contextual, famous and personal familiarity. Green rectangles indicate suggested comparison to ascertain whether representations in core regions (FFA, OFA and pSTS) would become more enriched for personally familiarised faces due to top-down processes from additional regions, compared to that of contextual familiarisation. Double-headed arrow indicates the horizontal plane of the brain from posterior (back) to anterior (front).

Additionally, results for representational similarity between core ROIs in the present study may be indicative of connectivity between the regions, where the ROIs co-operate to encode a face. Thus, our findings may provide support for proposals from previous models suggesting the OFA and FFA are connected via feedforward processes during face perception (Haxby et al., 2000), even for familiar face processing, where back propagation of higher-level may also occur (Gobbini & Haxby, 2007). Findings may also point towards potential eventual connectivity

between the OFA and pSTS. Duchaine & Yovel (2015) propose that these regions may be connected via a visual pathways implicated in familiar face perception. Therefore, studies could explore regional connectivity more directly between the core face ROIs based on the representational structure of increasingly familiar faces.

Lastly, investigations for dyadic person similarity and relative similarities in face perception could be examined with an extended set of personality trait items, social trait dispositions and face traits, but can also include other types of dispositions that appear to shape our judgements. This may include upbringing and culture (Sofer et al., 2017), particularly when considering ethnic similarities, clinical conditions (Willis et al., 2013) such as schizophrenia, one of which has typically been difficult to classify due to the heterogeneity (and internal experience) of symptoms (Fang et al., 2022). IS-RSA would be a useful way to assess patterns of similarity where there is increased heterogeneity, capturing the structure of highly idiosyncratic patterns of internal cognitive and perceptual experiences (i.e., unique patterns and combinations of positive, negative [Crow, 1982], and cognitive symptoms [Beck & Rector, 2002]). Dyadic brain similarities among those with schizophrenia (Fang et al., 2022) and threat detection in faces, for instance (Huang et al., 2011) could be compared to that of the neurotypical population. These studies would aid our understanding of extended underlying structures that may contribute to the variance in face perception, including that of both neurotypical and neurodivergent populations with higher heterogeneity. This may in turn have implications for healthcare and treatment.

Studying dyadic friendship formation may also provide insights for the extent to which people who gravitate towards each other are either already similar or become similar based on their dispositions. Participants from a first-year undergraduate cohort could be tested whilst they are strangers, and then as they continue to attend university throughout the three years and as the trajectory of any friendships are formed, it can be assessed as well as their similarity through self-reports of their behavioural dispositions and similarities in brain representations. This investigation would tie together all facets of the present thesis: we would learn about whether

inter-subject similarity in the form of both behavioural and brain dispositions (Matz et al., 2022; Finn et al., 2020; Parkinson et al., 2018; Shen et al., 2025) predicts friendship formation. Specifically, we could measure whether explicit face trait representations *as well as* brain representations change in a similar direction between these dyads, as they become familiar with each other in-person (Kovács, 2020; Ambrus et al., 2021) and potentially develop a friendship (Parkinson et al., 2018; Shen et al., 2025). This could be tested over a sufficient period which includes 8 months (Shen et al., 2025), based on the frequency, recency and closeness of contact (Popova & Wiese, 2022).

6.3 Conclusion

To conclude, the present thesis presents three empirical studies that investigated factors that moderate face perception in humans, including familiarity (from a behavioural as well as a neuroscientific point of view), and individual differences, namely the idiosyncratic structure of personality and social dispositions on face trait perception. Essentially, the thesis has been concerned with the way we represent faces, and (IS-)RSA has allowed for assessing the representational structure of faces in relevant contexts. When addressing the effects of familiarisation, we were interested in how face representations change as we become familiar with someone. We were interested in the explicit perceptions of social traits in faces, as well as the underlying substrates from which face perception arises, investigated as a separate domain. We found that both social trait representations and brain representations of faces do indeed change as we become familiar with others, and that these changes begin to happen within a 2-3 week period. We then turned the focus to individual differences in face perception, which is an important aspect considering that not all face information is perceived in the same way across people. We found that pairs of people who have similar personality and social trait characteristics also tend to perceive others in a similar way, and this association is meaningful, however, there remains much individual variance unaccounted for by the variables used in the present thesis.

These findings demonstrate that different factors influence the way we perceptually organise faces, whether they change with time or are influenced by our own characteristics.

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological review*, *128*(2), 290.

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of personality and social psychology*, *93*(5), 751.

Akdeniz, G., Toker, S., & Atli, I. (2018). Neural mechanisms underlying visual pareidolia processing: an fMRI study. *Pakistan journal of medical sciences*, *34*(6), 1560.

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of personality and social psychology*, *55*(3), 387.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological monographs*, *47*(1), i.

Allport, G. W. [1954] (1966). *The Nature of Prejudice.* Reading, MA: Addison-Wesley.

Ambrus, G. G., Eick, C. M., Kaiser, D., & Kovács, G. (2021). Getting to know you: Emerging neural representations during face familiarization. *Journal of Neuroscience*, *41*(26), 5687-5698.

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage*, *23*(3), 905-913.

Archer, J. (1988). The behavioural biology of aggression (Vol. 1). Cup Archive.

Asch, S. E. (1946). Forming impressions of personality. The Journal of Abnormal and Social Psychology, 41(3), 258–290.

Baek, E., Hyon, R., López, K., Du, M., Porter, M., & Parkinson, C. (2023). Lonely individuals process the world in idiosyncratic ways. *Psychological Science, 34(6)*, 683-695.

Banse, R. (2001). Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. Cognition & Emotion, 15(4), 501-520.

Baran, F. D. E., & Cetin, M. (2025). AI-driven early diagnosis of specific mental disorders: a comprehensive study. *Cognitive Neurodynamics*, *19*(1), 70.

Barr, A. (2003). Trust and expected trustworthiness: experimental evidence from Zimbabwean villages. *The Economic Journal*, *113*(489), 614-630.

Barragan-Jason, G., Cauchoix, M., & Barbeau, E. J. (2015). The neural speed of familiar face recognition. *Neuropsychologia*, *75*, 390-401.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1)*, 1-48.

Batres, C., & Shiramizu, V. (2023). Examining the "attractiveness halo effect" across cultures. Current Psychology, 42(29), 25515-25519.

Bayer, M., Berhe, O., Dziobek, I., & Johnstone, T. (2021). Rapid neural representations of personally relevant faces. *Cerebral Cortex*, *31*(10), 4699-4708.

Beck, A. T., & Rector, N. A. (2002). Delusions: A cognitive perspective. *Journal of Cognitive Psychotherapy*, *16*(4), 455-468.

Berkowitz, L., & Frodi, A. (1979). Reactions to a child's mistakes as affected by her/his looks and speech. *Social Psychology Quarterly*, 420-425.

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

Bible Hub. (2025). Ethiopic Language. https://biblehub.com/topical/e/ethiopic_language.htm#:~:text=While%20the%20Bible%20does%20not,Cultural%20Legacy.

Bilalić, M. (2016). Revisiting the role of the fusiform face area in expertise. *Journal of cognitive neuroscience*, *28*(9), 1345-1357.

Bond, Jr., C. F., Berry, D. S., & Omar, A. (1994). The Kernel of Truth in Judgments of Deceptiveness. *Basic and Applied Social Psychology*, *15*(4), 523–534.

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychological bulletin*, *106*(2), 265.

Broche-Pérez, Y., Jiménez, L. H., & Omar-Martínez, E. (2016). Neural substrates of decision-making. *Neurología (English Edition)*, *31*(5), 319-325.

Bronstad, M., & Russell, R. (2007). Beauty is in the 'we' of the beholder: Greater agreement on facial attractiveness among close relations. *Perception*, *36*, 1674-1681.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77*(3), 305-327.

Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, *2*(1), 1-13.

Bukowski, H., Dricot, L., Hanseeuw, B., & Rossion, B. (2013). Cerebral lateralization of face-sensitive areas in left-handers: only the FFA does not get it right. *Cortex*, *49*(9), 2583-2589.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive psychology*, *51*(3), 256-284.

Campbell, R., Heywood, C. A., Cowey, A., Regard, M., & Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia*, *28*(11), 1123-1142.

Candidi, M., Stienen, B. M., Aglioti, S. M., & de Gelder, B. (2011). Event-related repetitive transcranial magnetic stimulation of posterior superior temporal sulcus improves the detection of threatening postural changes in human bodies. *Journal of neuroscience*, *31*(48), 17547-17554.

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.

Chagnon N. A. (1988). Life histories, blood revenge, and warfare in a tribal population. Science (New York, N.Y.), 239(4843), 985–992.

Chen, G., Shin, Y., Taylor, P., Glen, D., Reynolds, R., Israel, R., & Cox, R. (2016). Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *Neuroimage, 142*, 248-259.

Chen, G., Taylor, P., Shin, Y., Reynolds, R., & Cox, R. (2017). Untangling the relatedness among correlations, part II: Inter-subject correlation group analysis through linear mixed-effects modeling. *Neuroimage, 147*, 825-840.

Clifford, M. M. (1975). Physical attractiveness and academic performance. *Child Study Journal*.

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

Costa, P. T., & McCrae, R. R. (1999). A five-factor theory of personality. *Handbook of personality: Theory and research*, *2*(01), 1999.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO Personality Inventory-Revised (NEO-PI-R)*. Psychological Assessment Resources.

Costa Jr, P. T., & McCrae, R. R. (2000). *Neo Personality Inventory*. American Psychological Association.

Crow, T. J. (1982). Two syndromes in schizophrenia?. *Trends in Neurosciences*, *5*, 351-354.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. In M. P. Zanna (Ed.), Advances in experimental social psychology, Vol. 40, pp. 61–149).

Dash, R., Udgata, S., Mohapatra, R. K., Dash, V., & Das, A. (2025). A Deep Learning Approach to Unveil Types of Mental Illness by Analyzing Social Media Posts. *Mathematical and Computational Applications*, *30*(3), 49.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral cortex*, *25*(11), 4596-4609.

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., & Orban, G. A. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study. *Journal of Neuroscience*, *24*(10), 2551-2565.

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS computational biology, 13(4), e1005508.

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature communications*, *10*(1), 1258.

Dornbusch, S. M., Hastorf, A. H., Richardson, S. A., Muzzy, R. E., & Vreeland, R. S. (1965). The perceiver and the perceived: Their relative influence on the categories of interpersonal cognition. *Journal of Personality and Social Psychology*, *1*(5), 434.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual review of vision science*, *1*(1), 393-416.

Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology*, *26*, 309-321.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research.

Eick, C. M., Kovács, G., Rostalski, S. M., Röhrig, L., & Ambrus, G. G. (2020). The occipital face area is causally involved in identity-related visual-semantic associations. *Brain structure & function*, *225*(5), 1483–1493.

Ellemers, N., Pagliaro, S., & Barreto, M. (2017). Morality and behavioural regulation in groups: A social identity approach. In *European Review of Social Psychology: Volume 24* (pp. 160-193). Routledge.

Elmore, A. L., & Crouch, E. (2020). The Association of Adverse Childhood Experiences With Anxiety and Depression for Children and Youth, 8 to 17 Years of Age. *Academic pediatrics*, *20*(5), 600–608.

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, *45*(14), 3234-3241.

Fan, X., Guo, Q., Zhang, X., Fei, L., He, S., & Weng, X. (2023). Top-down modulation and cortical-amg/hpc interaction in familiar face processing. *Cerebral Cortex*, *33*(8), 4677-4687.

Fang, K., Wen, B., Niu, L., Wan, B., & Zhang, W. (2022). Higher brain structural heterogeneity in schizophrenia. *Frontiers in psychiatry*, *13*, 1017399.

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.

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*, 116828.

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.

Földiák, P., Xiao, D., Keysers, C., Edwards, R., & Perrett, D. I. (2004). Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Progress in brain research*, *144*, 107-116.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678.

Francis, R. C. (1988). On the relationship between aggression and social dominance. Ethology, 78(3), 223-237.

Friends of the Academy of the Hebrew Language in Jerusalem. (2025). Hebrew Through The Ages. https://friendsoftheacademyofthehebrewlanguage.org/blog/history-of-hebrew-language-through-the-ages/#:~:text=Mishnaic%20Hebrew%20vocabulary%20evolved%20from,of%20the%20fifth%20century%20CE.

Frith, C. D., & Frith, U. (2021). Mapping mentalising in the brain. In *The neural basis of mentalizing* (pp. 17-45). Cham: Springer International Publishing.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 459-473.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of experimental psychology. Human perception and performance*, *28*(2), 431–446.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *Journal of cognitive neuroscience*, *12*(3), 495-504.

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

Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A., Smoller, J. W., Kwok, H., Anthony, S. E., Nakayama, K., Rhodes, G., & Wilmer, J. B. (2015). Individual Aesthetic Preferences for Faces Are Shaped Mostly by Environments, Not Genes. *Current Biology: CB*, *25*(20), 2684–2689.

Gheorghiu, A. I., Callan, M. J., & Skylark, W. J. (2017). Facial appearance affects science communication. *Proceedings of the National Academy of Sciences*, *114*(23), 5970-5975.

Gilaie-Dotan, S., & Malach, R. (2007). Sub-exemplar shape tuning in human face-related areas. *Cerebral cortex*, *17*(2), 325-338.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*(1), 32-41.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504-528.

Graham, J. R., Harvey, C. R., & Puri, M. (2017). A corporate beauty contest. *Management Science*, *63*(9), 3044-3056.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta psychologica*, *107*(1-3), 293-321.

Gross C. G. (2002). Genealogy of the "grandmother cell". *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, *8*(5), 512–518.

Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, *24*(4), 369-376.

Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I recognise you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *The Quarterly Journal of Experimental Psychology: Section A*, *51*(1), 179-195.

Haxby, J. V., & Gobbini, M. I. (2011). *Distributed neural systems for face perception* (pp. 93-110). The Oxford handbook of face perception.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223-233.

Hehman, E., Sutherland, C. A., 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.

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology*, *53*(2), 221.

Hocking, J., & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex*, *18*(10), 2439-2449.

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.

Huang, J., Xu, T., & Chan, R. C. (2011). Do patients with schizophrenia have a general or specific deficit in the perception of social threat? A meta-analytic study. *Psychiatry Research*, *185*(1-2), 1-8.

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145-E9152.

JASP Team (2025). JASP (Version 0.95.1)[Computer software].

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.), Clarendon Press, Oxford.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323.

John, O. P., Robins, R. W., & Pervin, L. A. (Third Edition). (2010). *Handbook of personality: Theory and research*. Guilford Press.

Jonas, J., Brissart, H., Hossu, G., Colnat-Coulbois, S., Vignal, J. P., Rossion, B., & Maillard, L. (2018). A face identity hallucination (palinopsia) generated by intracerebral stimulation of the face-selective right lateral fusiform cortex. *Cortex*, *99*, 296-310.

Jonas, J., & Rossion, B. (2021). Intracerebral electrical stimulation to understand the neural basis of human face identity recognition. *European Journal of Neuroscience*, *54*(1), 4197-4211.

Jones, B. C. et al. (2021). To which world regions does the valence–dominance model of social perception apply?. *Nature human behaviour*, *5*(1), 159-169.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302-4311.

Karimi-Rouzbahani, H., Ramezani, F., Woolgar, A., Rich, A., & Ghodrati, M. (2021). Perceptual difficulty modulates the direction of information flow in familiar face recognition. *NeuroImage*, *233*, 117896.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22-63.

Kietzmann, T. C., Poltoratski, S., König, P., Blake, R., Tong, F., & Ling, S. (2015). The occipital face area is causally involved in facial viewpoint perception. *Journal of Neuroscience*, *35*(50), 16398-16403.

Kim, M., Young, L., & Anzellotti, S. (2022). Exploring the representational structure of trait knowledge using perceived similarity judgments. *Social Cognition*, *40*(6), 549-579.

Kline MA, Shamsudheen R, Broesch T. (2018). Variation is the universal: making cultural evolution work in developmental psychology. *Philosophical Transactions Royal Society B*, 373.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of personality and social psychology*, *110*(5), 675.

Konorski, J. (1967). *Integrative activity of the brain.* University of Chicago Press: Chicago.

Kovács, G. (2020). Getting to know someone: Familiarity, person recognition, and identification in the human brain. *Journal of cognitive neuroscience*, *32*(12), 2205-2225.

Kramer, R. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46-58.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863-3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 249.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126-1141.

Kühn, S., Brick, T. R., Müller, B. C., & Gallinat, J. (2014). Is this car looking at you? How anthropomorphism predicts fusiform face area activation when seeing cars. *PloS one*, *9*(12), e113885.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82(13)*, 1-26.

Kwantes, C. T., & Kuo, B. C. H. (2021). Trust and Trustworthiness across Cultures. Springer.

Kwart, D. G., Foulsham, T., & Kingstone, A. (2012). Age and beauty are in the eye of the beholder. *Perception*, *41*(8), 925-938.

Lachman, S. J., & Bass, A. R. (1985). A direct study of halo effect. *The journal of psychology*, *119*(6), 535-540.

Landsiedel, J., Daughters, K., Downing, P. E., & Koldewyn, K. (2022). The role of motion in the neural representation of social interactions in the posterior temporal cortex. *Neuroimage*, *262*, 119533.

Landsiedel, J., & Koldewyn, K. (2023). Auditory dyadic interactions through the "eye" of the social brain: How visual is the posterior STS interaction region?. *Imaging Neuroscience*, *1*, 1-20.

Lavan, N., Mileva, M., & McGettigan, C. (2021). How does familiarity with a voice affect trait judgements?. *British Journal of Psychology*, *112*(1), 282-300.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of personality and social psychology*, *93*(2), 234.

Li, C., Burton, A. M., Ambrus, G. G., & Kovács, G. (2022). A neural measure of the degree of face familiarity. *cortex*, *155*, 1-12.

Lin, T., Lendry, R., & Ebner, N. C. (2016). Face likeability mediates the memory-enhancing effect of face attractiveness in young but not older adults. *Memory*, *24*(10), 1396-1406.

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: an fMRI study. *Journal of cognitive neuroscience*, *22*(1), 203-211.

Ma, Correll, & Wittenbrink (2015). The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. *Behavior Research Methods, 47*, 1122-1135.

Ma, Kantner, & Wittenbrink, (2020). Chicago Face Database: Multiracial Expansion. *Behavior Research Methods.*

Manahova, M. E., Spaak, E., & de Lange, F. P. (2020). Familiarity increases processing speed in the visual system. *Journal of cognitive neuroscience*, *32*(4), 722-733.

Maples-Keller, J. L., Williamson, R. L., Sleep, C. E., Carter, N. T., Campbell, W. K., & Miller, J. D. (2019). Using item response theory to develop a 60-item representation of the NEO PI–R using the International Personality Item Pool: Development of the IPIP–NEO–60. *Journal of personality assessment*, *101*(1), 4-15.

Marcus, D. K., & Miller, R. S. (2003). Sex differences in judgments of physical attractiveness: A social relations analysis. *Personality and Social Psychology Bulletin*, *29*(3), 325-335.

Markus, H., Smith, J., & Moreland, R. L. (1985). Role of the self-concept in the perception of others. *Journal of Personality and social Psychology*, *49*(6), 1494.

Masud, Y., & Ajmal, M. A. (2012). Left-handed people in a right-handed world: A phenomenological study. *Pakistan Journal of Social and Clinical Psychology*, *10*(1), 49-60.

MathWorks. (2024). *Geoplot - Plot line in geographic coordinates*. https://uk.mathworks.com/help/matlab/ref/geoplot.html#mw_24aee5b3-b3ae-4a22-8fe2-9e4543ea87eb

MathWorks. (2024). *Geoplot - Plot points, lines, and polygons on map - Since R2022a*. https://uk.mathworks.com/help/map/ref/geopointshape.geoplot.html

MathWorks. (2024). *regress - Multiple linear regression*. https://uk.mathworks.com/help/stats/regress.html

Mattarozzi, K., Todorov, A., Marzocchi, M., Vicari, A., & Russo, P. M. (2015). Effects of gender and personality on first impression. *PloS one*, *10*(9), e0135529.

Matz, S. C., Hyon, R., Baek, E. C., Parkinson, C., & Cerf, M. (2022). Personality similarity predicts synchronous neural responses in fMRI and EEG data. *Scientific Reports*, *12*(1), 14325.

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's" adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of personality and social psychology*, *49*(3), 710.

McKone, E., & Kanwisher, N. (2005). 17 Does the Human Brain Process Objects of Expertise Like Faces? A Review of the Evidence. *From monkey brain to human brain*, 339.

Merola, G. P., Patti, A., Catania, G., Boy, O. B., Noschese, I., Becerra, Á. R., ... & Ricca, V. (2025). Phenomenological psychopathology meets machine learning: A multicentric retrospective study (Mu. St. ARD) targeting the role of Aberrant Salience assessment in psychosis detection. Schizophrenia Research, 281, 147-156.

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.

Mileva, M., Kramer, R. S., & Burton, A. M. (2019). Social evaluation of faces across gender and familiarity. *Perception*, *48*(6), 471-486.

Milner, D., & Goodale, M. (2006). *The visual brain in action* (Vol. 27). Oup Oxford.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*(1), 57-77.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, *6*, 414-417.

Morgane, P. J., Galler, J. R., & Mokler, D. J. (2005). A review of systems and networks of the limbic forebrain/limbic midbrain. *Progress in Neurobiology*, *75*(2), 143-160.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of cognitive neuroscience*, *9*(5), 555-604.

Mundel, T., Milton, J. G., Dimitrov, A., Wilson, H. W., Pelizzari, C., Uftring, S., Torres, I., Erickson, R. K., Spire, J., & Towle, V. L. (2003). Transient inability to distinguish between faces: electrophysiologic studies. *Journal of clinical neurophysiology*, *20*(2), 102-110.

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British journal of psychology*, *102*(4), 726-747.

Natu, V. S., & O'Toole, A. J. (2015). Spatiotemporal changes in neural response patterns to faces varying in visual familiarity. *Neuroimage*, *108*, 151-159.

Nguyen, H. T., Isaacowitz, D. M., & Rubin, P. A. (2009). Age-and fatigue-related markers of human faces: an eye-tracking study. *Ophthalmology*, *116*(2), 355-360.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS computational biology*, *10*(4), e1003553.

Office for National Statistics. (2021). *England.* https://explore-local-statistics.beta.ons.gov.uk/areas/E92000001-england.

Office for National Statistics. (2021). Ethnic group variable: Census 2021. https://www.ons.gov.uk/census/census2021dictionary/variablesbytopic/ethnicgroupnationalidentitylanguageandreligionvariablescensus2021/ethnicgroup.

Oh, D., Martin, J. D., & Freeman, J. B. (2022). Personality across world regions predicts variability in the structure of face impressions. *Psychological Science*, *33*(8), 1240-1256.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of nonverbal behavior*, *34*, 83-110.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092.

Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in cognitive sciences*, *8*(7), 315-324.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in cognitive sciences*, *6*(6), 261-266.

Pandeirada, J. N., Madeira, M., Fernandes, N. L., Marinho, P., & Vasconcelos, M. (2022). Judgements of Social Dominance From Faces and Related Variables. Frontiers in Psychology, 13, 873147.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1), 332.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L. & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, *394*(6696), 884-887.

Perrett, D. I. & Mistlin, A. J. (1990). Perception of facial attributes. In Stebbins, W. C., & Berkley, M. A. (Eds.). (1990). *Comparative perception, Vol. 2. Complex signals.* John Wiley & Sons.

Pérusse, D. (1993). Cultural and reproductive success in industrial societies: Testing the relationship at the proximate and ultimate levels. Behavioral and Brain Sciences, 16(2), 267-283.

Peskin, M., & Newell, F. N. (2004). Familiarity breeds attraction: Effects of exposure on the attractiveness of typical and distinctive faces. *Perception*, *33*(2), 147-157.

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in cognitive sciences*, *25*(2), 100-110.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental brain research*, *209*, 481-493.

Popova, T., & Wiese, H. (2023). Developing familiarity during the first eight months of knowing a person: A longitudinal EEG study on face and identity learning. *cortex*, *165*, 26-37.

Popova, T., & Wiese, H. (2023). How quickly do we learn new faces in everyday life? Neurophysiological evidence for face identity learning after a brief real-life encounter. *Cortex*, *159*, 205-216.

Popova, T., & Wiese, H. (2022). The time it takes to truly know someone: Neurophysiological correlates of face and identity learning during the first two years. *Biological Psychology*, *170*, 108312.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cerebral cortex*, *9*(5), 445-458.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Ramon, M., & Gobbini, M. I. (2018). Familiarity matters: A review on prioritized processing of personally familiar faces. *Visual Cognition*, *26*(3), 179-195.

Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. Progress in Brain Research, 253, 243-262.

Rhodes, G., Morley, G., & Simmons, L. W. (2013). Women can judge sexual unfaithfulness from unfamiliar men's faces. *Biology letters*, *9*(1), 20120908.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*(1), 32-47.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. Journal of personality and social psychology, 9(4), 283.

Rossion, B. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in bioscience*, *6*(258), e307.

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, *126*(11), 2381-2395.

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain and cognition*, *79*(2), 138-157.

Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., & Crommelinck, M. (2001). How does the brain discriminate familiar and unfamiliar faces?: a PET study of face categorical perception. *Journal of cognitive neuroscience*, *13*(7), 1019-1034.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: behavioral and neural correlates. *Journal of personality and social psychology*, *104*(3), 409–426.

Ryali, C. K., Goffin, S., Winkielman, P., & Yu, A. J. (2020). From likely to likable: The role of statistical typicality in human social assessment of faces. Proceedings of the National Academy of Sciences, 117(47), 29371-29380.

Ryan, W. B. F., Carbotte, S. M., J, Coplan., O'Hara, S., Melkonian, A., Arko, R., Weissel, R.A., Ferrini, V., Goodwillie, A., Nitsche, F., Bonczkowski, J., and Zemsky, R. (2009). Global Multi-Resolution Topography (GMRT) synthesis data set, *Geochemistry*. *Geophysics*. *Geosystems*., 10, Q03014, doi:10.1029/2008GC002332.

Selfhout, M., Burk, W., Branje, S., Denissen, J., Van Aken, M., & Meeus, W. (2010). Emerging late adolescent friendship networks and Big Five personality traits: A social network approach. *Journal of Personality*, *78*(2), 509-538.

Selfhout, M., Denissen, J., Branje, S., & Meeus, W. (2010). In the eye of the beholder: Perceived, actual, and peer-rated similarity in personality, communication, and friendship intensity during the acquaintanceship process. *Journal of Personality and Social Psychology*, *96*(6), 1152-1165.

Shen, Y. L., Hyon, R., Wheatley, T., Kleinbaum, A. M., Welker, C. L., & Parkinson, C. (2025). Neural similarity predicts whether strangers become friends. *Nature Human Behaviour*, 1-14.

Schwartz, L., & Yovel, G. (2016). The roles of perceptual and conceptual information in face recognition. *Journal of Experimental Psychology: General*, *145*(11), 1493.

Shultz, S., Lee, S. M., Pelphrey, K., & McCarthy, G. (2011). The posterior superior temporal sulcus is sensitive to the outcome of human and non-human goal-directed actions. *Social cognitive and affective neuroscience*, *6*(5), 602-611.

Shultz, S., & McCarthy, G. (2012). Goal-directed actions activate the face-sensitive posterior superior temporal sulcus and fusiform gyrus in the absence of human-like perceptual cues. *Cerebral Cortex*, *22*(5), 1098-1106.

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.

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H., & Todorov, A. (2017). For your local eyes only: Culture-specific face typicality influences perceptions of trustworthiness. *Perception*, *46*(8), 914-928.

Solnick, S. J., & Schweitzer, M. E. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. Organizational behavior and human decision processes, 79(3), 199-215.

Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Urrutia, A., Natalicio, L. F., & Natalicio, D. S. (1971). The state-trait anxiety inventory. *Revista Interamericana de Psicologia/Interamerican journal of psychology*, *5*(3 & 4).

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 of the United States of America*, 117(19), 10218–10224.

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105-118.

Sutherland, C. A., Rhodes, G., Burton, N. S., & Young, A. W. (2020). Do facial first impressions reflect a shared social reality?. *British Journal of Psychology*, *111*(2), 215-232.

Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of cognitive neuroscience*, *18*(9), 1488-1497.

The Holy Bible: The Amplified Bible. (2015). The Lockman Foundation. (Original work published 1987).

Thorndike, E. L. (1920). A constant error in psychological ratings. Journal of Experimental Psychology, 4, 25-29.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623-1626.

Tomlinson, E. C., & Schnackenberg, A. (2022). The effects of transparency perceptions on trustworthiness perceptions and trust. Journal of Trust Research, 12(1), 1–23.

Tootell, R. B., Tsao, D., & Vanduffel, W. (2003). Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. *Journal of Neuroscience*, *23*(10), 3981-3989.

Tsantani, M., Kriegeskorte, N., McGettigan, C., & Garrido, L. (2019). Faces and voices in the brain: a modality-general person-identity representation in superior temporal sulcus. *NeuroImage*, *201*, 116004.

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L. (2021). FFA and OFA Encode Distinct Types of Face Identity Information. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *41*(9), 1952–1969.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nature neuroscience*, *6*(9), 989-995.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670-674.

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications biology*, *1*, 62.

Van Essen, D. C., & Dierker, D. L. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron*, *56*(2), 209-225.

Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B., Bakircioglu, M., & Miller, M. I. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision research*, *41*(10-11), 1359-1378.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K. the WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage, 80*(2013), 62-79.

Varela, V. P. L., Towler, A., Kemp, R. I., & White, D. (2023). Looking at faces in the wild. *Scientific Reports, 13*, 783.

Vernon, R., Sutherland, C., Young, A., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences, 111(32)*, E3353-E3361.

Verosky, S. C., Todorov, A., & Turk-Browne, N. B. (2013). Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia*, *51*(11), 2100-2108.

Visconti di Oleggio Castello, M., & Gobbini, M. I. (2015). Familiar face detection in 180ms. *PloS one*, *10*(8), e0136548.

Visconti di Oleggio Castello, M., Halchenko, Y. O., Guntupalli, J. S., Gors, J. D., & Gobbini, M. I. (2017). The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Scientific reports*, *7*(1), 12237.

Visconti di Oleggio Castello, M., Haxby, J. V., & Gobbini, M. I. (2021). Shared neural codes for visual and semantic information about familiar faces in a common representational space. *Proceedings of the National Academy of Sciences*, *118*(45), e2110474118.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, *14*(5), 779-804.

Walster, E., Aronson, V., Abrahams, D., & Rottman, L. (1966). Importance of physical attractiveness in dating behavior. *Journal of personality and social psychology*, *4*(5), 508.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, *137*, 188-200.

Willems, R. M., Peelen, M. V., & Hagoort, P. (2010). Cerebral lateralization of face-selective and body-selective visual areas depends on handedness. *Cerebral cortex*, *20*(7), 1719-1725.

Willis, M. L., Dodd, H. F., & Palermo, R. (2013). The relationship between anxiety and the social judgements of approachability and trustworthiness. *PLoS One*, *8*(10), e76825.

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.

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, *59*(2), 189-202.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological science*, *26*(8), 1325-1331.

Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. Behavioral Ecology and Sociobiology, 64(10), 1673-1683.

Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): Image or person?. *Vision research*, *49*(23), 2800-2807.

Yan, X., Volfart, A., & Rossion, B. (2023). A neural marker of the human face identity familiarity effect. *Scientific reports*, *13*(1), 16294.

Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron*, *44*(5), 889-898.

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not competent but warm... really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, *8*(3), 291-308.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, *9*(2p2), 1.

Zebrowitz, L. A., Bronstad, P. M., & Lee, H. K. (2007). The contribution of face familiarity to ingroup favoritism and stereotyping. *Social Cognition*, *25*(2), 306-338.

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.

Appendices

Appendix A

Research Ethics Application Form for the experiments in Chapter 3 and Chapter 4 (approved on 21/09/2018)

Brunel University London

Research Ethics Application Form

A1 Project Details

A1 Project short title

The neural underpinnings of face and voice identity perception

A2 Project full title

The neural underpinnings of face and voice identity perception

A3 Proposed Start Date 01/10/2018

NOTE: If you are using human participants, their data or their tissue, you must ensure you have research ethics approval BEFORE you commence your research.

A4 Proposed End date: 01/10/2019

Applicant Details

Appendix B

Consent form and participant information provided for participants in the experiments in Chapter 3 and Chapter 4

ROYAL HOLLOWAY, UNIVERSITY OF LONDON - MAGNETIC RESONANCE IMAGING UNIT

CONSENT FORM

NAME OF PARTICIPANT……

Please read the following statement carefully and then add your signature. If you have any questions, please ask the person who gave you this form. You are under no pressure to give your consent and you are free to withdraw from the MRI examination at any time.

I agree to participate in an MRI examination conducted for research purposes

by …… (name of researcher)

on ……(name of project).

- I understand that the examination is not part of any medical treatment.
- I have completed two screening forms and I have been given an opportunity to discuss any issues arising from them.
- The nature of the examination has been explained to me and I have had an opportunity to ask questions about it.
- I consent to a specialist and my UK general practitioner being contacted in the unlikely event that the scan reveals any suspected abnormality.

Signature …… Date……

(for children under 18 years: signature by child and a parent or guardian)

WITNESS:

Statement by a witness, who must be either an authorised person or a scientific collaborator who is familiar with the experimental procedure and is able to answer questions about it.

I certify that the above participant signed this form in my presence. I am satisfied that the participant fully understands the statement made and I certify that he/she had adequate opportunity to ask questions about the procedure before signing.

Signature…… Date……

Name ……..

Address of witness (if not an Authorised Person):

COMBINED UNIVERSITIES BRAIN IMAGING CENTRE CUBIC

ROYAL HOLLOWAY, UNIVERSITY OF LONDON - MAGNETIC RESONANCE IMAGING UNIT

INITIAL SCREENING FORM

NAME OF PARTICIPANT ..... Sex: M / F

Date of birth..... Approximate weight in kg..... Approximate height in cm.....

Please read the following questions CAREFULLY and provide answers. For a very small number of individuals, being scanned can endanger comfort, health or even life. The purpose of these questions is to make sure that you are not such a person.

You have the right to withdraw from the screening and subsequent scanning if you find the questions unacceptably intrusive. The information you provide will be treated as strictly confidential and will be held in secure conditions.

	Delete as appropriate
1. Have you been fitted with a pacemaker or artificial heart valve?	YES/NO
2. Have you any aneurysm clips, shunts or stents in your body or a cochlear implant?	YES/NO
3. Have you ever had any metal fragments in your eyes?	YES/NO
4. Have you ever had any metal fragments, e.g. shrapnel in any other part of your body?	YES/NO
5. Have you any surgically implanted metal in any part of your body, other than dental fillings and crowns (e.g. joint replacement or bone reconstruction)	YES/NO
6. Have you ever had any surgery that might have involved metal implants of which you are not aware?	YES/NO
7. Do you wear a denture plate or brace with metal in it?	YES/NO
8. Do you wear a hearing aid?	YES/NO
9. Do you use drug patches attached to your skin?	YES/NO
10. Have you ever suffered from any of: epilepsy, diabetes or thermoregulatory problems?	YES/NO
11. Have you ever suffered from any heart disease?	YES/NO
12. Is there any possibility that you might be pregnant?	YES/NO
13. Have you been sterilised using clips?	YES/NO
14. Do you have a contraceptive coil (IUD) or other contraceptive implants installed?	YES/NO
If yes, please provide details: _____	
15. Are you currently breast-feeding an infant?	YES/NO

Please enter below the name and address of your UK doctor (general practitioner).

I have read and understood the questions above and have answered them correctly.

SIGNED..... DATE.....
(for children under 18 years: signature by a parent or guardian)

In the presence of ..... (name) .....(signature)

Address of witness, if not the experimenter:

COMBINED UNIVERSITIES BRAIN IMAGING CENTRE

CUBIC

ROYAL HOLLOWAY, UNIVERSITY OF LONDON - MAGNETIC RESONANCE IMAGING UNIT

SECOND SCREENING FORM

This form should be completed and signed immediately before your scan, after removal of any jewellery or other metal objects and (if required by the operator) changing your clothes.

NAME OF PARTICIPANT .....

Date of birth..... Sex: M / F

Please read the following questions CAREFULLY and provide answers. For a very small number of individuals, being scanned can endanger comfort, health or even life. The purpose of these questions is to make sure that you are not such a person.

You have the right to withdraw from the screening and subsequent scanning if you find the questions unacceptably intrusive. The information you provide will be treated as strictly confidential and will he held in secure conditions.

BEFORE YOU ARE TAKEN THROUGH FOR YOUR SCAN IT IS ESSENTIAL THAT YOU REMOVE **ALL METAL OBJECTS** INCLUDING:-WATCHES, PENS, LOOSE CHANGE, KEYS, HAIR CLIPS, ALL JEWELLERY, METALLIC COSMETICS, CHEQUE/CASH POINT CARDS.

Delete as appropriate

1. Are you wearing or carrying any metal items such as those listed above? YES/NO
2. Have your answers to any of the questions in the initial screening form changed?
   (The initial screening form must be shown to you before you answer this question.) YES/NO

Specifically, please confirm:

3. Have you been fitted with a pacemaker, artificial heart valve or cochlear implant? YES/NO
4. Are you wearing a drug patch attached to your skin? YES/NO
5. Is there any possibility that you might be pregnant? YES/NO

I have read and understood the questions above and have answered them correctly.

SIGNATURE..... DATE.....
(for children under 18 years: signature by a parent or guardian)

FOR STAFF USE:

I certify that the initial screening form and the consent form have been completed by the person named above and I have attached them to this form. The volunteer has been given the standard information sheet about MRI experiments, together with any necessary study-specific information, and has been given an opportunity to ask questions. I am satisfied that the volunteer is adequately informed and understands the content of the consent form. I have taken adequate steps to ensure that the volunteer has no ferro-magnetic metal in or on his/her person and I am satisfied that the scan can proceed.

SIGNATURE..... **NAME (print) .....**

College of Health and Life Sciences
Department of Life Sciences

CONSENT FORM

Title of study: The neural underpinnings of face and voice identity perception

The participant should complete the whole of this sheet	***Please tick the appropriate box*** YES	NO
Have you read the Research Participant Information Sheet?		
Have you had an opportunity to ask questions and discuss this study?		
Have you received satisfactory answers to all your questions?		
Who have you spoken to?		
Do you understand that you will not be referred to by name in any report concerning the study?		
Do you understand that you are free to withdraw from the study: • at any time? • without having to give a reason for withdrawing?		
I agree to the use of non-attributable direct quotes when the study is written up or published.		
I agree that anonymised research data may be shared with other researchers for further analyses, and made available as 'open data'. This means that research data will be publicly available and may be used for purposes unrelated to this study. It will not be possible to identify me from this data.		
I agree that my personal data will be separated from the research data and password-protected. Both personal and research data will be kept indefinitely for scientific research purposes.		
Do you agree to take part in this study?		
Signature of Research Participant:		
Date:		
Name in capitals:		

Researcher name:	Signature:
Supervisor name:	Signature:

College of Health and Life Sciences

Department of Life Sciences

PARTICIPANT INFORMATION SHEET

Study title: The neural underpinnings of face and voice identity perception

You are being invited to take part in a research study conducted by Dr Lúcia Garrido and Dr Olga Korolkova. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study?

This study is concerned with the ability to recognise faces and voices, focussing on how we learn new faces and voices. We want to better understand the neural mechanisms involved in visual and auditory recognition and how we represent identities in the brain. Participation will involve five different parts:

- Part 1: MRI session 1 (taking approximately 2 hours)
- Part 2: Behavioural session 1 (taking approximately 1 hour and 30 minutes)
- Part 3: Watching a season of The Apprentice at home and completing quick online questionnaires about each episode (12 hours of watching + 1h of questionnaires in total)
- Part 4: MRI session 2 (taking approximately 2 hours)
- Part 5: Behavioural session 2 (taking approximately 1 hour and 30 minutes)

These sessions will take place on different days and will last a total of up to 20 hours. We expect that all five parts will be completed within 3 to 4 weeks.

Why have I been invited to participate?

Because you are between 18-35 years of age, you are a native English speaker, you are right handed, you are able to see and hear our stimuli in the scanner (you have normal or corrected-to-normal vision and normal hearing), and you are able to remove all metal from your body. For this study, we plan to invite up to 40 participants.

Do I have to take part?

As participation is entirely voluntary, it is up to you to decide whether or not to take part. You were given two information sheets (this one, plus one information sheet on MRI scanning). We ask you to read both forms carefully before deciding to take part. If you do decide to take part, you will need to sign two consent forms (one for the entire study and one specifically for the MRI scanning). If you decide to take part you are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part?

If you agree to take part you will be asked to fill in an initial MRI screening form and will be asked some basic questions to ensure you are not familiar with the season of the Apprentice already.

If you do fulfil our requirements you will be invited to come to the MRI Unit at Royal Holloway for a behavioural testing session and an MRI scanning session. For the **behavioural session (part 1)** you will

be asked to make some simple judgements about faces and voices, such as rating them for their trustworthiness, memorising them, or rating pairs according to their perceived similarity. This session will take approximately 1 hour and 30 minutes.

For the **MRI session (part 2),** you will be asked to complete a second MRI screening form before you enter the scanner. Whilst inside the scanner you will be presented with visual stimuli (such as faces, scenes, objects), auditory stimuli (such as voices, natural sounds, noise), and audiovisual stimuli, while we measure your neural activity. You will be asked to pay attention to the stimuli and make responses by pressing buttons. Detailed instructions will be given just before the scan, and from time to time during it. You will not spend more than 1 hour and 30 minutes inside the scanner and the additional 30 minutes of the 2-hour session will be used to explain the study to you, fill in forms and answer your questions. We will ask you not to move your head while inside the scanner.

For **part 3** of the study, you will be given a season of the Apprentice that you will be asked to watch within the next 2-3 weeks while completing a short online questionnaire (5 minutes each) after each season.

After you have watched all episodes, you will be invited back to complete another **MRI session (part 4)** and another **behavioural testing session (part 5)**, and they will be very similar to the initial ones.

You will receive **£120** for participating in all five parts of the study. You will only be reimbursed at the end of the study, after you complete all five parts. If you dropout before the final session, you will be reimbursed with **£10** for each session that you have completed (watching the show and completing the online questionnaires are considered to be one session).

What do I have to do?

There are no dietary or lifestyle restrictions for the behavioural study. That means you can continue to take your regular medications, but we will ask you not to drink alcohol before the study. For fMRI scanning sessions, if your vision should be corrected, we will ask you to use contact lenses instead of your usual glasses, or to use our special MRI-compatible glasses. Before entering the scanner room we will ask you to remove all metal objects including: watches, pens, change, keys, hair clips, jewellery, piercing, metallic cosmetics, credit cards (they can be damaged by the magnetic field).

What are the possible disadvantages and risks of taking part?

As far as we know, fMRI poses no direct health risks. However, the Department of Health advises that certain people should NOT be scanned. Prior to going in the scanner, you will be screened in order to ensure that it is safe for you to enter. Please carefully read through the MRI participant information sheet and MRI consent form that will be provided.

There are no anticipated disadvantages and risks associated with the behavioural tasks you will be asked to perform in this study. The effects of participating should be comparable to those you would experience being in a lecture for an equivalent length of time. To avoid possible fatigue, you will be given a number of opportunities to take a break.

What are the possible benefits of taking part?

There is no immediate benefit to the participant from taking part in the study. The research data collected from a group of participants and analysed may provide new insights about organising face and voice information in the brain.

What if something goes wrong?

You may stop the experiment at any time without providing a reason. If you have any questions, concerns, suggestions, or complaints that are not being addressed by the researcher, please contact the Chair of the College Research Ethics Committee (please see contact information below).

Will my taking part in this study be kept confidential?

All information that is collected about you during the course of the research will be kept strictly confidential. The researchers will save your personal data (such as your name, date of birth, or e-mail) separately from your research data. Your responses in behavioural tasks, questionnaires, and your MRI data (research data) will be solely assigned a code. The researchers will keep spreadsheets linking the codes for each participant with their names and other personal information. These spreadsheets will be password protected, saved securely, and only available to researchers in the project.

All data collected during the MRI scans will also be stored according to the rules of operation of the Combined Universities Brain Imaging Centre (CUBIC). These rules require that your name, sex, and age are saved. This information is held in locked filing cabinets, and it is retained at least for ten years.

Your anonymised research data may be shared with other researchers at other universities and research centres for further analyses, and made available as "open data" in a repository. This means that the data will be publicly available and may be used for other purposes not related to this study. However, it will not be possible to identify you from these data, and your personal data will not be shared at any point.

What will happen to the results of the research study?

The research data will be coded (for anonymity) and analysed by the researchers before being reported. The results may also be presented in appropriate scientific journals and conferences. If you take part in this research, you can obtain copies of these publications from the researchers. The data will be stored by the Chief Investigator Dr Lúcia Garrido at Brunel University. The research data may be shared with other researchers in the world and made available as "open data" but at no point will any personal data be shared.

Who is organising and funding the research?

This research is organised by Dr Lúcia Garrido and Dr Olga Korolkova, Division of Psychology at Brunel University, and funded by a Leverhulme Trust grant awarded to Dr Lúcia Garrido.

What are the indemnity arrangements?

Brunel University holds Public liability insurance policies that apply to this study. If you can demonstrate that you experienced harm as a result of you participation in this study, you may be able to claim compensation. Please contact Professor Peter Hobson, the Chair of the University Research Ethics Committee ([redacted]) if you would like further information about the insurance arrangements that apply to this study.

Who has reviewed the study?

This study has been reviewed and approved by the College Research Ethics Committee on 21/09/2018.

Brunel University is committed to compliance with the Universities UK Research Integrity Concordat. You are entitled to expect the highest level of integrity from our researchers during the course of their research.

Contact for further information and complaints

If you have any questions concerning this study, please contact the researchers: Dr Olga Korolkova ([redacted]) and Dr Lúcia Garrido ([redacted]), Division of Psychology, Department of Life Sciences, Brunel University.

For complaints and questions about the conduct of the research, please contact Professor Christina Victor, Chair of the College of Health and Life Sciences Research Ethics Committee ([redacted]).

Thank you for reading this information sheet.

Appendix C

Contents of the USB drive given to participants during the interim session between pre- and post-familiarisation rating and scanning sessions

Name	Date Modified	Size	Kind
INSTRUCTIONS.docx	23 October 2018 at 12:11	78 KB	Microso...(.docx)
INSTRUCTIONS.pdf	23 October 2018 at 12:12	28 KB	PDF Document
Links_to_all_questionnaires.docx	23 October 2018 at 11:48	147 KB	Microso...(.docx)
Links_to_all_questionnaires.pdf	23 October 2018 at 11:53	35 KB	PDF Document
The_Apprentice	23 October 2018 at 14:30	--	Folder
EP01	24 July 2019 at 16:34	--	Folder
ep_01.mp4	9 March 2017 at 17:19	319.3 MB	MPEG-4 movie
Link_to_questionnaire_EP01.txt	23 October 2018 at 12:21	65 bytes	Plain Text
EP02	23 October 2018 at 12:25	--	Folder
ep_02.mp4	25 August 2017 at 13:20	392.7 MB	MPEG-4 movie
Link_to_questionnaire_EP02.txt	23 October 2018 at 12:22	65 bytes	Plain Text
EP03	24 July 2019 at 16:35	--	Folder
ep_03.mp4	9 May 2017 at 11:03	766.7 MB	MPEG-4 movie
Link_to_questionnaire_EP03.txt	23 October 2018 at 12:22	65 bytes	Plain Text
EP04	23 October 2018 at 12:25	--	Folder
ep_04.mp4	23 February 2017 at 09:42	330.2 MB	MPEG-4 movie
Link_to_questionnaire_EP04.txt	23 October 2018 at 12:22	65 bytes	Plain Text
EP05	24 July 2019 at 16:35	--	Folder
ep_05.mp4	21 April 2017 at 16:08	717.4 MB	MPEG-4 movie
Link_to_questionnaire_EP05.txt	23 October 2018 at 12:22	65 bytes	Plain Text
EP06	23 October 2018 at 12:25	--	Folder
ep_06.mp4	20 April 2017 at 13:34	745 MB	MPEG-4 movie
Link_to_questionnaire_EP06.txt	23 October 2018 at 12:22	65 bytes	Plain Text
EP07	24 July 2019 at 16:36	--	Folder
ep_07.mp4	24 April 2017 at 13:03	718.7 MB	MPEG-4 movie
Link_to_questionnaire_EP07.txt	23 October 2018 at 12:23	65 bytes	Plain Text
EP08	23 October 2018 at 12:26	--	Folder
ep_08.mp4	24 April 2017 at 11:31	736.5 MB	MPEG-4 movie
Link_to_questionnaire_EP08.txt	23 October 2018 at 12:23	65 bytes	Plain Text
EP09	24 July 2019 at 16:36	--	Folder
ep_09.mp4	21 April 2017 at 18:56	692.7 MB	MPEG-4 movie
Link_to_questionnaire_EP09.txt	23 October 2018 at 12:23	65 bytes	Plain Text
EP10	23 October 2018 at 12:26	--	Folder
ep_10.mp4	25 April 2017 at 11:31	652.1 MB	MPEG-4 movie
Link_to_questionnaire_EP10.txt	23 October 2018 at 12:23	65 bytes	Plain Text
EP11	24 July 2019 at 16:37	--	Folder
ep_11.mp4	8 May 2017 at 11:40	606.2 MB	MPEG-4 movie
Link_to_questionnaire_EP11.txt	23 October 2018 at 12:23	65 bytes	Plain Text
EP12	23 October 2018 at 12:26	--	Folder
ep_12.mp4	25 August 2017 at 14:37	480.8 MB	MPEG-4 movie
Link_to_questionnaire_EP12.txt	23 October 2018 at 12:24	65 bytes	Plain Text