UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Face Perception in Natural Images

Permalink

https://escholarship.org/uc/item/73m0111j

Author

Hao, Susan

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Face Perception in Natural Images

by

Susan Hao

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sonia Bishop, Chair
Professor David Whitney
Professor Frederic Theunissen
Profressor Kevin Weiner

Fall 2022

Face Perception in Natural Images

Copyright 2022
by
Susan Hao

Abstract

Face Perception in Natural Images

by

Susan Hao

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Sonia Bishop, Chair

Face perception is an integral part of social relationships and how humans view the world. Although we typically see faces that range in demographics and expressions under natural conditions, many studies in the field of vision have researched face perception using unnaturalistic, posed, or even artificial stimuli. In this dissertation, I aim to understand how individuals perceive emotional faces as they would in the real world – by using naturalistic stimuli and further examine what features contribute to individuals' overall perception. Chapter 1 provides a brief overview of the prior literature describing traditional models of face perception in the fields of psychology and neuroscience. These models have historically used controlled, artificial stimuli potentially leading to biases in results specific to stimuli used in those studies. In lieu of using highly controlled face stimuli, I propose using naturalistic images varying in age, gender, expression, etc. to study face perception behaviorally and in the brain through functional magnetic resonance imaging (fMRI). Chapter 2 describes a behavioral study that investigates how individuals judge emotions of naturalistic scenes of crowds. Examining what features in the scene contribute to the overall perception of affect in the crowd, this study revealed that the reliance of the mean emotion of the crowd vs the most emotional face of the crowd varies with the dimension of affect under consideration. Additionally, background context was shown to have an increased influence on reliance of the most emotional face for certain affect dimensions. In the next section, Chapter 3, I investigate how mid-level structural and textural features in naturalistic images of faces are represented in the brain using fMRI. This chapter's findings show that tuning to mid-level features occurs primarily in retinotopic regions, contradicting traditional hierarchical models of face processing. Furthermore, there exists idiosyncratic differences in mid-level representation – specifically, those with more severe Autistic traits have better representation of mid-level features in retinotopy. In the last experiment, Chapter 4, I seek to understand how high-level semantic features are organized in the brain and whether there exist individual differences in tuning to semantic features. Participants who were worse at face identity recognition were shown to have worse higher identity feature tuning in face processing re-

gions. In the last section of this dissertation, Chapter 5, I summarize my findings from all three experiments and offer insights into these results in the context of the face processing literature.

To Dong Hao

You would have loved this.

Contents

Contents ii

List of Figures v

List of Tables vii

1 General Introduction 1

2 Judging Emotion in Natural Images of Crowds 4
2.1 Introduction . . . 4
2.2 Experiment 1 . . . 6
Methods . . . 6
Participants . . . 6
Stimuli . . . 6
Experimental Procedure . . . 6
Data Analysis . . . 7
Regression Analyses . . . 9
Results . . . 9
Stepwise Regression . . . 9
Drop Column Feature Importance . . . 10
Control Analyses . . . 11
2.3 Experiment 2 . . . 12
Methods . . . 12
Participants . . . 12
Stimuli . . . 13
Experimental Procedure . . . 14
Data Analysis . . . 14
Results . . . 14
Control Analyses . . . 16
2.4 Discussion . . . 17

3 Cortical Representation of Mid-Level Face Features 21

3.1 Introduction . . . 21
3.2 Methods . . . 23
Participants . . . 23
Stimuli . . . 23
fMRI Experimental Design . . . 23
Retinotopic Localizer . . . 23
Dynamic Localizer . . . 23
Experimental Procedure . . . 24
fMRI Experimental Task . . . 24
Retinotopic and Dynamic Localizer . . . 24
Behavioral Face Perception Tasks and Questionnaires . . . 24
fMRI Data Acquisition . . . 25
Data Pre-Processing . . . 25
Functional Pre-Processing . . . 25
Structural Pre-Processing . . . 26
ROI Localizer . . . 26
fMRI Data Modeling . . . 26
Models Constructed . . . 26
Model Estimation . . . 27
Model Validation . . . 29
Statistical Analysis . . . 29
Model Prediction Significance Testing . . . 29
Visualization of Group Model Performance in fsaverage . . . 29
3.3 Results . . . 30
Group Significance Maps and Mean Validation Correlation . . . 30
Individual Differences in Model Fit . . . 32
Relating Mid-Level Models to High-Level Semantic Models . . . 35
3.4 Discussion . . . 36

4 Cortical Representation of High-Level Semantic Face Features 39
4.1 Introduction . . . 39
4.2 Methods . . . 40
fMRI Data Modeling . . . 40
Models Constructed . . . 40
Statistical Analysis . . . 42
Variance Partitioning . . . 42
4.3 Results . . . 42
Group Significance Maps and Mean Validation Correlation . . . 42
Individual Differences in Model Fit . . . 43
Variance Partitioning . . . 46
4.4 Discussion . . . 50

5 General Discussion **53**

6 References **55**

A Supplemental Materials for Chapter 2: Judging Emotion in Natural Image **68**
A.1 Instructions for Task . . . 68
Ratings of the Full Image . . . 68
Ratings of Context . . . 68
Ratings of Individual Faces . . . 69
Faces Without Context and Scrambled Faces . . . 69
A.2 Tables & Figures . . . 69

B Supplemental Materials for Chapter 3: Cortical Representation of Mid-Level Face Features **77**
B.1 Methods . . . 77
Eyetracking . . . 77
Individual Differences . . . 77
Relationship to Model Performance . . . 77
Independent Component Analysis . . . 78
PCA . . . 78
ICA of Fiducial Coordinate Model Feature Weights . . . 79
Interpreting Top Group ICs . . . 79
B.2 Tables & Figures . . . 80

C Supplemental Materials for Chapter 4: Cortical Representation of High-Level Semantic Face Features **92**
C.1 Figures . . . 92

List of Figures

2.1 Example Stimuli for Experiment 1 7
2.2 Rating procedure for Full Images 8
2.3 Drop column feature importance analysis for Experiment 1 11
2.4 Example Stimuli for Experiment 2 13
2.5 Drop column feature importance analysis for Experiment 2a 15
2.6 Drop column feature importance analysis for Experiment 2b 16

3.1 Mid-Level Models Constructed 27
3.2 Voxelwise modeling framework 28
3.3 Freesurfer fsaverage template maps 30
3.4 Group mid-level model performance in fsaverage 31
3.5 Comparison of original mid-level model performance with mid-level models fit on Gabor residuals. 32
3.6 Correlation plots of model fit and AQ scores (top panel) and CFMT scores (bottom panel) in fsaverage. 33
3.7 Mid-level model performance correlated with AQ and CFMT 34
3.8 Correlation matrix of mid-level model performance and high-level model performance 36

4.1 Group semantic model performance in fsaverage 43
4.2 Correlation plots of model performance and AQ and CFMT 45
4.3 Model performance correlated with CFMT for the Identity sub-models 46
4.4 Variance partitioning for Semantic Binary, Identity, and Emotion models 48
4.5 Correlation between the unique contribution of higher identity features and CFMT 49
4.6 The relative contribution of higher identity features vs lower identity features and CFMT 50

A.1 Is the effect of max face emotion explainable by face size? A control analysis for Experiment 1. 72
A.2 Is the effect of max face arousal driven by max dominance or vice versa? A control analysis for Experiment 1. 73
A.3 Does the min face for valence contribute significantly to full image ratings for Experiment 1? A control analysis. 74

A.4 Is the effect of max face emotion driven by face size? A control analysis for Experiment 2. 75
A.5 Does the MIN (negative) valence regressor (i.e., the maximally negative face) contribute significantly to Faces without Context and Scrambled Faces ratings in Experiment 2? 76

B.1 Correlation matrix of eyetracking summary statistics correlated with AQ and CFMT 80
B.2 Correlation matrix of eyetracking summary statistics correlated with model performance 80
B.3 ROI size correlated with AQ and CFMT 81
B.4 Functional ROIs across participants plotted in fsaverage space 81
B.5 Individual Fiducial Coordinate model performance maps 83
B.6 Individual Fiducial Attribute model performance maps 84
B.7 Individual Fiducial Texture model performance maps 85
B.8 Mid-level model performance in fsaverage with low level features regressed out . 86
B.9 Mid-level model fit on Gabor residuals performance correlated with AQ and CFMT 87
B.10 Gabor model performance correlated with AQ and CFMT 88
B.11 Correlation matrix of model performance in each ROI for mid-level and high-level models 89
B.12 PC Explained Variance 90
B.13 Hypothetical Dimension Group Regression Analyses 91

C.1 Individual Semantic Binary model performance maps 93
C.2 Individual Identity model performance maps 94
C.3 Individual Emotion model performance maps 95
C.4 Model performance correlated with AQ and CFMT for the three main semantic models 96
C.5 Model performance correlated with AQ and CFMT for the Identity sub-models . 97
C.6 Model performance correlated with AQ and CFMT for the Emotion sub-models 98
C.7 Shared variance between higher identity features and lower identity features correlated with CFMT 99

List of Tables

2.1 Stepwise Regression 10

4.1 Semantic model features 41

A.1 Demographics for Experiment 1 70
A.2 Demographics for Experiment 2a 71
A.3 Demographics for Experiment 2b 71

B.1 Semantic sub-models 82

Acknowledgments

Although my name stands alone on the title page, I am not by any means the sole contributor. This dissertation would not exist without the support of my friends, family, mentors and colleagues. I hope the words that follow will give those around me a small glimmer of how appreciative I am.

First and foremost, I am deeply grateful for my supervisor, Sonia Bishop, for taking a chance on me and supporting me while I navigated my PhD. Thank you for the countless hours spent discussing research ideas, analyses, figures, and papers. I would not be the researcher I am today without your help and encouragement. I would also like to thank past and present members of the Bishop lab who I've relied on for both collaboration as well as friendships. Specifically, I would like to express my gratitude to Ivan Alvarez and Barbora Jurigova for their contribution to the research in this dissertation. Ivan, you are an amazing researcher, and I appreciate all that you've taught me in the short time we were able to work together. Barbora, you were one of the first research assistants I hired, and I consider that one of the best decisions I made during my PhD. This dissertation would not exist without the thousands of hours you've dedicated.

I would like to extend my appreciation to my secondary advisor David Whitney for continuously encouraging me and pushing me to think more deeply about my research. To the Whitney lab members, past and present – thank you for making me feel so welcomed. I will miss our Thai lunches, boba breaks, and most of all, our VSS dance parties.

I'd like to thank the other members of my dissertation committee, Frederic Theunissen and Kevin Weiner, for their time and invaluable advice. Special thanks to Alex Huth for meeting with me every week to flesh out analyses and research questions, and thanks to my other collaborators Bhisma Chakrabarti and Brad Duchaine for their guidance and helpful feedback.

When I first started grad school, I naively thought that my work and social life should be independent. Now, I realize how inextricably linked they are – I've met so many amazing people at Berkeley, and without their constant support, I would not have been able to complete my PhD. To Teresa Canas Bajo and Sanam Mozaffari, thank you for your continued friendship and emotional support for the last five years. You two have shaped my Berkeley experience, and I can never put into words how much our patio drinks, late night chats, and not so infrequent shenanigans mean to me. Thank you to Alisa Braun, Mandy Chen, Angie Godinez, and Maedbh King for being amazing friends – I can always count on each of you to be there for me sharing together laughs, cries, and everything in between. I would also like to express my appreciation for my Bay Area friends Hari Kuppusamykrishnan, Jason Lerner, Somon Nader, and Arthur Sfez for providing a respite from grad school and encouraging me to have fun.

To my college roommates Caroline Chung, Hannah Sweeney, and Alicia Yang, thank you for supporting me all these years even from afar. Pursuing grad school wouldn't even be an option if it weren't for you three, and I'm forever grateful for our friendship.

To my mother, Yongping Dong, words cannot express how thankful I am for your love and encouragement but most of all, the sacrifices you've made to support my dreams. Without you, I would not be where I am today. To my father, Jifa Hao, thank you for pushing me to be curious, to ask questions, and to pursue my passions. I am deeply influenced by your love for research and devotion to learning. Thank you to my sister, Grace Hao, for continuously supporting me even when I'm being unreasonable. I'm so lucky to have a sister like you who I can count on no matter the circumstance. I would also like to thank my brother, Jonathan Hao, for being mature beyond his years and offering me advice far wiser than anyone would expect of a 15 year old.

Finally, to my husband James Emerick – you were the most integral part of this journey, and I owe a lot of this PhD to you. Throughout this program, you've been a huge source of emotional support – listening patiently to me when I needed to vent and offering just the right advice that encouraged me to keep going. Looking back, my PhD was not just a profound time in my life professionally, but also personally. Together, we celebrated many milestones from moving into our Berkeley cottage after a lifetime on the East Coast to getting married in the redwoods of Santa Cruz. So cheers to this final milestone and closing this chapter of our lives. I can't wait to see what comes next.

Chapter 1

General Introduction

Faces are an essential part of what makes us human. They play a key role in our social interactions and can tell us invaluable information about someone's social status, internal emotional state, or their relationship to others. Thus, it is important to understand how we perceive faces as it gives us insights into how we interact with the world. In the fields of psychology and psychophysics, studies have typically used extremely controlled stimuli to investigate how individuals perceive facial emotion. Specifically in the field of face ensemble perception, research looking at crowd face perception have used cropped, posed, or unnatural faces arranged without background context (Haberman et al., 2009; Haberman & Whitney, 2007, 2009; Im et al., 2017; Li, et al., 2016; Sun & Chong, 2020). This neglects the impact of varied, natural faces in relation to background context and may reduce the ecological validity of results.

This preference of using unnaturalistic, controlled stimuli is not exclusive to the field of psychology. Previous studies in the field of neuroscience studying face identity and emotion perception have also traditionally used posed faces with limited ranges of emotional expressions and demographics. Results in this field are somewhat contradictory with some arguing that face processing is hierarchical (Haxby et al, 2000) and domain specific (Kanwisher, 2000) with two independent pathways (one for emotion and one for identity) whereas others argue against a strictly hierarchical framework (Duchaine & Yovel, 2015) and show that emotion and identity can be co-tuned in some regions (Baseler et al., 2014; Fox et al., 2011; Ganel et al., 2005; Said et al., 2011; Wegrzyn et al., 2015). These conflicting results may be a consequence of using restricted stimulus sets with findings being potentially specific to the stimuli used in individual studies.

Given this, I attempt to address these potential methodological issues in the field by using naturalistic stimuli to study face and emotion perception. Using this method, I address the following questions:

1. When a natural image of a scene has multiple faces present, how is the perception of scene emotion determined? Do we use summary statistics for some or all emotional dimensions (i.e., averaging face emotion)? What effect does the context have on

judgement of scene emotion and on use of summary statistics? (Chapter 2)

2. When real-world faces with a large range of expressions, races, genders, and ages are used, do we see a hierarchical framework of face processing in the brain? Where do we see structural face features represented and do we see differences in tuning to these face features across individuals? (Chapter 3)

3. When using diverse, naturalistic face stimuli to investigate semantic face representation in the brain, do we see co-tuning to identity and emotion? Does representation of facial identity or emotion differ as a function of individual differences in those known to have issues with face identity or emotion processing? (Chapter 4)

With the use of naturalistic faces as stimuli in our neuroimaging studies, we need to move away from traditional statistical design and use a regularized, predictive encoding modeling framework that can circumvent a lot of issues we see in the previous functional magnetic resonance imaging (fMRI) literature such as issues with overfitting and generalizability. Additionally, interpretability techniques should be utilized to understand what features are driving our model results. Variance partitioning was one such technique which investigated aspects of face features that contribute to cortical tuning to semantic representation on a group basis. We also used techniques such as drop column feature importance in our behavioral experiments to find what aspects of scene content influence crowd emotion across participants.

While it is important to understand on a group basis the perception of emotional faces, we know that there exist individual differences in how people represent face identity and emotion. Autism and Developmental Prosopagnosia are two specific disorders that are characterized by deficits in emotion and identity processing compared to normal populations. Those who have Developmental Prosopagnosia exhibit perceptual difficulty recognizing face identity without any specific structural damage or injury to the brain. Autistic individuals have problems with social interaction and potential issues with emotion processing. Studies in the field of Autism are quite mixed with conflicting results relating to the absence or presence of perceptual emotion or identity deficits (Baron-Cohen et al., 2001a; Golan et al., 2006; Hooper et al., 2018; Harms et al., 2010; Loth et al., 2018; Philip et al., 2010; Uljarevic & Hamilton 2013; Weigelt et al., 2012; Wright et al., 2008). In many of these studies, the symptom severity of Autism is ignored and group-level analysis is performed with an Autism group and a control group. Thus, inconsistencies in the literature might be attributed to ignoring individual differences that may be coupled with symptom severity. In Chapters 3 and 4, I leverage sub-clinical assessments of Autism and Developmental Prosopagnosia traits to examine feature encoding of emotional faces in cortex as a function of severity of these traits. We use the Autism Quotient (AQ) (Baron-Cohen et al., 2001b) and the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006) to gauge the severity of Autistm and Developmental Prosopagnosia traits, respectively, in each participant. AQ and CFMT scores were then related to the strength of encoding for structural and semantic

features in cortex allowing us to understand the neural correlates of perceptual deficits in face processing.

Chapter 2

Judging Emotion in Natural Images of Crowds

2.1 Introduction

In the visual world, there is an overabundance of information in the environment around us. Our visual system is incapable of processing a large amount of this information, yet our percept of the world is rich and undisrupted. Previous studies have suggested that this rich, gist level percept relies, in part, on the use of ensemble information. Ensemble perception, that is the perception of groups of similar objects using summary statistics, is argued to happen at every level of visual processing (Whitney & Yamanashi Leib, 2018). In the case of facial expressions, findings have suggested that when adult human participants judge the emotion of a crowd, they implicitly take the mean across individual facial expressions (Haberman et al., 2009; Haberman & Whitney, 2007, 2009; Im et al., 2017; Li, et al., 2016; Sun & Chong, 2020).

Most of the studies to date that have explored the potential use of summary statistics in the perception of facial expression have asked participants to make judgements regarding basic emotions (e.g., happiness, sadness, anger). There have been far fewer studies using dimensional measures of affect. In one recent exception to this, Han et al. (2021) demonstrated that participants were able to correctly judge the average emotional valence of Mooney face crowds.

To our knowledge, there has been no work addressing whether judgments of the emotional arousal of groups is consistent with the use of summary statistics. Evidence from the attention literature suggests that highly arousing stimuli may be preferentially processed. In their arousal-biased competition model, Mather and Sutherland (2011) argue that arousal modulates the strength of competing mental representations in a manner that favors highly arousing stimuli. This is held to explain phenomena such as "pop out" which is observed when a single high arousal stimulus is rapidly detected regardless of the number of surrounding neutral stimuli (Lundqvist et al., 2015; Pinkham et al., 2010). According to Mather and

Sutherland's model (2011), arousal influences selective attention regardless of stimulus valence. In line with this, studies examining attentional orienting and disengagement have found that arousal modulates selective attention to a greater extent than valence (Ossenfort & Isaacowitz, 2021; Vogt et al., 2008). However, it remains an open question as to whether the relative salience of individual faces in a group setting will also impact the perception of group affect. One way this might be observed is in terms of a greater reliance on 'max' than 'mean' emotion for judgements of facial arousal than for face valence. In other words, participants might be disproportionately influenced by the highest arousal face when judging the arousal of groups of individuals, but this might not be observed to an equivalent extent for facial valence or other dimensions of affect.

Another important question pertains to whether reliance on summary statistics might differ for natural faces versus posed faces presented without background context. To date, most ensemble perception studies that have examined facial expression have only used a limited number of emotional expressions with faces in isolation without context. Further, while there are studies which have shown that context affects face emotion perception (Aviezer et al., 2011; Alwis & Haberman 2020; Barrett & Kensinger, 2010; Kret et al., 2013; Righart & de Gelder, 2008a, 2008b; Sasson et al., 2016), many of these have artificially inserted faces into a scene, ignoring context's orienting role in the perception of facial emotion. In an initial study using naturalistic stimuli to understand context's influence on participants' perception of individual facial expressions, Chen and Whitney (2019) investigated the extent to which affective ratings of background context - derived by taking a scene and blurring out a central character - predicted affective ratings of the complete scene. A trackball was used by participants to rate affect in a two-dimensional valence by arousal space. Across both movies and naturalistic footage, context explained nearly as much unique variance in the ratings of scene affect as that explained by affect ratings of the central character with the background blurred out. To the extent that other characters were present in the scene, they were included within the background context and ensemble perception processes were not investigated. Here, we build on this prior work by examining the extent to which background context influences the processing of the facial expression of groups, asking whether background context alters reliance on summary statistics, and whether the impact of context varies as a function of the dimension of affect in question.

With these questions in mind, our aims were as follows: 1) to investigate participants' relative reliance on mean versus maximal emotion when judging crowd emotion, 2) to determine if this varies across different dimensions of affect (namely valence, arousal, and dominance), 3) to investigate the influence of background context on judgements of crowd emotion – in particular to examine if context has an additive effect in predicting crowd emotion ratings and/or alters reliance on mean versus maximal emotion, and 4) to understand whether the extent to which statistical dependencies in visual scenes are intact or disrupted impacts the use of summary statistics.

2.2 Experiment 1

Methods

Participants

317 participants (189 male, 128 female) between the ages of 18 and 40 were recruited from Amazon Mechanical Turk (see Supplemental Table A.1 for additional demographic details). All participants provided informed consent (online tick box). Effect size estimates were not used to determine sample size as regression analyses were conducted across images with ratings averaged across participants; separate groups of participants provided ratings for predictor and dependent variables. Experimental procedures and recruitment were approved by the University of California, Berkeley Committee for the Protection of Human Subjects. All participants were required to reside in the United States and could only complete one experimental condition. Any participant who had an IP address that was outside of the United States was excluded. Participants who had greater than 50 trials in which the stimuli didn't load because of internet connection issues, and participants who responded with the same rating for greater than 75% of trials or who used a range of 4 points or less (out of a 9-point scale) were excluded ($n = 8$ total).

Stimuli

Stimuli were obtained from Google Images search queries. The stimulus set was formed using natural scenes of crowds photographed in real-world settings containing individuals of different genders, races, and ages showing varying expressions. Stimuli used as *Full Images* ($n = 168$) comprised crowd scenes containing 3 to 5 emotional faces, with each image being 500 x 500 pixels in size (Figure 2.1, column 1). Stimuli used in the *Context Condition* ($n = 168$) comprised the same images with the individual faces cropped out to leave only the background context of the scene (Figure 2.1, column 2). Stimuli used in the *Individual Face Condition* comprised the individual faces cropped out from the original *Full Images* and were presented separately (Figure 2.1, right hand columns). These face images were divided into three sets (set 1: $n = 214$, set 2: $n = 222$, set 3: $n = 200$). All faces from the same original image were kept within the same set. The order of faces within a given set were randomized. Sixty-two participants rated the *Full Image* stimuli, 59 participants rated the *Context* stimuli,and 196 participants rated the *Individual Face* stimuli ($n = 59$ for set 1, $n = 72$ for set 2, $n = 65$ for set 3).

Experimental Procedure

Ratings of stimuli were conducted online using Qualtrics. At the beginning of the experimental session, participants were asked to give consent online. Participants were then asked to maximize the window in which the stimuli were presented and to sit an arm's length distance away from the display. Participants viewed a fixation square for 1 second followed

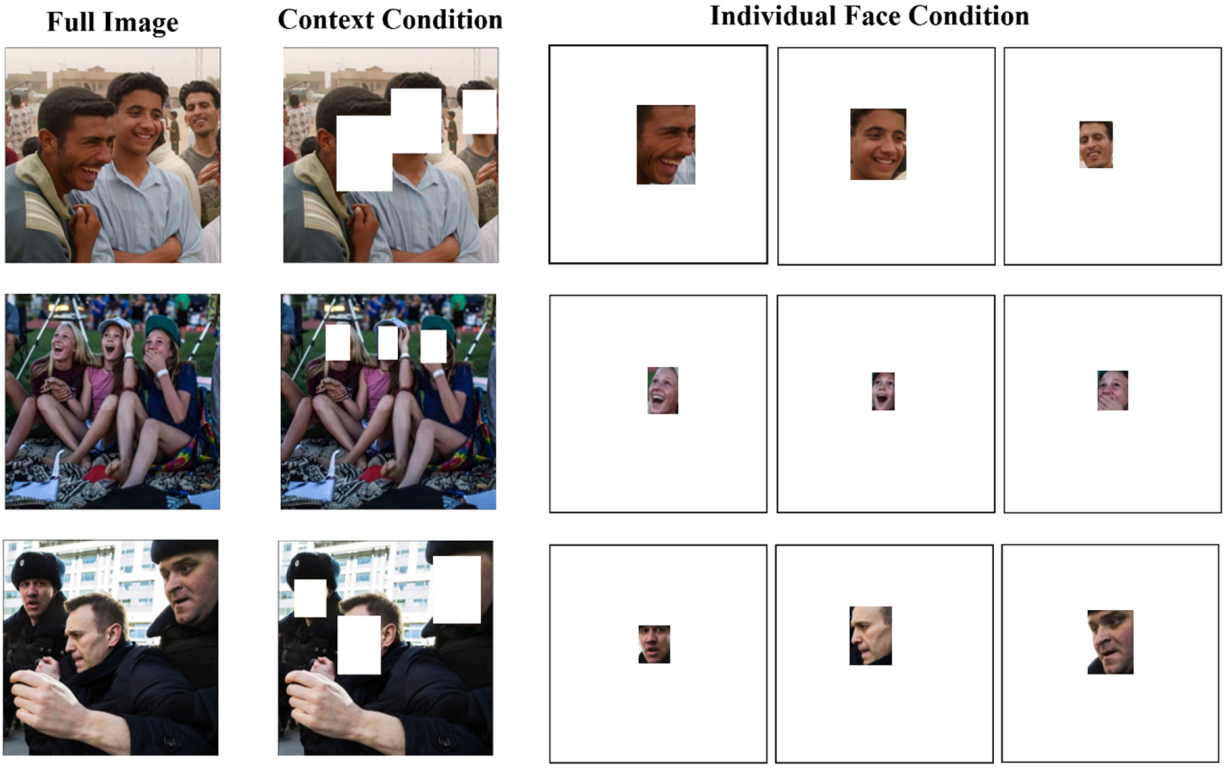


Figure 2.1: Example Stimuli for Experiment 1. In the left-hand column, we show three example Full Images stimuli. Full Images comprised 500 x 500 pixel images of naturalistic scenes with three to five emotional faces. The second column shows the corresponding stimuli presented in the Context Condition. Here, the individual faces were cropped out of each full image leaving just the background context. The remaining columns show the corresponding stimuli presented in the Individual Face Condition. Here, each face in the full image has been cropped out and presented separately.

by a stimulus image for 1 second. Valence, arousal, and dominance were rated on a 9-point Likert scale. Participants rating the *Full Image* were asked to rate the average valence, arousal, and dominance of the faces in the crowd (see Figure 2.2). In the *Context Condition*, participants were asked to rate the average valence, arousal, and dominance of the scene. In the *Individual Face Condition,* participants were asked to rate the valence, arousal, and dominance of the face. Participants were given unlimited time to make their ratings. All participants were given instructions defining each emotional dimension along with examples at the beginning of the experimental session (see Appendix A for instructions). There were 4 practice trials prior to the start of the task.

Data Analysis

Each participant's ratings were z-scored across images. Next, for each *Full Image* stimulus, scores on each dimension were averaged across participants to obtain a single score for that

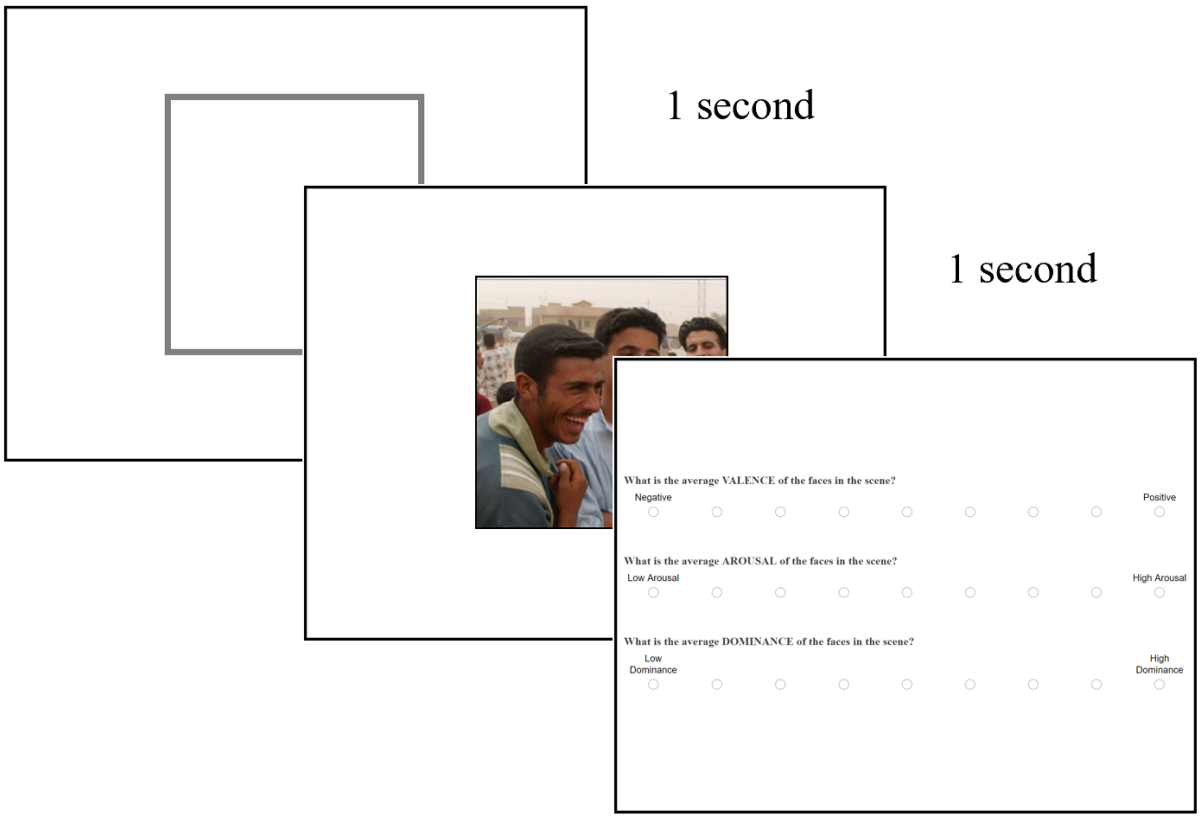


Figure 2.2: Rating procedure for Full Images. A fixation square was shown for 1 second followed by the stimulus, an image of a naturalistic scene of a crowd, for 1 second. Participants were then asked to rate the average valence, arousal, and dominance of the faces in the scene. Response time was not constrained.

image on the given dimension. Following the approach used in prior ensemble studies (e.g., Leib et al., 2016), the rating of crowd emotion for each *Full Image* was used as the dependent variable in each of the regressions described below. Here, this was conducted separately for each of the three dimensions of affect under consideration: valence, arousal, and dominance. Subsequently, for each individual face in each image, ratings from participants who completed the *Individual Faces Condition* were averaged across participants to obtain a single score for that face on each dimension of affect.

These scores were used to create the following regressors. The MEAN facial emotion regressor for each dimension of affect was calculated using the mean of the scores for all individual faces present in a given image on the dimension in question. For the dimensions of arousal, valence, and dominance, the MAX facial emotion regressors used the score for the most arousing face, the most positive face, and the most dominant face in each image, respectively. The MIN facial emotion regressors used the score for the least arousing face, the least positive (most negative) face, and the least dominant face in each image. Finally, the rating on each dimension of each image with the faces cropped out (as obtained from

participants who completed the *Context condition*) was averaged across participants. This was used to create the CONTEXT regressor for each dimension of affect.

Regression Analyses

For each of the three dimensions of affect, we conducted a forward stepwise regression with ratings of crowd emotion (*Full Images* ratings) as the dependent variable and MEAN, MAX, MIN facial emotion regressors and the CONTEXT regressor entered as potential predictors. A stopping threshold of $p < .05$ was used (i.e., the predictor that explained most variance was entered at each step, with only features that significantly improved model fit at $p < .05$ being allowed to enter each model). Forward stepwise regression gives insight into which variable out of several candidates is the strongest predictor of the dependent variable of interest and also addresses the extent to which other variables explain additional significant variance. However, each predictor variable entered may contain shared variance with other variables, and if two predictors are highly correlated, this does not allow for understanding of the unique variance explained by each. Hence, to understand the unique contribution of each predictor variable, we also conducted a drop column feature importance analysis for each dimension of affect. Here, we included MAX facial emotion, MEAN facial emotion, and CONTEXT regressors as features of interest. We did not include MIN facial emotion as this was not a significant predictor in any of the three stepwise regressions (see Results and Table 2.1). In the drop column feature importance analyses conducted, all features of interest were entered into a regression model predicting *Full Image* ratings on a given dimension of affect. Each feature, in turn, was dropped from the regression, and the model retrained with all features except the one dropped. The importance of that feature was then estimated as the difference between the model performance (adjusted R^2) with all features as predictors and model performance with all features except the one dropped. Permutation tests were performed to obtain the null distribution of feature importance scores to determine the significance of the unique contribution of each feature (significance was assessed using $p < .05$).

Results

Stepwise Regression

For valence, the features that entered the final stepwise regression model ($R^2_{adj} = .87$) (i.e., that predicted ratings of crowd valence for each *Full Image*) were the MEAN facial emotion regressor ($p < .001$), the CONTEXT regressor ($p < .001$), and the MAX facial emotion regressor ($p = .020$), see Table 2.1. For arousal, the features included in the final stepwise regression model ($R^2_{adj} = .80$) were the MAX facial emotion regressor ($p < .001$) and the CONTEXT regressor ($p < .001$), see Table 2.1. Similarly, for dominance, the MAX facial emotion regressor ($p < .001$) and the CONTEXT regressor ($p < .001$) were included in the final stepwise regression model ($R^2_{adj} = .60$), see Table 2.1.

Feature	Beta	t	Sig (t-statistic)
Valence			
Intercept	-0.07	-1.89	$p = .059$
MEAN	0.80	9.58	$p < .001$
CONTEXT	0.39	7.45	$p < .001$
MAX	0.17	2.35	$p = .020$
Arousal			
Intercept	-0.38	-14.99	$p < .001$
MAX	0.70	19.82	$p < .001$
CONTEXT	0.31	6.30	$p < .001$
Dominance			
Intercept	-0.14	-5.27	$p < .001$
MAX	0.41	7.88	$p < .001$
CONTEXT	0.47	7.67	$p < .001$

Table 2.1: Stepwise Regression. Three forward stepwise regression analyses were conducted with ratings of *Full Image* crowd affect as the dependent variable. The features entered into the final model for each dimension of affect (valence, arousal, and dominance) are presented here together with feature beta weights, t-values for the estimated weights, and associated p-values. The variables considered included the maximal facial emotion (MAX) regressor, the mean facial emotion (MEAN) regressor, the minimum facial emotion regressor (MIN), and the CONTEXT regressor for the dimension in question. The MIN variable did not enter the regression model for any of the three dimensions of affect.

Drop Column Feature Importance

For each dimension of affect, we conducted a drop column feature importance analysis to examine the unique variance explained by the MAX facial emotion, MEAN facial emotion, and CONTEXT regressors (see Figure 2.3). Permutation tests (10,000 shuffles) were performed to test whether each feature's unique R^2_{adj} was significant. For valence, we found that both the MEAN facial emotion regressor (unique $R^2_{adj} = .07$, $p < .001$) and the CONTEXT regressor (unique $R^2_{adj} = .04$, $p = .003$) made significant unique contributions to the prediction of crowd valence ratings. For arousal, the MAX facial emotion regressor (unique $R^2_{adj} = .18$, $p < .001$) and the CONTEXT regressor (unique $R^2_{adj} = .04$, $p = .004$) both made significant unique contributions to the prediction of crowd arousal ratings. For dominance, the CONTEXT regressor (unique $R^2_{adj} = .14$, $p < .001$) and the MAX facial emotion regressor (unique $R^2_{adj} = .05$, $p = .003$) both made significant unique contributions to the prediction of crowd dominance ratings.

These results suggest that CONTEXT influences judgement of crowd emotion across all three dimensions of affect. The influence of CONTEXT was especially notable for dominance ratings. In addition to CONTEXT, MEAN facial emotion explained significant unique variance in the judgements of crowd valence while the maximum (MAX) facial emotion regressor explained significant unique variance in judgments of crowd arousal and dominance.

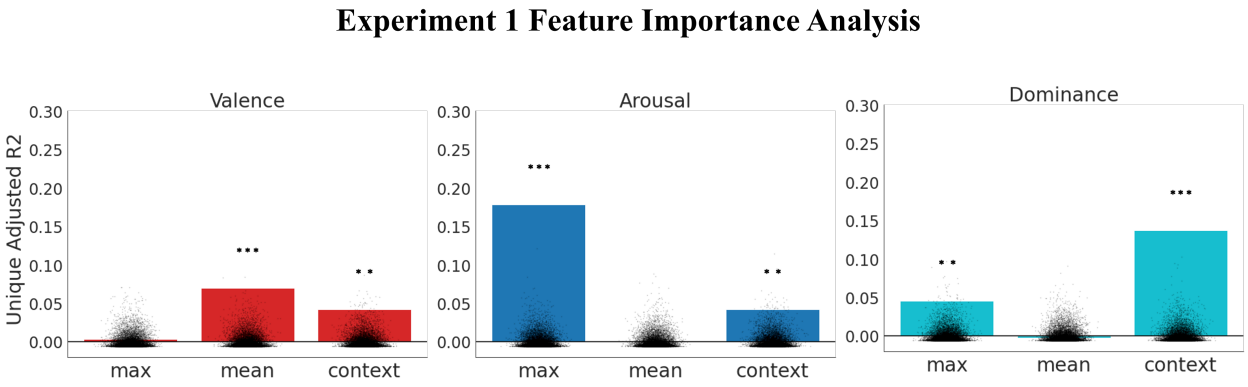


Figure 2.3: Drop column feature importance analysis for Experiment 1. A drop column feature importance analysis was performed for each dimension of affect using the max facial emotion (MAX), mean facial emotion (MEAN), and context regressors as features of interest. The results for each dimension of affect are illustrated in the three panels presented here. In each panel, each bar represents the unique variance explained by a given feature; this is calculated as the difference in adjusted R^2 for the model with every feature versus that for the model with all features except the feature of interest. Scattered dots overlaying each bar show the permuted null distribution of the unique adjusted R^2 values. Left panel: mean facial valence and context contributed significantly to ratings of *Full Image* crowd valence. Center panel: the most arousing face (max facial arousal) and context contributed significantly to ratings of *Full Image* crowd arousal. Right panel: the most dominant face (max facial dominance) and context contributed significantly to ratings of *Full Image* crowd dominance. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Control Analyses

As reported above, we observed an influence of MAX facial emotion as opposed to MEAN facial emotion for two of our three dimensions of affect. This might reflect reduced reliance on summary statistics when judgments of crowd affect are made in relation to natural scenes. However, using natural images raises the potential for confounds that are not present for artificially generated experimental stimuli. In particular, the faces within our images often varied in size. To ensure that the apparent effect of MAX arousal and MAX dominance were not simply driven by a confounding relationship with face size, we took MAX size arousal and MAX size dominance ratings (that is, ratings of arousal and dominance for the largest face in the image) and used them to predict affect ratings for the full image, retaining residual scores in each case. We then repeated the drop column analysis described above using the residual full image ratings as the dependent variable. In this manner we controlled for the influence of the largest face on the ratings of the full image. These analyses revealed that the maximally arousing and maximally dominant face in each image contributed significantly to the prediction of full image ratings even when the variance explained by the arousal or dominance ratings of the largest face had been residualized out, ($R^2_{adj} = .12$, $p < .001$; $R^2_{adj} = .03$, $p = .012$, respectively), Figure A.1.

Another important question is whether the influence of the maximally dominant face on judgements of crowd dominance might be secondary to the findings for arousal, or indeed vice versa. If the most dominant face is also often the most arousing, then attentional orient-

ing towards the most arousing face might have a knock-on influence on dominance ratings. The reverse could also apply. To address this, we re-conducted the drop column feature importance analysis for dominance on the residual scores obtained after using dominance ratings for the maximally arousing face to predict full image dominance ratings. We additionally conducted a drop-column analysis for arousal on the residual scores obtained after using arousal ratings for the most dominant face to predict full image arousal ratings. These analyses revealed that the maximally arousing face and the maximally dominant face still contributed significantly to the prediction of arousal and dominance ratings, respectively, for the full image ($R^2_{adj} = .18$, $p < .001$; $R^2_{adj} = .02$, $p = .046$, respectively, Figure A.2.)

In the case of valence, the scale's ends are defined as negative (rating of 1) to positive (rating of 9). In the analyses reported above, we used the face with the highest score (i.e., most positive expression) as the MAX face and the face with the lowest score (i.e., most negative face) as the MIN face. MIN face ratings did not enter into the model identified by our initial stepwise regression. A supplementary drop feature analysis replacing the valence MAX with the valence MIN regressor confirmed that entering the MIN (most negative) face did not account for a significant amount of unique variance in the valence regressions (Figure A.3).

2.3 Experiment 2

Most prior summary statistic experiments have not used background context. An interesting question is whether the presence of background context influences participants' reliance on the maximally emotional face versus mean emotion when rating group affect. This might arise as a result of information in the background context or in the relative positioning and pose of faces leading to differential reliance on individual faces within the image. We sought to address this through two follow up experiments. We hypothesized that removing background context (Experiment 2a) and disrupting the relationship between faces in each image (Experiment 2b) would result in greater reliance on mean emotion and reduced reliance on max emotion but that for arousal there would still be a unique contribution of the most arousing face no matter where in the image it was located. This hypothesis builds on findings from the attention literature indicating that high arousal stimuli capture spatial attention with this resulting in enhanced perceptual processing (Mather & Sutherland, 2011; Ossenfort & Isaacowitz, 2021; Vogt et al., 2008).

Methods

Participants

Separate groups of participants were recruited from Amazon Mechanical Turk for Experiments 2a and 2b. Quality assurance and exclusion criteria were identical to that for Experiment 1. Two participants were excluded from Experiment 2a and one from Experiment 2b due to responding with the same rating for greater than 75% of trials, using a range of 4

points or less (out of a 9-point scale), or having more than 50 trials fail to load because of internet issues. After these exclusions, 72 participants (52 males, 20 females) aged between 18 and 40 years took part in Experiment 2a and 62 participants (32 males, 30 females) aged between 18 and 40 years took part in Experiment 2b. See Supplemental Table A.2 & A.3 for additional demographic details.

Stimuli

In Experiment 2a, the *Full Image* stimuli used in Experiment 1 had the context removed such that only the faces that were in the image were visible (see Figure 2.4, top panel) – we named these stimuli *Faces without Context.* In Experiment 2b, faces from the *Individual Face Condition* were randomly drawn and placed in the positions held by faces in the Experiment 2a stimuli (see Figure 2.4, bottom panel). This was to examine the effect of disrupting the dependencies between faces in natural images. We called the Experiment 2b stimuli *Scrambled Faces.*



Figure 2.4: Example stimuli for Experiment 2. The top panel gives three examples of Experiment 2a stimuli: *Faces without Context* and the bottom panel gives the corresponding Experiment 2b stimuli: *Scrambled Faces.* For the latter, individual faces were randomly drawn from the stimulus set and were used to replace the original faces in each image while keeping the relative positioning of the faces intact.

Experimental Procedure

Participants were asked to judge the average valence, arousal, and dominance of the group of faces for the *Faces without Context* stimuli in Experiment 2a and the *Scrambled Faces* stimuli in Experiment 2b. Rating scales were as described in Experiment 1. Participants were given unlimited time to make their ratings, and all participants were given instructions on each dimension and examples together with 4 practice trials (as for Experiment 1).

Data Analysis

In Experiment 2a, ratings for all the *Faces Without Context* stimuli on a given dimension of affect were z-scored within participants. Following this, ratings for each stimulus for a given dimension of affect were averaged across participants. In Experiment 2b, a similar procedure was followed for the *Scrambled Faces* stimuli. For Experiment 2a, drop column feature importance analyses were performed with scores for the *Faces Without Context* stimuli on each dimension of affect as the dependent variable. MAX and MEAN facial emotion scores for each image were entered as predictor variables. These scores were calculated using the *Individual Face Condition* ratings from Experiment 1. For Experiment 2a, stimulus MAX and MEAN facial emotion scores were identical to those in Experiment 1.

For Experiment 2b, drop column feature importance analyses were similarly performed with scores for the *Scrambled Faces* stimuli on each dimension of affect as the dependent variable and MAX and MEAN facial emotion scores for each image entered as predictor variables. These scores were calculated using the *Individual Face Condition* ratings from Experiment 1. Because the faces within each image for Experiment 2b were no longer the same as in Experiment 1 (as a result of the scrambling procedure), these scores had to be re-estimated.

Results

In Experiment 2a, the MEAN facial emotion regressor explained unique variance in valence and dominance ratings for the *Faces Without Context* stimuli (valence: unique $R^2_{adj} = .10$, $p < .001$, dominance: unique $R^2_{adj} = .06$, $p < .001$), Figure 2.5. The MAX facial emotion regressor did not make a significant unique contribution to explained variance in these ratings (valence: unique $R^2_{adj} < .01$, $p = .194$, dominance: unique $R^2_{adj} = .02$, $p = .064$). In contrast, for arousal ratings, both the MAX facial emotion regressor and the MEAN emotion regressor explained significant unique variance in ratings of the *Faces Without Context* stimuli (MAX: unique $R^2_{adj} = .10$, $p < .001$, MEAN: unique $R^2_{adj} = .06$, $p < .001$), Figure 2.5. These findings are in line with our hypotheses and contrast with the results for Experiment 1 where mean facial emotion ratings did not significantly contribute to arousal or dominance ratings for the full image and a larger proportion of variance was explained by the maximally arousing and maximally dominant faces. Despite this increased reliance on mean emotion, the maximally

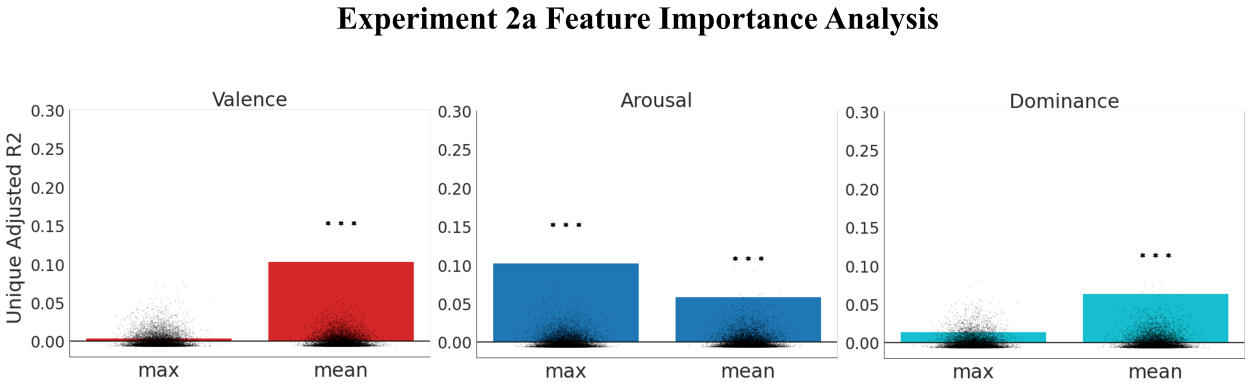


Figure 2.5: Drop column feature importance analysis for Experiment 2a. Drop column feature importance analysis was performed for each dimension of affect using max facial emotion (MAX) and mean facial emotion (MEAN) for each image as features and ratings of crowd affect for the *Faces without Context* stimuli as the dependent variable. Bars represent the unique variance explained by each regressor or 'feature'; this is calculated as the difference in adjusted R^2 for the model with both features versus that for the model with the feature in question removed. Scattered dots overlaying each bar show the permuted null distribution of the unique adjusted R^2 values. Mean facial emotion explained unique variance in *Faces without Context* ratings for all three dimensions (left panel: valence, center panel: arousal, right panel: dominance). The maximally arousing face also explained unique variance in *Faces without Context* arousal ratings. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

arousing face in each image continued to significantly influence arousal ratings for groups of natural faces with the background context removed.

In Experiment 2b, in addition to removing the background, we randomly replaced the faces in each image to eliminate any cues from one face to other faces (i.e., perhaps less arousing faces are turned towards the maximally arousing face) and to disrupt other possible spatial relationships (i.e., perhaps the maximally arousing face tends to be in the center). Drop column feature importance analyses revealed that, as in Experiments 1 and 2a, the MEAN facial emotion regressor explained unique variance in valence ratings for the *Scrambled Faces* stimuli ($R^2_{adj} = .27$, $p < .001$), Figure 2.6. In contrast, both the MEAN facial emotion regressor and the MAX facial emotion regressor explained unique variance in arousal ratings (MEAN: unique $R^2_{adj} = .07$, $p < .001$; MAX: unique $R^2_{adj} = .07$, $p < .001$) and in dominance ratings (MEAN: unique $R^2_{adj} = .06$, $p < .001$; MAX: unique $R^2_{adj} = .08$, $p < .001$). Together, these findings suggest that when background context is removed, participants show greater reliance on mean emotion when judging the emotion of a group of faces. This is also observed when the relationship between faces in a given image is disrupted. However, in both of these cases, the maximally arousing face continues to significantly influence judgements of group arousal. The results are more equivocal for dominance, with a significance effect of the most dominant face on group dominance ratings in Experiment 2b (context removed and faces from the whole set randomly allocated to each face position in a given image) but not Experiment 2a (background context removed, original faces retained in original positions).

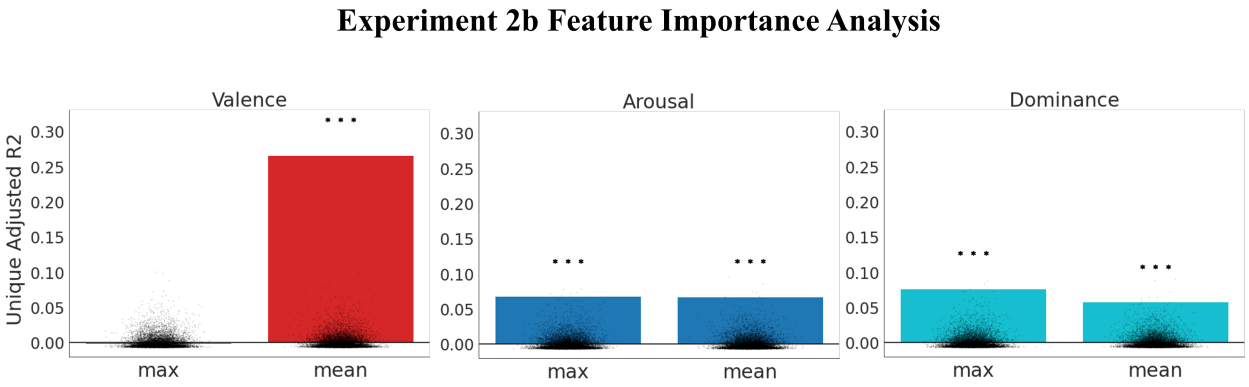


Figure 2.6: Drop column feature importance analysis for Experiment 2b. Drop column feature importance analysis was performed for each dimension of affect using max facial emotion (MAX) and mean facial emotion (MEAN) for each image as features and ratings of crowd affect for the *Scrambled Faces* stimuli as the dependent variable. Bars represent the unique variance explained by each regressor or 'feature'; this is calculated as the difference in adjusted R^2 for the model with both features versus that for the model with the feature in question removed. Scattered dots overlaying each bar show the permuted null distribution of the unique adjusted R^2 values. Mean facial emotion explained unique variance in *Scrambled Faces* ratings for all three dimensions (left panel: valence, center panel: arousal, right panel: dominance). The maximally arousing and the maximally dominant face also explained unique variance in the *Scrambled Faces* arousal and dominance ratings, respectively. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Control Analyses

As in Experiment 1, we conducted a control analysis to ensure that face size was not influencing our observed findings. We used the ratings on each dimension of affect for the largest face (MAX size) in each image to predict judgements of group emotion for both the *Faces without Context* stimuli (Exp 2a) and the *Scrambled Faces* stimuli (Exp 2b). The residuals were then used as dependent variables in an additional set of drop column analyses (see Figure A.4). The maximally arousing face continued to contribute significantly to the prediction of group arousal ratings for both the *Faces without Context* stimuli (Exp 2a) and the *Scrambled Faces* stimuli (Exp 2b) when the variance explained by the arousal ratings of the largest face were residualized out of these group ratings ($R^2_{adj} = .05$, $p = .002$; $R^2_{adj} = .03$, $p = .016$, respectively). The maximally dominant face also contributed significantly to the prediction of group dominance ratings for both the *Faces without Context* stimuli (Exp 2a) and the *Scrambled Faces* stimuli (Exp 2b) when the variance explained by the dominance ratings of the largest face were residualized out of the group ratings. For the *Faces without Context* stimuli (Exp 2a), removal of variance explained by the largest face led to the contribution of the maximally dominant face becoming more apparent ($R^2_{adj} = .02$, $p = .024$). For the *Scrambled Faces* stimuli (Exp 2b), the maximally dominant face also explained a significant amount of unique variance in group dominance ratings after the variance explained by the dominance ratings of largest face was regressed out ($R^2_{adj} = .09$, $p < .001$).

2.4 Discussion

Results from our first experiment indicate that when participants view natural static photographs of crowds, the perceived emotion of the crowd is predicted significantly both by the facial expressions of the individuals in the scene and by the background context. The extent to which a MEAN or MAX rule is used for judgements of crowd emotion appears to depend upon the specific emotional dimension being judged. In the case of valence, mean valence (as calculated across all faces in the image) explains unique variance in the ratings of crowd facial valence alongside background context. In contrast, for both arousal and dominance, the most arousing and most dominant face explain unique variance in ratings of crowd emotion whereas mean arousal and mean dominance regressors do not have a unique effect. These effects survived controlling for the affective ratings of the largest face in each image, suggesting that it isn't simply that photographers have focused on the most arousing or dominant face in each group.

In our second experiment, we conducted additional manipulations, both removing background context (Experiment 2a) and additionally removing the faces from a given image and replacing them with faces randomly drawn from other images (Experiment 2b). The latter manipulation ensures that the maximally emotional face, for any given dimension of affect, will not be consistently in a given position across images. In addition, it disrupts the relationship between faces in a given image, for example ensuring that there is no information from other faces such as gaze or head orientation that can cue the maximally emotional face. Under both these manipulations, group valence ratings continued to be predicted by the mean of the valence of faces within each image. For arousal and dominance ratings, a shift towards partial reliance on both MEAN and MAX facial emotion occurred. For arousal ratings, both the mean arousal of faces in a given image and the maximally arousing face in a given image predicted ratings of group arousal. This was observed regardless of whether ratings of the largest face in the image were regressed out before performing the drop column analysis or not. In the case of dominance, a stronger shift towards reliance on mean dominance was observed, with the maximally dominant face explaining unique variance in group ratings for Experiment 2b (context removed and face relationships disrupted) but only in Experiment 2a (context removed) after controlling for variance explained by the dominance ratings of the largest face in the image.

The findings from Experiment 1 indicate that across all affective dimensions (valence, arousal, dominance), background context explains unique variance in ratings of crowd emotion when considered alongside MEAN and MAX facial emotion. Meanwhile, the findings from Experiments 1 and 2 taken together suggest that the extent to which participants use the average affect shown by individuals in a group to inform their judgements of group affect depends both on the dimension of affect under consideration and upon the presence of background context. Contextual information exerted an especially strong influence on judgments of crowd dominance both in terms of direct prediction of crowd ratings and in modulating reliance on the most dominant face versus mean dominance of the faces in the group image. Arguably, dominance might be thought of as more of a relative judgement than arousal or

valence, where someone is more or less dominant than someone else. In addition, dominance might, to some extent, be contextually bound (dominance in a board room setting might be predicted by very different cues than in an athletics setting). This might potentially explain the stronger influence of context in judgements of dominance.

Ratings of crowd emotional valence (highly negative to highly positive) were strongly influenced by the average valence of faces in the image, and there was no unique contribution of the most positive (MAX) face, or indeed of the most negative (MIN) face. This held across Experiment 1, Experiment 2a, and Experiment 2b (see Figures 2.4, 2.6, 2.7, A.3, and A.5). This finding falls in line with prior summary statistic studies which have shown that mean emotion is a strong predictor of implicit or explicit evaluation of group emotion for valence (Han et al., 2021) and simple basic emotions such as 'happy' or 'sad' (Haberman & Whitney, 2007, 2009).

In contrast, across Experiments 1, 2a, and 2b, we observed a significant unique influence of the maximally arousing face upon judgements of arousal for groups of faces. This held for natural full images (Experiment 1), faces taken from these images with the background context removed (Experiment 2a), and similarly positioned groups of faces comprising faces randomly selected from across our image set (Experiment 2b). This unique influence of the maximally arousing face was greatest in Experiment 1 where no unique influence of mean arousal across faces was observed, and less, though still significant, in Experiments 2a and 2b. In these latter experiments, mean arousal also contributed to predictions of ratings of group facial emotion.

The attentional literature may potentially help explain this difference in findings for arousal versus valence. Specifically, findings from the attention literature suggest that highly arousing stimuli are preferentially processed leading to earlier break-through in binocular rivalry experiments, automatic orienting, slowed disengagement in dot-probe and anti-saccade studies, and increased pop-out in visual search paradigms (Bradley et al., 2012; Lundqvist et al., 2015; Ossenfort & Isaacowitz, 2021; Sheth & Pham, 2008; Vogt et al., 2008). This had led to the theory that emotionally arousing stimuli exert stronger bottom-up salience effects upon attentional competition (Mather & Sutherland, 2011). There is also some similar, though less extensive, evidence that this might also hold for stimuli indicative of dominance (Foulsham et al., 2011; Ohlsen et al., 2013; Ratcliff et al., 2011). In contrast, it has been argued that stimulus valence has a far lesser effect than stimulus arousal upon attentional competition (Mather & Sutherland, 2011) with findings from studies examining the relative effects of stimulus valence and arousal on visual search and attentional disengagement supporting this contention (Lundqvist et al., 2015; Ossenfort & Isaacowitz, 2021; Vogt et al., 2008). Heightened selective attention to the most arousing face in each image could result in enhanced perceptual processing and increased influence of this face on ratings of group arousal, potentially explaining our findings of a significant unique influence of the maximally arousing face upon judgements of arousal across all three experiments.

Our findings also suggest a potentially heightened reliance on the maximally arousing face and reduced reliance on mean arousal when background context is present (Experiment 1) relative to when it is absent (Experiments 2a and 2b). While we need to be cautious in

drawing conclusions across experiments, one possible explanation is that, in Experiment 1, background contextual cues as to which part of the image was most salient further focused selective attention on the most arousing face. In line with this, many studies have found that when there is a plethora of visual information, context can cue participants into attending to the most salient part of the scene (Chun, 2000; Brockmole & Henderson, 2006). There is also evidence that context's orienting role is especially apparent in high arousal scenes (Bradley et al., 2012; Calbi et al., 2017).

Future work incorporating eye-tracking would enable us to determine if background context alone can orient participants to the position of the most arousing face and if, in the absence of background context, participants also preferentially orient to the most arousing face, regardless of its position in the image. Here, it would also be of interest to vary viewing duration. In the experimental work reported here, each image was presented for one second. This is sufficient for more than one fixation. We used this duration of presentation to match that in some of the prior work on ensemble perception (Haberman & Whitney, 2011; Im et al., 2017) and out of a desire to simulate comfortable natural viewing. We did not control fixation given that the variability of the position of faces in natural images does not allow for easy selection of a single fixation point and out of concern for ecological validity. At shorter presentation times, participants might rely to a greater extent on summary statistics. However, it is also possible that early orienting to maximally arousing stimuli, and potentially maximally dominant stimuli, would exert an even stronger influence. This might also vary depending on the presence or absence of background context.

Our findings may be of pertinence not only for the field of psychology but also for the computer vision and machine learning literature. Machine learning and computer vision are being integrated into our everyday lives through social media and security-oriented face recognition systems. Many models have only been trained to categorize either singleton faces or groups of faces in isolation and not in context. Those models that categorize groups of faces in context (Dhall et al., 2016, 2017; Kosti et al., 2020; Mou et al., 2015; Rassadin et al., 2017) fail to simulate human ratings of crowd emotion and perform worse than models that categorize faces in isolation. The findings reported here may indicate how models can be refined to better mimic human perception of crowd emotion in uncontrolled, natural conditions.

In conclusion, our results show that in natural scenes, background context both directly influences judgements of crowd emotion and also influences the extent to which participants rely on the maximally emotional face as opposed to mean facial emotion, potentially through directing attentional resources to the most salient parts of the scene. It has been well noted that context contributes to the perception of single facial expressions (Aviezer et al., 2017; Chen & Whitney, 2019, 2020; Righart & de Gelder, 2008a, 2008b; Sasson et al., 2016) and that context directs attention to salient parts to the scene (Chun, 2000; Bradley et al., 2012; Brockmole & Henderson, 2006; Calbi et al., 2017). Here, we show that, when judging a group of faces, context contributes significantly to the perception of group facial affect and increases the influence of the maximally arousing and dominant faces within the scene upon the perception of group affect. In the absence of context, there is a greater reliance on

summary statistics across emotional dimensions. However, it is of note that the maximally arousing face significantly influences judgements of crowd arousal under all conditions. This is in line with emotional arousal influencing the perception of group facial affect by biasing of selective attention, and subsequent perceptual processing, in favor of the maximally arousing face.

Chapter 3

Cortical Representation of Mid-Level Face Features

3.1 Introduction

Face perception is an integral part to social relationships and how we view the world. It enables us to judge expressions and recognize people in our social networks. Previous studies have often used Haxby's model (Haxby et al., 2000) as a way of understanding how visual face information is represented in the cortex. Early level contrast information is processed in Early Visual Cortex or retinotopic regions (V1-V4), mid-level structural information in Occipital Face Area (OFA) (Gauthier et al., 2000; Haxby et al., 1999), high-level identity information in Fusiform Face Area (FFA) (Kanwisher et al., 1997), and emotion information in Superior Temporal Sulcus (STS) (Hoffman & Haxby, 2000). However, there have been several accounts that argue against a strict hierarchical model (Rossion, 2008; Duchaine & Yovel, 2015; Pitcher et al., 2014; Weiner et al., 2016) and against domain specificity in face processing regions (Baseler et al., 2014; Fox et al., 2011; Ganel et al., 2005; Said et al., 2011; Wegrzyn et al., 2015).

One possibility for this conflicting literature is that many of these previous studies have used extremely limited stimuli that have only a few conditions and only a few exemplars per condition. This results in a restricted stimulus set potentially accounting for inconsistencies where findings may be telling us something about the stimuli and task rather than what the brain is representing (Westfall et al., 2016). Here, we aim to understand where mid-level structural and textural information is represented in Occipital Temporal Cortex (OTC) by circumventing previous limitations in the field by using a varied, large number of natural real-world stimuli similar to that used by the machine learning field. By using this large number of stimuli, we can employ a multi-feature voxelwise encoding model framework (Naselaris et al., 2009; Stansbury et al., 2013; Huth et al., 2012; Huth et al., 2016) that captures many different abstractions of mid-level structural tuning to faces such as local face structure as annotated by dozens of fiducial points and higher-level abstractions such as face parts and

texture.

Another advantage of using these voxelwise encoding models is that we can look at model fit and face representation on an individual participant basis. This allows us to investigate individual differences in people who may have face processing deficits such as sub-clinical populations that exhibit Developmental Prosopagnosia traits or Autistic traits. Individuals with Developmental Prosopagnosia are unable to recognize face identity as well as individuals with normal perception (Duchaine & Nakayama, 2006), and it has been shown that there is both reduced grey matter (Garrido et al., 2009) and reduced face selectivity (Jiahui et al., 2018) in the fusiform gyrus. It is unclear, however, whether these abnormalities scale as a function of severity and whether mid-level structural and textural organization might be impaired in individuals with scores on the face recognition tasks such as the Cambridge Face Memory Task (Duchaine & Nakayama, 2006) indicative of mild to moderate Developmental Prosopagnosia. In the Autism literature, there have been many studies that show that participants with Autism tend to have preferential processing of local features (Lahaie et al., 2006, Jolliffe and Baron-Cohen 1997). The Enhanced Perceptual Functioning model builds on this literature arguing that Autistic individuals have enhanced low-level discrimination, preferential locally oriented perception, and utilize more posterior regions in complex visual tasks (Mottron et al., 2006). While there has been some conflicting literature focusing on whether there exist abnormalities in face parts processing in ASD literature (Golarai et al., 2006; Hernandez et al., 2009; Harms et al., 2010), there hasn't been a more global exploration of how individuals with Autism encode mid-level face features in visual processing areas. Our study aims to understand how individuals differ in mid-level face processing across OTC and whether this falls in line with previous theories in the literature.

To better understand the mechanisms underlying these deficits, we propose using several mid-level multi-feature encoding models to explore how different mid-level information is extracted from the brain and whether this differs as a function of severity of face processing deficits. Additionally, we can investigate how these different mid-level features map onto higher-level semantic information such as emotion and identity. This will allow us to understand where mechanisms of face processing break down for individuals with Developmental Prosopagnosia and Autistic traits.

For our study, our aims are as follows: 1) investigate how the brain organizes mid-level structural and textural information, 2) whether mid-level information is represented exclusively in OFA per Haxby's model, 3) whether there is differential tuning to mid-level features for those with Autistic traits (as measured by the Autism Quotient) and those who have Developmental Prosopagnosia traits (as measured by the Cambridge Face Memory Test), and 4) how mid-level face information is related to higher-level semantic information.

3.2 Methods

Participants

Participants were recruited from University of California, Berkeley and the greater San Francisco Bay Area community. 41 healthy participants completed all sessions of the study (16 male, 21 female, mean age = 24, age range = 18-39). 4 participants were excluded because of abnormal fixation patterns (fixating outside of the stimuli for more than 10% of the trials). Experimental procedures and recruitment were approved by the University of California, Berkeley Committee for the Protection of Human Subjects. All participants were required to be raised in the United States or a country with similar demographics (one participant was from the United Kingdom) so that they were exposed to faces that comprised similar demographics.

Stimuli

fMRI Experimental Design

A total of 936 natural images of faces were used. Images were tightly cropped so only internal and external features (such as hair) of faces were visible. 840 images were presented in the model estimation runs (70 images presented twice per run for 12 runs). 96 images were presented in the model validation runs (96 images presented twice per run). There were four validation runs in total, and all images were the same across each run. The order of the images was shuffled for each validation run. Images were displayed at a visual angle of 6° x 8° on the display screen.

Retinotopic Localizer

To localize retinotopic locations in the cortex, we used high contrast retinotopic mapping stimuli (Engel et al., 1994; Sereno et al., 1995; DeYoe et al., 1996). Stimuli consisted of a radial checkerboard (12 radial segments, 9 eccentric segments), reversing contrast at a frequency of 6 Hz up to a maximum eccentricity of 8.72° visual angle. Two apertures were used – (1) a clockwise or counter-clockwise 45° wedge and (2) an expanding or contracting ring. A single revolution of the aperture had a duration of 8 seconds. The stimuli were shown at 100% contrast inside the aperture and 2% contrast outside the aperture. Stimuli were presented in three runs – (1) wedge apertures cycling counter-clockwise for 8 revolutions, (2) wedge apertures cycling clockwise for 8 revolutions, and (3) ring apertures contracting (4 revolutions) and expanding (4 revolutions).

Dynamic Localizer

To localize functional face locations in the cortex, we used a dynamic localizer (Jiahui et al., 2018). The localizer comprised brief movie clips divided into five categories: faces, scenes,

bodies, objects, and scrambled objects. Video clip stimuli were presented at a visual angle of 21° x 16° visual angle. There were 10 unique stimuli within each category. Five runs were presented in total with each run consisting of the same stimuli. Stimuli blocks were shuffled across all runs.

Experimental Procedure

fMRI Experimental Task

There were 16 runs (12 estimation and 4 validation) total across two fMRI sessions (2-10 days apart). In each session, participants were presented with 3 estimation runs followed by a validation run and then another 3 estimation runs followed by a validation run. For all runs, participants were shown a fixation box (white box outlining where the stimuli would be) for 10 seconds at the beginning of each run and 16 seconds at the end of each run. For the estimation runs, stimuli were presented 2 seconds at a time with 8 seconds of fixation box occurring every 46-47 faces. For validation runs, stimuli were also presented 2 seconds at a time with 8 seconds of fixation box occurring every 64 faces. Estimation runs lasted approximately 328 seconds and validation runs lasted approximately 432 seconds. All stimuli were outlined with the fixation box. Participants were instructed to freeview within the fixation box, but not outside of it. Participants who had fixations outside the fixation box for more than 10% of trials were excluded ($n = 4$). A simple memory task was used for each run where participants were asked to press '1' when seeing a face for the first time and '2' when seeing a face for the second time within run.

Retinotopic and Dynamic Localizer

Participants were asked to fixate at the center of the screen on a fixation cross across all runs for both the retinotopic and dynamic localizer. For the retinotopic localizers, runs were preceded with 5 seconds of fixation. Each retinotopic run lasted 288 seconds. For the dynamic localizer, runs were preceded with 12 seconds of fixation. Runs were divided into 10 stimuli blocks. Stimuli were presented for 1.5 seconds with the fixation cross being presented 0.5 seconds following the stimuli. After every 5 blocks, there was a 12 second fixation cross. Each block contained 5 trials of stimuli from the same category. Each dynamic localizer run lasted 156 seconds.

Behavioral Face Perception Tasks and Questionnaires

A behavioral session preceded the fMRI sessions. During the behavioral session, participants were administered several questionnaires: Autism-Spectrum Quotient Questionnaire or AQ (Baron-Cohen et al., 2001b) to measure severity of Autistic traits and the Cambridge Face Memory Test or CFMT (Duchaine & Nakayama, 2006) to measure severity of Developmental Prosopagnosia traits. Those with higher AQ scores had more severe Autistic traits, and those who were low on CFMT had worse face recognition ability.

fMRI Data Acquisition

fMRI and MRI data were acquired on a Siemens 3T Tim Trio MRI system using a 32-channel phased-array head coil. 33 participants wore custom headcases to reduce participant motion during scanning (CaseForge, Inc). For 4 participants, custom headcases were not available at the time of scanning, and standard padding was used instead. Functional images were acquired with a gradient echo EPI sequence (TR = 1,000 ms, TE = 28.4 ms, flip angle = 57°, slices = 45, resolution = 2.5 mm isotropic) with factor 4 parallel multiband acceleration (Moeller et al., 2010).

B_0 field maps were acquired in-plane in each run to estimate and correct distortions due to field inhomogeneity (TR = 400 ms, TE1 = 5.19 ms, TE2 = 7.65 ms, resolution = 2.5 mm isotropic). Finally, a T1-weighted whole-brain anatomical image was acquired in order to reconstruct the cortical surface and anatomically localize functional data (MPRAGE, TR = 2,300 ms, TE = 2.96 ms, TI = 900 ms, flip angle = 9°, 208 slices = 208, resolution = 1 mm isotropic).

Cardiac and respiratory cycles were recorded during scanning with a pulse oximeter and respiratory bellows (BioPac Systems MP150, BioPac Systems Inc.). Participants responded via a 4-button MRI-compatible button box. Eyetracking was acquired for participants using Eyelink 1000 tracking the left eye using a sampling rate of 1000 Hz.

Data Pre-Processing

Functional Pre-Processing

fMRI data was pre-processed with FSL (Jenkinson et al., 2012) and custom code. Functional volumes with abnormally high variance, attributed to image acquisition artifacts or large participant motion events, were identified and replaced with the Tsvarana tool (https://github.com/IvanAlvarez/Tsvarana). Voxelwise variance was calculated between each volume and the median within-run volume and normalized by the mean voxel intensity across the timeseries. Volumes with mean normalized variance greater than 5 were removed and replaced with the mean of the first preceding and following volumes that did not violate the variance threshold. This process was repeated until no volumes displayed mean normalized variance above 5. Runs with greater than 10% of volumes removed were excluded from further analysis. fMRI data were then registered to the first volume acquired for a given participant, while simultaneously performing spatial correction for magnetic field inhomogeneities as estimated by B_0 field maps (Jenkinson and Smith, 2001; Jenkinson et al., 2002; Greve and Fischl, 2009), followed by brain extraction (Smith, 2002). Timeseries were denoised by performing independent component analysis (Beckmann and Smith, 2004) on segmented non-brain voxels, retaining the 10 components with highest variance explained, and performing a regression on brain voxels, where the residual timeseries were retained. Data was then subjected to a high-pass Savitzky-Golay temporal filter (window = 120s). Finally, cardiac and respiratory signals were modeled and removed with FSL PNM (Brooks et al., 2008).

Structural Pre-Processing

T1-weighted anatomical scans were analyzed with FreeSurfer (Fischl, 2012). Anatomical data underwent removal of non-brain tissue, Talairach transformation, intensity normalization and automated classification of white and gray matter tissue in order to reconstruct the cortical surface (Dale et al., 1999; Fischl et al., 1999). Following semi-automated correction for tissue misclassification and topological errors, a flattened cortical representation or flatmap was obtained by placing five relaxation cuts on the cortical mesh and applying FreeSurfer's automated flattening algorithm. The resulting cortical surface reconstructions were imported into Pycortex (Gao et al., 2015), and used to display functional activity maps.

ROI Localizer

Visual category localizer data were analyzed with FSL FEAT (v6.0) (Woolrich et al., 2001, 2004). BOLD data were spatially smoothed (Gaussian kernel FWHM = 3 mm), z-scored, pre-whitened, and subjected to a run-wise GLM with five stimulus categories as explanatory variables, and motion regressors as nuisance variables. Resulting contrasts were summarized across runs with a fixed-effects model, and outputs displayed on the cortical surface to aid manual delineation of category-selective regions.

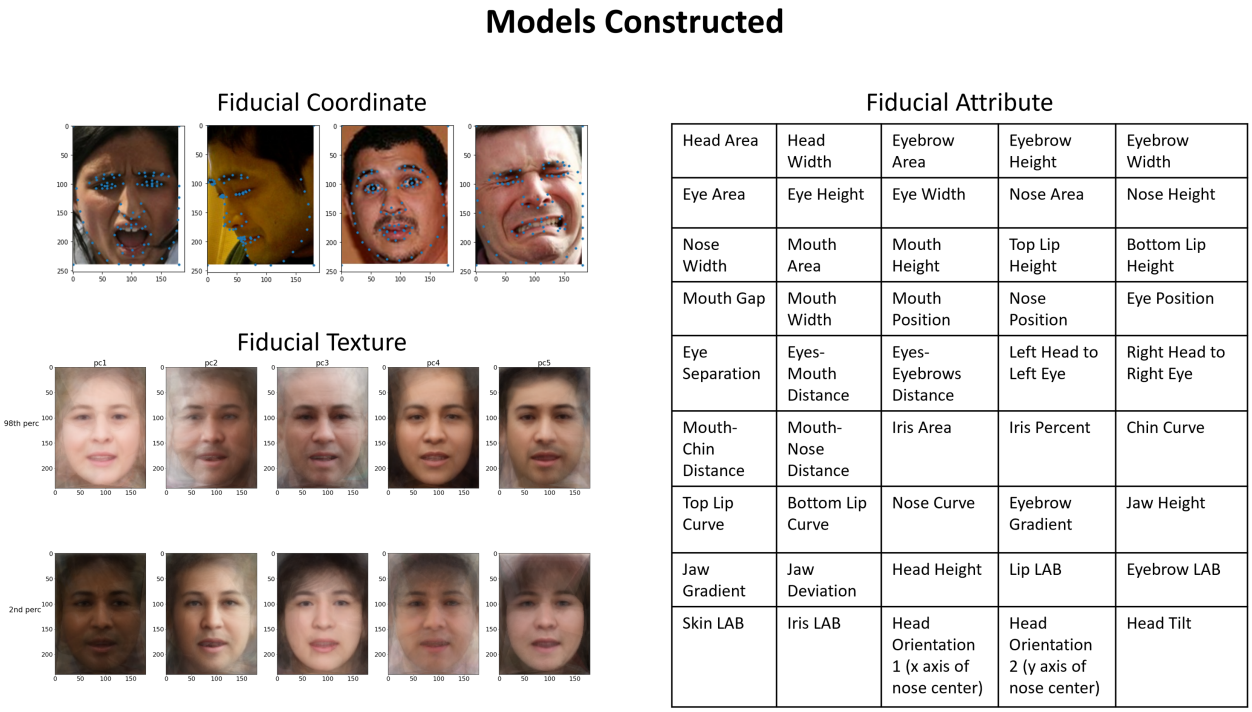
Retinotopic mapping data were analyzed with the traveling wave phase-encoded approach (Engel et al., 1997). Wedge and ring apertures were analyzed independently to derive polar angle and retinotopic eccentricity estimates, respectively. Voxelwise timeseries were Fourier-transformed, averaged across runs of opposite polarity, and the phase of the signal at the stimulation frequency (0.125 Hz) taken as the signal of interest. Outputs were displayed on the cortical surface to aid manual delineation of retinotopic areas.

Combining the two approaches above, we delineated the following cortical areas based on their modelled responses to localizing stimuli: visual areas V1,V2, V3, hV4, LO-1, and LO-2, category-selective occipital face area (OFA), fusiform face area (FFA), posterior, medial and anterior portions of face-selective superior temporal sulcus (pSTS, mSTS, aSTS), anterior temporal face patch (ATFP), parahippocampal place area (PPA) and the extrastriate body area (EBA). For simplicity, we grouped visual areas together to form the following ROIs: retinotopy (V1-V4) and STS (pSTS, mSTS, aSTS). We also delineated a larger area, occipital-temporal cortex (OTC), that followed the parieto-occipital sulcus, intraparietal sulcus, and a portion of the postcentral sulcus to the Sylvian fissure. All analyses were constrained within OTC.

fMRI Data Modeling

Models Constructed

Face stimuli were labeled via a semi-automatic process using Face++ (Zhou et al., 2015). We used the Face++ landmarks model (83 x, y points) to label key features of the face such as jaw line, eyebrows, eyes, nose, and mouth. These points were then manually adjusted



Head Area	Head Width	Eyebrow Area	Eyebrow Height	Eyebrow Width
Eye Area	Eye Height	Eye Width	Nose Area	Nose Height
Nose Width	Mouth Area	Mouth Height	Top Lip Height	Bottom Lip Height
Mouth Gap	Mouth Width	Mouth Position	Nose Position	Eye Position
Eye Separation	Eyes-Mouth Distance	Eyes-Eyebrows Distance	Left Head to Left Eye	Right Head to Right Eye
Mouth-Chin Distance	Mouth-Nose Distance	Iris Area	Iris Percent	Chin Curve
Top Lip Curve	Bottom Lip Curve	Nose Curve	Eyebrow Gradient	Jaw Height
Jaw Gradient	Jaw Deviation	Head Height	Lip LAB	Eyebrow LAB
Skin LAB	Iris LAB	Head Orientation 1 (x axis of nose center)	Head Orientation 2 (y axis of nose center)	Head Tilt

Figure 3.1: Mid-Level Models Constructed. The top left panel shows the 83 points on example faces in which we used the x, y pixel location for these points as features for the Fiducial Coordinate model. The bottom left panel shows the 98th percentile and 2nd percentile face loadings for the first 5 regressors for the Fiducial Texture model which was constructed by morphing all faces to the average face template and then doing PCA on the pixels to create 100 PC features. The right panel shows the defined attributes for the Fiducial Attribute model that were calculated using the 83 Fiducial Coordinate points.

using WebMorph (DeBruine, 2018) to ensure that the coordinate points were precisely on the key features of the face.

Three models were created from these x, y coordinates. The Fiducial Coordinate model simply used the raw x, y coordinates in pixel space as features. The Fiducial Attribute model made linear and non-linear transformations to aggregate information about facial attributes (i.e., head area, head width, etc.) based off the attributes Vernon and colleagues used (2014). The Fiducial Texture model used an approach similar to that of Chang and Tsao (2017) to investigate textural face information. All face stimuli were morphed to the average fiducial coordinate template and PCA was performed on pixel values of the morphed faces. The first 100 PCs were kept and the left singular matrix (image loadings onto PCs) was used as features for the model.

Model Estimation

Model estimation was performed using ridge regression on the 12 estimation runs BOLD time series for each voxel. For each participant, a feature matrix was constructed for each model

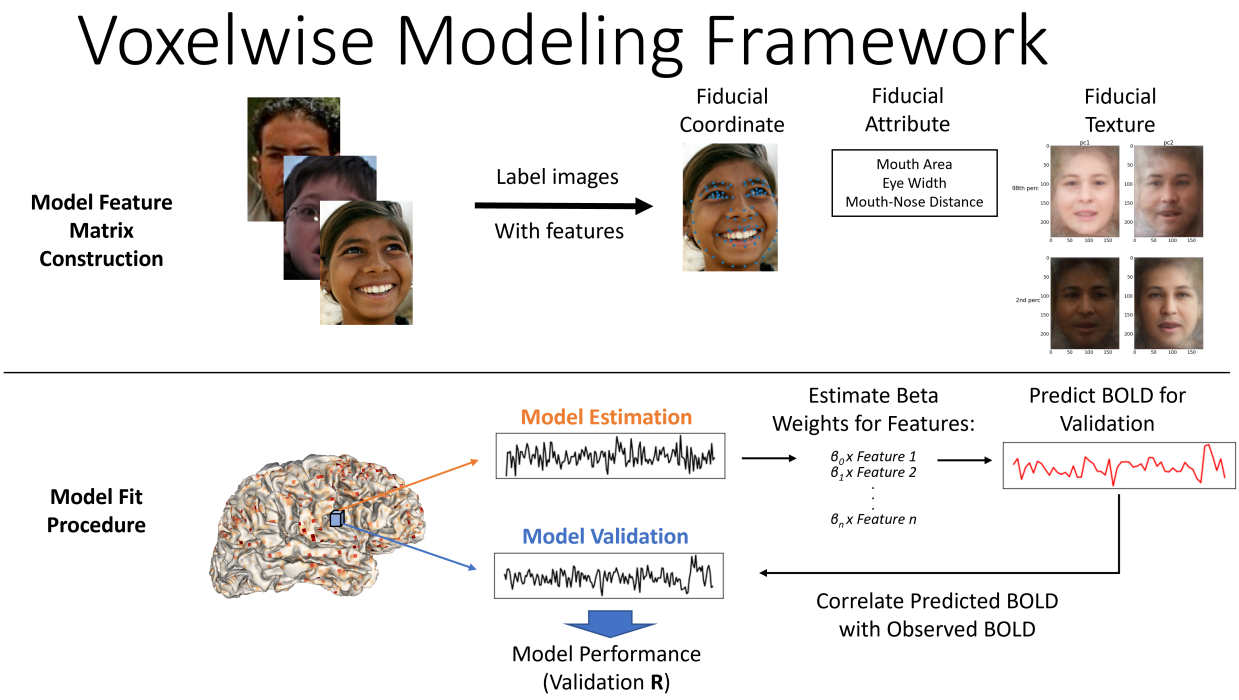


Figure 3.2: Voxelwise modeling framework. In the top panel, we show the process of constructing the model feature matrix. 936 images were labelled with a separate set of features for each Fiducial model to create the model feature matrix. In the bottom panel, we show model fit procedure. Images were divided into Model Estimation and Model Validation sets. For each voxel, we fit a ridge regression model on the BOLD time series observed for the Model Estimation set and estimate the beta weights for each of the features in the model feature matrix. We then multiply those beta weights with the feature matrix for the Model Validation set to get a predicted BOLD signal for Model Validation. Finally, to get a metric of model performance, we correlate the predicted BOLD signal with the actual observed BOLD signal to get a validation correlation value.

using the features (i.e., x, y coordinates) for each stimulus in the estimation runs. Features were then convolved with a finite impulse response (FIR) filter using delays of 2-4s, 4-6s, 6-8s, and 8-10s. For each participant and for each voxel in OTC, we performed cross validation to select a L2 parameter. A range of 20 lambdas scaled logarithmically from 10^{-2} to 10^{3} were tested. Bootstrap cross validation was performed (30 iterations). For each lambda and for each iteration of cross validation, each model was fit on 10 of the 12 estimation runs with the remaining 2 estimation runs held out for prediction. We then took the highest mean correlation score across all voxels and across all participants to obtain a single L2 lambda parameter for all voxels and all participants. The final feature weights were constructed by re-fitting the ridge regression model using the single L2 lambda across all estimation data.

Model Validation

For each participant and for each voxel, the ridge model fit on the model estimation BOLD time series was used to predict the BOLD time series for the 4 validation runs. As with model estimation, a features matrix was constructed for each model using the features for each stimulus in the validation runs. This matrix was then convolved with 4 FIR bins (4s, 4-6s, 6-8s, and 8-10s). The feature weights that were estimated during model estimation for each voxel were then multiplied with the validation feature matrix to obtain a BOLD time series prediction for that voxel. We then correlated these predicted BOLD time series with the observed BOLD time series for the validation data to obtain a measure of model fit. Permutation testing was then performed (see Statistical Analysis) to see whether the model correlation values were significant.

Statistical Analysis

Model Prediction Significance Testing

To assess model significance, we performed permutation tests on the model validation correlation. For each permutation, the rows of the validation feature matrix were shuffled and multiplied by the model estimation weights. This was done 5000 times and a one-tailed significance analysis was performed for each voxel to produce a null distribution of correlation scores. Significance was determined for each voxel by generating a p-value by comparing the true correlation with the null distribution correlation distribution. We used a threshold of $p < .05$.

Visualization of Group Model Performance in fsaverage

To better visualize all 37 participants' model performance in a common space, we transformed each participant's vertex space (2D mapping of voxel space) into Freesurfer's fsaverage space – a surface-based template. We then plotted several model performance metrics onto this space. To get the mean correlation for each vertex, we transformed each participant's validation correlations into Fisher Z values and averaged the Fisher Z values across participants for each vertex in OTC. We then re-transformed the mean Fisher Z for each vertex back into R correlation values. To get group level significance maps, we labeled each vertex as an array of 0s and 1s with 1 representing that this vertex was significant for a given participant and 0 meaning that this vertex was not significant using a significance level of $p < 0.05$ obtained using the permutation significance methods discussed in the previous section. For each vertex, we had an array length that equaled the number of participants ($n = 37$). We then performed a one tailed binomial t-test with a chance of success equal to .05 and obtained group level p-values for each vertex. To correct for multiple comparisons, false discovery rate (FDR) correction using the Benjamini-Hochberg procedure was applied for voxels in OTC ($q[FDR] < .05$). We also visualized differences in model performance as a function of AQ and CFMT in the fsaverage template across OTC. For each vertex, the transformed

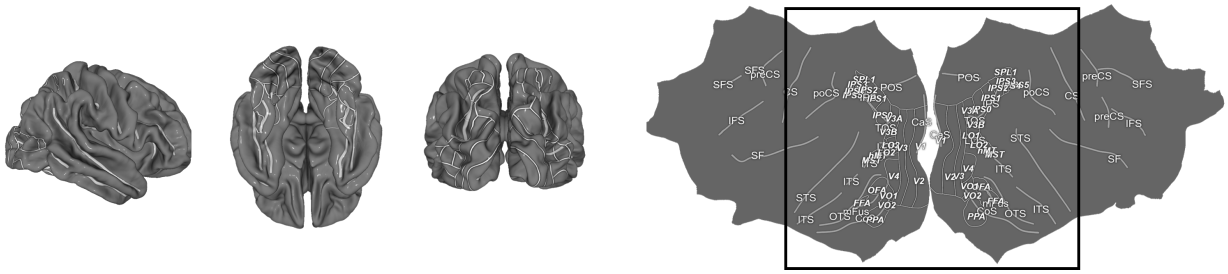


Figure 3.3: Freesurfer fsaverage template maps. Fsaverage is a surface-based template created by averaging anatomical maps. On the left we have three viewpoints (sagittal, transverse, and coronal, respectively) of the volumetric 3D fsaverage cortex. On the right, we show a flatmap representation of the 3D volumetric fsaverage brain. Results will be displayed within the space of the black bounding box which zooms in on OTC.

validation Fisher Z values were correlated with the individual participant's score obtained from the AQ and CFMT task.

3.3 Results

Group Significance Maps and Mean Validation Correlation

To examine model performance on a group level and better visualize all 37 participants' data in a common space, we transformed each participant's cortical space into the fsaverage template space. For each model, binomial t-tests with a significance threshold of $q[FDR] < .05$ were performed to find which voxels had significant model fit across participants (see Methods for details). For each of the three mid-level models tested, we found that voxels with significant model fit were mainly clustered in retinotopic regions and extended into intermediate-higher face processing areas such as OFA and FFA (Figure 3.4A, top panel). Although transforming participants' structural space into a common template can help us visualize model fit, individual functional ROIs may not overlap perfectly. Thus, we also looked at summary statistics of model fit within individuals' functionally defined ROI (Figure 3.4B) and calculated the average percentage of voxels that had significant model performance ($q[FDR] < .05$). The highest number of voxels with significant model fit were in retinotopy across all participants.

These findings suggest that retinotopic areas represent mid-level face features. This contrasts with traditional theories arguing that V1-V4 only process low-level contrast information (Bruce & Young, 1986; Haxby et al., 2000). However, using natural images may present us with confounds that aren't present for highly controlled posed stimuli. In particular, our mid-level features may be highly correlated with low-level statistical information in the images. If this is the case, it could indicate that low-level information is driving our model performance. To control for low-level information, we fit a low-level visual Gabor model (Nishimoto et al., 2012) and obtained the residuals. The mid-level models were then re-fit

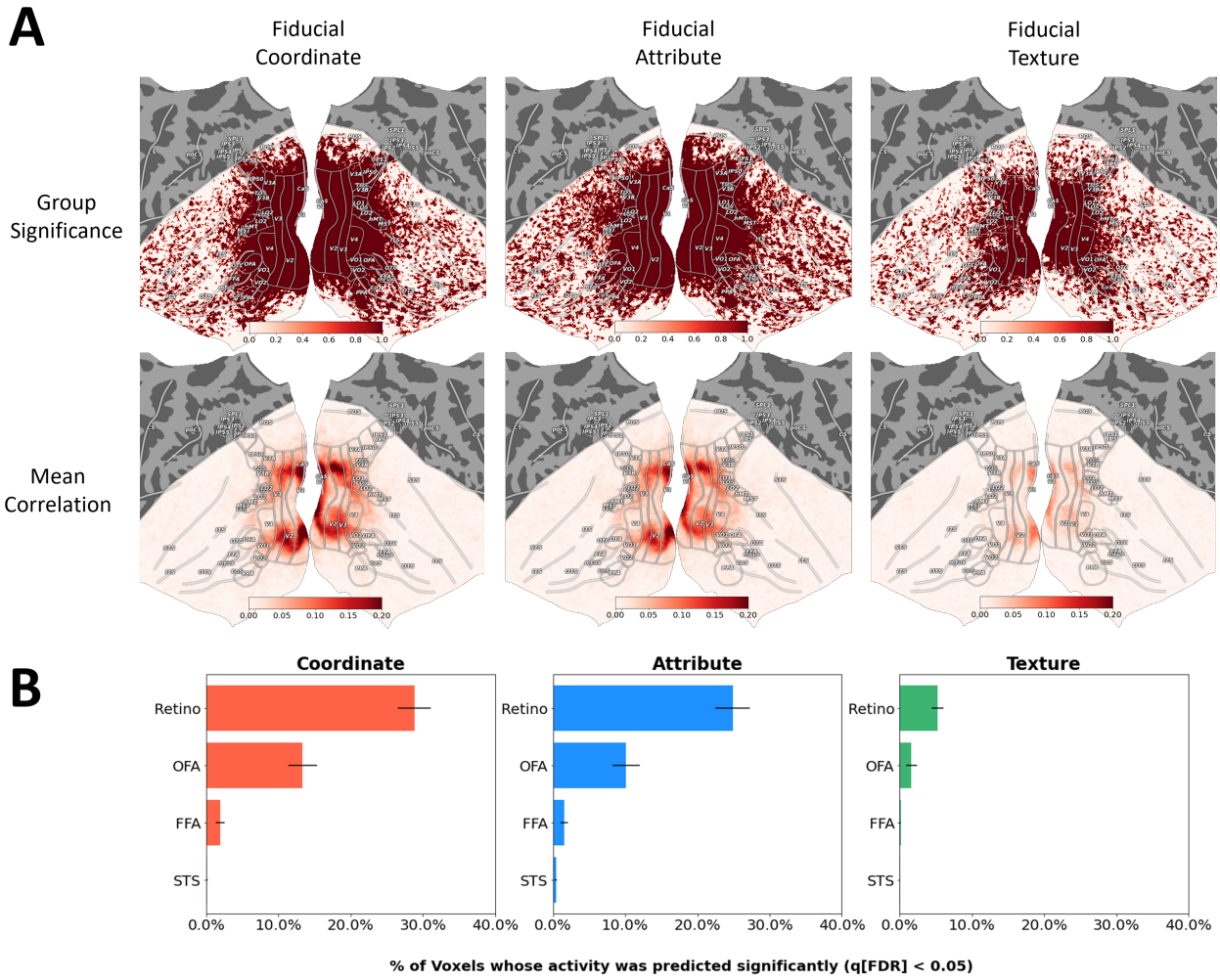


Figure 3.4: Group mid-level model performance in fsaverage. (A) From left to right, group-level model performance for the Fiducial Coordinate, Fiducial Attribute, and Fiducial Texture model plotted on fsaverage maps. Grey regions represent areas that were not used in analyses (regions outside OTC). In the top panel, we have corrected ($q[FDR] < .05$) group significance maps for model validation performance that were calculated using binomial t-tests. Uncorrected binary individual significance maps (as calculated by permutation tests, see Methods) were transformed into the common fsaverage space. A binomial t-test with a success threshold of .05 was then conducted across transformed individual maps to ascertain whether a voxel was significant on a group basis. FDR correction was then performed across OTC for the group map. Red indicates that the voxel was significant after FDR correction (across OTC) for the group. The bottom panel shows the group validation correlation across participants calculated by converting individual correlations for each vertex into Fisher Z values. Those Fisher Z values were averaged and then re-transformed into a correlation value. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction. (B) Proportion of voxels whose activity was significantly predicted by mid-level models ($q[FDR] < .05$) in Retino (V1-V4), OFA, FFA, and STS ROIs. Bars represent the average percentage of voxels that were significant as calculated by the permutation tests in each ROI (FDR corrected within ROI) across participants. Error bars represent standard error across participants.

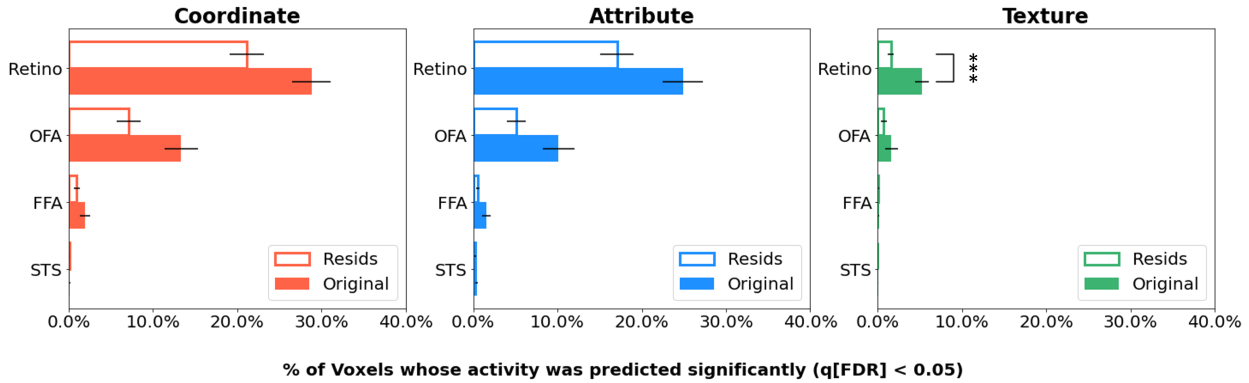


Figure 3.5: Comparison of original mid-level model performance with mid-level models fit on Gabor residuals. Proportion of voxels whose activity was significantly predicted by original mid-level models ($q[FDR] < .05$) and mid-level models with low-level information regressed out in Retino (V1-V4), OFA, FFA, and STS ROIs. Bars represent the average percentage of voxels that were significant as calculated by the permutation tests in each ROI (FDR corrected within ROI) across participants. Error bars represent standard error across participants. Stars represent significant difference between the original and residual model (t-tests, Bonferroni corrected across ROIs: $*p < .05$, $**p < .01$, $***p < .001$).

on those residuals to ensure that low-level information was regressed out. When controlling out for low-level contrast information, there was still high model performance across all models in retinotopic areas (Figure 3.5, Figure B.8). T-tests were then performed to investigate whether there were differences between the percentage of significant voxels in original mid-level models and residual models within ROIs. We found that there was only a significant difference in model performance in retinotopy for the Texture model ($p < .001$, Bonferroni corrected). Together, our results show that retinotopic areas represent mid-level features (especially structure) extremely well independent of low-level statistical dependencies.

Individual Differences in Model Fit

As a first step, we sought to examine where each model fits on a group basis within OTC. However, we know on a perceptual level that individuals have differences in face processing. Are there neural correlates of these perceptual idiosyncrasies? To investigate whether individuals have differential tuning to mid-level face features, we looked at correlations of model performance and scores on AQ and CFMT. To visualize this in a common space, individuals' cortex maps were again transformed into fsaverage space. For each voxel, model performance (Fisher Z transforms of validation correlation) was correlated with AQ and CFMT.

For CFMT, our results show visually in fsaverage that participants with lower CFMT scores (worse at face recognition) have somewhat stronger correlations in retinotopic areas for the Fiducial Texture model (Figure 3.6, bottom panel). However, when relating number of significant voxels within each ROI (while controlling for ROI size) and CFMT, there were no significant correlations (Figure 3.7).

For all mid-level models, we observe that participants who have higher AQ scores (more

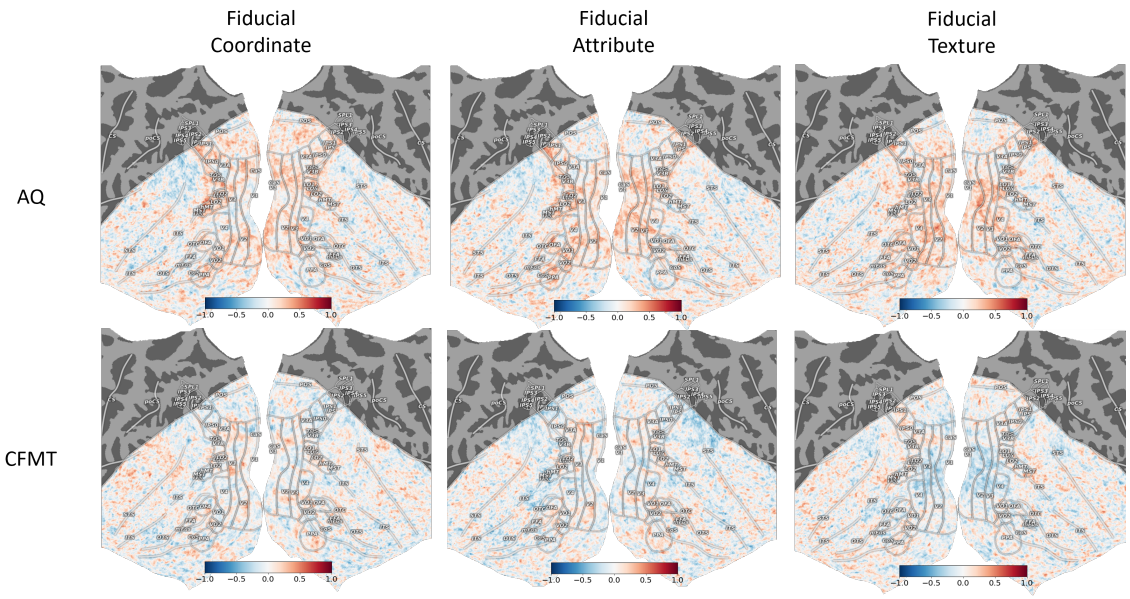


Figure 3.6: Correlation plots of model fit and AQ scores (top panel) and CFMT scores (bottom panel) in fsaverage. For each vertex, we correlated the model fit (Fisher Z transform of validation correlation) with an individual differences measure. In the top panel, positive red correlations mean that participants with higher AQ scores (more severe Autistic traits) had better model performance. Negative blue correlations indicate that participants with lower AQ scores (less Autistic traits) had better model performance. In the bottom panel, red indicates that participants with higher CFMT scores (better at face identity) had better performance whereas blue indicates that participants with lower CFMT scores (worse at face identity) had better model performance. We observed that in all three models, participants with more severe Autistic traits had better model performance in retinotopic areas.

severe Autistic traits) have higher model performance across retinotopic areas (Figure 3.6, top panel). This pattern is seen in OFA for the Fiducial Attribute and Texture model with high model performance extending to FFA for the Fiducial Texture model. To quantify these differences, we observed in the fsaverage maps, we took the number of significant voxels ($p < .05$, uncorrected) in an ROI for each participant and did a partial correlation with AQ while controlling for ROI size. Reinforcing what we observed in the fsaverage group maps, there were significant correlations with AQ scores and model performance in retinotopic areas across all models (Coordinate: $R_{partial} = .36$, $p = .029$; Attribute: $R_{partial} = .38$, $p = .022$; Texture: $R_{partial} = .43$, $p = .010$). There were no significant correlations across any other ROIs (Figure 3.7), and these results hold even when controlling for low-level features (Figure B.9). We conclude that individuals with more severe Autistic traits have enhanced representation of mid-level features. A similar pattern is observed for the low-level Gabor model (Figure B.10) where we see a significant correlation between Gabor model performance in retinotopy and AQ ($R_{partial} = .39$, $p = .020$) suggesting that this enhanced local processing not only occurs in mid-level features, but low-level contrast features as well.

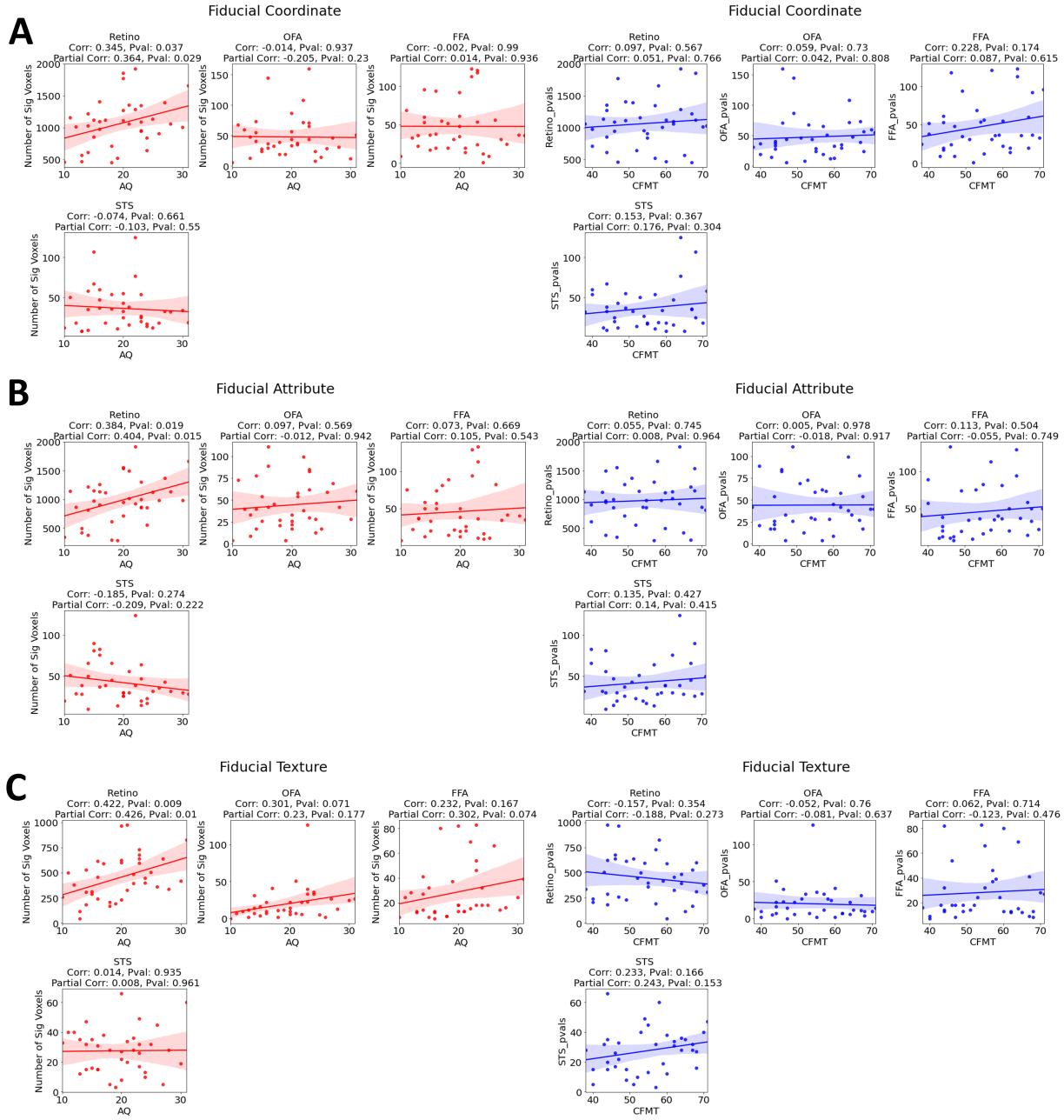


Figure 3.7: Mid-level model performance correlated with AQ (red) and CFMT (blue). For each participant, the number of significant voxels in each ROI were calculated from permutation tests ($p < .05$, uncorrected) for each model (top panel A: Fiducial Coordinate, middle panel B: Fiducial Attribute, bottom panel C: Fiducial Texture) and correlated with AQ and CFMT. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation as well as the partial correlation and partial p value when controlling for ROI size. We show that across all models that the number of significant voxels in retinotopy is significantly correlated with AQ when controlling out for ROI size.

Relating Mid-Level Models to High-Level Semantic Models

In the previous sections, we demonstrate that mid-level models are well represented in retinotopic areas, and model performance in retinotopy is correlated significantly with Autistic traits. While these results are useful in understanding how structure and texture are represented across cortex and whether these representations vary with traits of interest, it is important to also understand how mid-level models relate to higher level semantic models. This can give us insights into how mid-level model features transform into higher-level semantic representations. Furthermore, we can understand how idiosyncratic mid-level model performance propagates further down the visual processing stream affecting how individuals extract semantic information.

To address these questions, four high-level semantic models (Lower Emotion, Lower Identity, Higher Emotion, Higher Identity, see Table B.1) were created. Semantic terms were separated into emotion and identity features (44 and 45 features each, respectively) and further broken down into low-level semantic features (i.e., emotion: eyes closed, squinting; identity: bald, big eyes) and high-level semantic features (i.e., emotion: adoring, sad; identity: Asian, female). The goal of breaking semantic terms into higher vs lower features was to examine whether mid-level models mapped onto specific abstractions of semantic features rather than all semantic features as a group. We expected mid-level models would relate most closely to low-level semantic features. Features were labeled for each image by 3 in-lab raters and the agreement of two raters was used as the final value (0/1). Model fit procedures were identical to that described in the Methods section.

To understand how mid-level model features transform into higher-level semantic representations, we correlated the percentage of significant voxels ($q[FDR] < .05$) for the mid-level models in retinotopy with the percentage of significant voxels ($q[FDR] < .05$) for the four semantic sub-models in OFA and FFA. Texture model performance in retinotopy was significantly negatively correlated with the Higher Identity model in OFA ($R = -.47$, $q[FDR] = .037$), and all three mid-level models were significantly negatively correlated with the Higher Identity model in FFA (Coordinate: $R = -.47$, $q[FDR] = .014$; Attribute: $R = -.47$, $q[FDR] = .014$; Texture: $R = -.47$, $q[FDR] = .014$), see Figure 3.8. This negative relationship might indicate that participants who rely on mid-level representation in retinotopy have trouble extracting higher level representation of faces later on in the visual processing stream. Specifically, participants may be unable to convert mid-level structure and texture into higher identity information related to race, identity, and gender. This suggests an issue in bottom-up, feed-forward processing and thus begs the question of whether those with higher Autistic traits will also have poorer higher identity representation in OFA and FFA. In the next chapter of this dissertation, I will address this question as well as explore individual differences in semantic model performance in relation to face identity recognition.

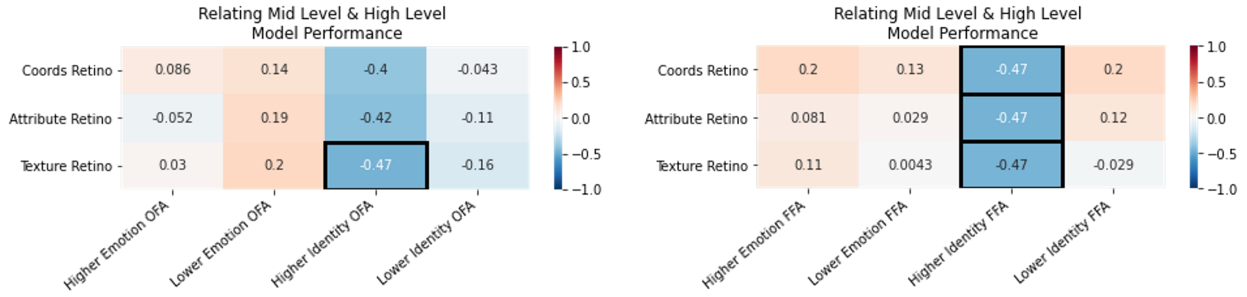


Figure 3.8: Correlation matrix of model performance in retinotopic regions for mid-level models and high-level model performance in OFA (left panel) and FFA (right panel). Model performance was calculated as the percentage of voxels that were significant ($q[FDR] < .05$). Each cell shows the correlation value between a given mid-level model with a given high-level model. Black boxes denote significant correlation values ($q[FDR] < .05$). There was a significant negative relationship between the Texture model performance in retinotopy and the Higher Identity model performance in OFA. All mid-level model performance in retinotopy had a significant negative correlation with Higher Identity model performance in FFA.

3.4 Discussion

Through a voxel-wise encoding modeling framework, we were able to see how mid-level face models (Fiducial Coordinates, Fiducial Attributes, Fiducial Texture) were represented in OTC on a group basis and how this representation differed in individuals. Across participants, we found that mid-level models performed the best in retinotopic areas. While there may have been a significant decrease in performance for the Texture model when fit on the Gabor model residuals, there was still high representation of this model in retinotopic regions on a group basis (Figure B.8). The significant difference in model performance may have been attributed to a strong relationship between the Texture and Gabor model features that was not present in our other mid-level models. Previous findings in the field have shown that texture is represented in retinotopic regions (Freeman et al., 2013; Kim et al., 2022) and that it can often be inferred from Gabor wavelet filters (Arivazhagan et al., 2006; Clausi & Jernigan, 2000; Fogel & Sagi, 1989). Thus, texture may be different than other types of mid-level representations and may have a more direct relationship with low-level representations.

For the mid-level models that included structural features (Fiducial Coordinate and Fiducial Attribute), model performance was the highest in retinotopic regions, and there were no significant differences in model performance even when controlling for low-level features. This contradicts the Haxby model (Haxby et al., 2000) demonstrating that mid-level face feature representation is not exclusive to OFA. These results further corroborate findings in previous literature showing that face processing isn't strictly hierarchical (i.e., there may be feed-forward connections from Early Visual Cortex to FFA), and OFA isn't a necessary entry point for recognition of higher-level features in faces (Rossion, 2008; Duchaine & Yovel, 2015; Pitcher et al., 2014; Weiner et al., 2016). Several studies in the field of object and scene processing show that anterior regions in Early Visual Cortex (V2-V4) may code for mid-level features such as shape, depth, and color (Pierce et al., 2015; Roe et al., 2012; Welchman et

al., 2005). Our results in the context of this literature suggest that face processing in Early Visual Cortex may share similarities to object and scene processing streams with mid-level representation being abstracted in retinotopic regions.

In this study, we first sought to understand how mid-level features are represented in OTC on a group level. As a second step, we aimed to investigate idiosyncratic differences in mid-level representation and whether this representation differs as a function of Autistic traits or performance on a face task that measures Developmental Prosopagnosia trait severity (CFMT). Although we observed some qualitative differences in model performance relating to Developmental Prosopagnosia traits, we ultimately did not find a significant relationship between performance on the CFMT and model performance in face ROIs.

On the contrary, we qualitatively and quantitatively observed that those with more severe Autistic traits had better model performance in retinotopic areas. That is, performance in all mid-level models had significant positive correlations with AQ in retinotopic areas. Our results fall in line with previous evidence in the Autism literature showing better local processing in vision for Autistic individuals (Chen et al., 2012; Dakin & Frith, 2005; Jemel etl al., 2006; Koldewyn et al., 2013; Mottron et al., 1999; Robertson et al., 2013). These results support the Enhanced Perceptual Functioning model (Mottron et al., 2006) theory which hypothesizes that people with severe Autistic traits have better local processing of features in more posterior regions in the brain. This model also claims that enhanced local perception is not a product of weaker global processing. One could argue, nonetheless, that our results may align with another theory in the Autism literature - the Weak Central Coherence theory (Happe & Frith, 2006). This theory suggests that local biases result from an inability to integrate information to form a more abstract global representation, coupling poor global processing with enhanced local processing. In the next chapter, we attempt to disentangle whether individuals with Autistic traits have poorer global perception, or weaker coherence, as a result of better local processing.

As the last step to understanding mid-level face representation in the visual processing stream, we investigated the relationship between mid-level model performance in retinotopy and high-level model performance in OFA and FFA to understand how biases in mid-level processing can affect higher-level face perception downstream. This can give us insights into how mid-level model features transform into higher-level semantic representations. We found that better model performance in retinotopy for all mid-level models was significantly related to worse semantic model performance for higher identity features (i.e., race, ethnicity, gender, age) in FFA. That is, individuals who have who have better performance of mid-level models in retinotopy may have difficulty extracting higher level identity information further up in the processing stream. This account supports a breakdown in feed-forward, or bottom-up, processing. However, there may be other theories as to why we see this negative relationship between retinotopy and high-level regions.

One theory stems from literature explaining predictive coding in the context of Autism. Predictive coding (Rao & Ballard, 1999) suggests the brain is constantly using priors to make inferences about sensory inputs. Lower-level regions receive predictions from higher-level regions in the brain in a top-down manner. If there is a mismatch with regards to

lower-level representation and the predictions, this residual error will be propagated through feed-forward connections from the lower-level regions to the higher-level regions. Van de Cruys and colleagues (2014) argue that individuals with certain face processing deficits, (i.e., individuals with Autism) tend to weight this residual error more halting perceptual inference progressing to a higher level to form more abstract representations. Thus, low-level predictions will be used at the expense of higher-level predictions. For our results, we might postulate that the poor performance in higher-level regions is a result of this over emphasis of residual error.

In conclusion, our study aimed to provide a glimpse into a lesser studied abstraction of face perception – mid-level features. We found that similar to the object and scene literature, there is strong mid-level representation of faces in retinotopic areas. Furthermore, we found this this representation was idiosyncratic – participants with more severe Autistic traits had better representation of mid-level features in retinotopy. Thus, individuals who have Autistic traits may have enhanced local processing that emerges in retinotopy. To better understand how biases in local processing propagates into abstractions of higher-level semantic representation, we related mid-level model performance in retinotopy and higher identity model performance in FFA. We found a significant negative relationship suggesting that biases toward local features in Early Visual Cortex may influence how participants extract holistic information in higher-level face processing regions. Our results give us invaluable insights into how face processing deviates in individuals known to have deficits in face perception.

Chapter 4

Cortical Representation of High-Level Semantic Face Features

4.1 Introduction

When we engage in social interactions, a person's identity or expression can indicate crucial cues about their social status, their internal emotional state, and their relationship with others. Thus, it is important that we understand the mechanisms of face identity and emotion perception in the brain as it gives us insights into the fabric of what makes us humans. Previous findings in the field of neuroscience that look at face identity and emotion perception are somewhat contradictory. Some argue that face processing is strictly hierarchical (Haxby et al, 2000) and domain specific (Gauthier et al., 2000; Hoffman & Haxby, 2000; Kanwisher et al., 1997; Kanwisher, 2000) while others argue that face processing has multiple pathways that are not defined in traditional hierarchical frameworks (Duchaine & Yovel, 2015), and face processing regions are not restricted in representing only one type of information (Baseler et al., 2014; Fox et al., 2011; Ganel et al., 2005; Said et al., 2011; Wegrzyn et al., 2015).

As discussed in the previous chapter, these inconsistencies may be a consequence of using restricted stimulus sets comprising posed faces with limited ranges of emotional expressions and demographics. In our study, we aim to use a large, varied naturalistic stimulus set of faces to investigate tuning to emotion and identity features in Occipital Temporal Cortex (OTC). Similar to the modeling framework discussed in the previous chapter, we employ multi-feature voxelwise modelling (Naselaris et al., 2009; Stansbury et al., 2013; Huth et al., 2012; Huth et al., 2016) to understand how emotion and identity are encoded in the brain.

In this chapter, we also seek to understand whether there are idiosyncrasies in tuning to emotion and identity features and whether these differences relate to Developmental Prosopagnosia and Autistic traits. Those with Developmental Prosopagnosia are known to have difficulty processing face identity (Duchaine & Nakayama, 2006) without any structural damage to the cortex. One study suggests that these deficits in face identity processing stem from worse face selectivity throughout the ventral stream (Jiahui et al., 2018). However, it

is unknown what aspects of faces lead to this suppressed face network in participants with Developmental Prosopagnosia. Autism, a developmental disorder characterized by issues with social communication and interaction, is typically seen as a disorder in which people struggle with global processing, but the literature on whether Autistic individuals actually demonstrate issues with higher-level semantic perception is quite mixed (Baron-Cohen et al., 2001a; Golan et al., 2006; Hooper et al., 2018; Harms et al., 2010; Loth et al., 2018; Philip et al., 2010; Uljarevic & Hamilton 2013; Weigelt et al., 2012; Wright et al., 2008). Furthermore, there is conflicting evidence as to whether abnormalities in the face processing stream occur functionally in cortex. There have been several studies that have theorized that issues in higher-level processing in Autism stem from less activation in fusiform face area (FFA) (Holt et al., 2014; Spencer et al., 2011) and superior temporal sulcus (STS) (Golan et al., 2006) when performing complex face tasks. However there exists literature that argue the contrary (Corbett et al., 2009; Ewbank et al., 2017; Harms et al., 2010).

In our study, we propose using our multi-feature voxelwise encoding modelling framework to both investigate tuning to semantic features on a group level and understand how this tuning differs as a function of severity of Developmental Prosopagnosia and Autism traits. The overall goals of our study are: 1) to understand how the brain organizes semantic features, 2) to investigate whether there is co-tuning in the brain to identity and emotion features, 3) to see if semantic representation differs as function of severity in Developmental Prosopagnosia traits or Autistic traits, and 4) to show at what level of representation (higher vs lower semantic features) these idiosyncrasies occur.

4.2 Methods

All methods relating to participants, stimuli, experimental procedure, fMRI data acquisition, and data pre-processing were identical to the methods described in the previous chapter of this dissertation. All sections of fMRI data modeling and statistical analysis with the exception of models constructed and variance partitioning were identical as well. To measure severity of Developmental Prosopagnosia traits, we used Cambridge Face Memory Task (CFMT) scores (lower scores equate to worse face recognition performance). To measure severity of Autistic traits, we used Autism Quotient (AQ) scores (higher scores imply more severe Autistic traits), see previous chapter methods.

fMRI Data Modeling

Models Constructed

Three main models (Table 4.1A) were created to visualize model performance in OTC on a group basis. The Semantic Binary model was created that comprised 89 binary (0/1) features which included both expression (e.g., happy, worried, annoyed) and identity-related features (e.g., Asian, female, gray hair). We further broke this model down into two other models – an Emotion only model which included expression terms and an Identity only

A

Emotion Model	Identity Model	Semantic Binary Model
Higher Emotion + Lower Emotion Models	Higher Identity + Lower Identity Models	Emotion + Identity Models

B

Higher Identity	Lower Identity	Higher Emotion	Lower Emotion
asian	5 o'clock shadow	adoring	arched eyebrows
black african	bags under eyes	afraid	eyes closed
child	bald	amazed	mouth closed
female	bangs	amused	mouth slightly open
hispanic or latino	beard	angry	mouth wide open
indian	big eyes	annoyed	squinting
middle aged	big lips	anticipating	teeth not visible
middle eastern	black hair	cautious	teeth very prominent
senior	blond hair	confident	
teenager	blue eyes	confused	
white caucasian	brown eyes	contented	
young adult	brown hair	curious	
	bushy eyebrows	dejected	
	curly hair	delighted	
	double chin	disapproving	
	facial hair	disgusted	
	goatee	distressed	
	gray hair	excited	
	green eyes	frustrated	
	mustache	grief-struck	
	oval face	happy	
	receding hairline	joyful	
	red hair	neutral	
	round face	offended	
	short hair	pained	
	small eyes	peaceful	
	square face	playful	
	straight hair	pleased	
	thin eyebrows	sad	
	very skinny	shocked	
	wavy hair	smiling	
	waxed eyebrows	stern	
	white hair	surprised	
		suspicious	
		thoughtful	
		worried	

Table 4.1: Semantic model features. (A) The three main semantic models: Emotion model which comprises the terms from Lower Emotion and Higher Emotion, the Identity model which includes terms from the Lower Identity and Higher Identity models, and the Semantic Binary model which includes all emotion and identity related terms. (B) The four sub-models used in individual differences analyses: the Higher Identity model which contains age, race/ethnicity, gender terms, the Lower Identity model which includes lower-level descriptors of identity, the Higher Emotion model which includes categorical emotion terms, and the Lower Emotion model which comprises structural semantic terms relating to emotion.

model which only comprised identity terms. When investigating idiosyncrasies in tuning to semantic features across OTC, we sought to understand at which levels of abstractions these individual differences emerge. We split the Identity only model into lower-level features that describe attributes of the face (i.e., bald, big eyes, black hair) and higher-level features that describe more abstract representations of identity (i.e., Asian, female, old). We repeated this procedure for the Emotion only model (see Table 4.1B). Three in-lab raters were used to label the face for the absence versus presence of each feature across all models. An agreement of two raters was used as the final value (0/1).

Statistical Analysis

Variance Partitioning

Variance partitioning (Nimon & Reio, 2011) was performed separately on the three main models: Semantic Binary, Identity, and Emotion. This analysis allows to understand the unique variance and shared variance across groups of features. For a given feature space A, feature space B, and the union of those feature spaces AB, we defined the following:

$$Unique\ R^2_A = R^2_{AB} - R^2_B$$
$$Unique\ R^2_B = R^2_{AB} - R^2_A$$
$$Shared\ R^2 = R^2_A + R^2_B - R^2_{AB}$$

We limited analysis on voxels that had significant model fit ($p < .05$, uncorrected) for the union model. All analyses were done on a voxel-wise basis. Unique variances and shared variances were averaged across voxels in a given ROI for each participant. We sought to find the unique variance of emotion and identity feature spaces for the Semantic Binary model. For the Identity model, higher identity terms were determined to be one feature space, and lower identity terms were determined to be another feature space. In the Emotion model, we identified higher emotion terms for a feature space and lower emotion terms for another feature space. For group level results, we averaged unique variances and shared variances for each ROI across participants.

4.3 Results

Group Significance Maps and Mean Validation Correlation

Participants' cortex maps were transformed into a fsaverage template to look at model (Semantic Binary, Emotion, and Identity) performance across OTC on a group basis. Binomial t-tests were conducted (significance threshold: $q(FDR) < 0.05$) to see which voxels had significant model fit for each model (Figure 4.1, top panel). Mean group validation correlation was also obtained by averaging individual participants' Fisher Z values (obtained from transforming validation correlation into Fisher Z) and then re-transforming the average

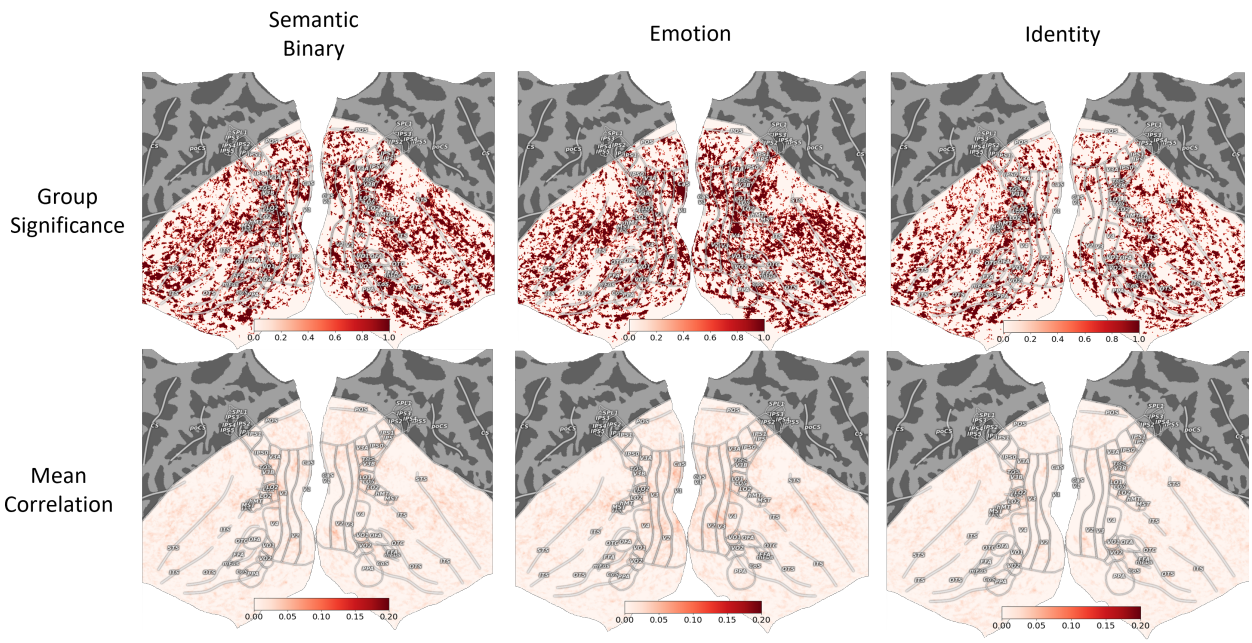


Figure 4.1: Group mid-level model performance in fsaverage. From left to right, group-level model performance for the Semantic Binary, Emotion only, and Identity only model plotted on fsaverage maps. Grey regions represent areas that were not used in analyses (regions outside OTC). In the top panel, we have corrected ($q[FDR] < .05$) group significance maps for model validation performance that were calculated using binomial t-tests. Uncorrected binary individual significance maps (as calculated by permutation tests, see previous chapter methods) were transformed into the common fsaverage space. A binomial t-test with a success threshold of .05 was then conducted across transformed individual maps to ascertain whether a voxel was significant on a group basis. FDR correction was then performed across OTC for the group map. Red indicates that the voxel was significant after FDR correction (across OTC) for the group. The bottom panel shows the group validation correlation across participants calculated by converting individual correlations for each vertex into Fisher Z values. Those Fisher Z values were averaged and then re-transformed into a correlation value. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction.

Fisher Z value into an average correlation (Figure 4.1, bottom panel). At a group level, there were no clear organization of model performance in OTC. This could, however, be due to high individual differences in model fit in which effects are wiped out when averaging data across participants. In the next section of our results, we seek to understand whether there exist individual differences and whether this relates to measures of Autistic traits or face recognition ability.

Individual Differences in Model Fit

As in previous chapter, we sought to examine individual differences by correlating model performance and AQ/CFMT scores. As a first step, results were visualized on a group basis for

our three main models (Semantic Binary, Emotion, Identity). Participants' cortex flatmaps were transformed into the fsaverage template space. For each voxel, participants' model performance metrics (Fisher Z transforms of validation correlation) were correlated with scores obtained from our individual differences measures: AQ and CFMT. For AQ, we did not qualitatively see a relationship between the semantic models within OTC (Figure 4.2, top panel). To quantify this, we took the number of significant voxels ($p < .05$, uncorrected) in each participant's functionally defined ROI and partially correlated this with AQ while covarying out ROI size. There was a significant partial correlation between the number of significant voxels in STS and AQ ($R_{partial} = .35$, $p = .037$), see Figure C.4A. However, when examining what abstraction of identity (higher/lower identity) or emotion (higher/lower emotion) contributed to this significant relationship, we did not find any significant correlations (Figure C.5, C.6). This may indicate that those with more severe Autistic traits (higher AQ scores) do not a preference for specific semantic features, but rather represent all semantic features well in STS.

For individuals differences in Developmental Prosopagnosia traits, we observed in Figure 4.2 that participants with higher CFMT scores (better at face identity recognition) had a pattern of higher model performance in OFA and FFA in the Semantic Binary and Identity Model. To statistically analyze these differences, the number of significant voxels in each participant's functional ROI were correlated with CFMT while controlling for ROI size. We found a significant relationship between the number of significant voxels in OFA and CFMT scores ($R_{partial} = .45$, $p = .010$) for the Semantic Binary model as well as a significant correlation between the number of significant voxels in OFA and CFMT ($R_{partial} = .44$, $p = .007$) for the Identity model. Although there was a significant correlation between FFA and CFMT for both the Semantic Binary ($R = .33$, $p = .04$) and Identity model ($R = .34$, $p = .038$) before controlling out for ROI size, this did not hold once we covaried out size (Semantic Binary: $R_{partial} = .03$, $p = .225$; Identity: $R_{partial} = .22$, $p = .192$). We did not observe any significant results when relating the Emotion model and CFMT (Figure C.4).

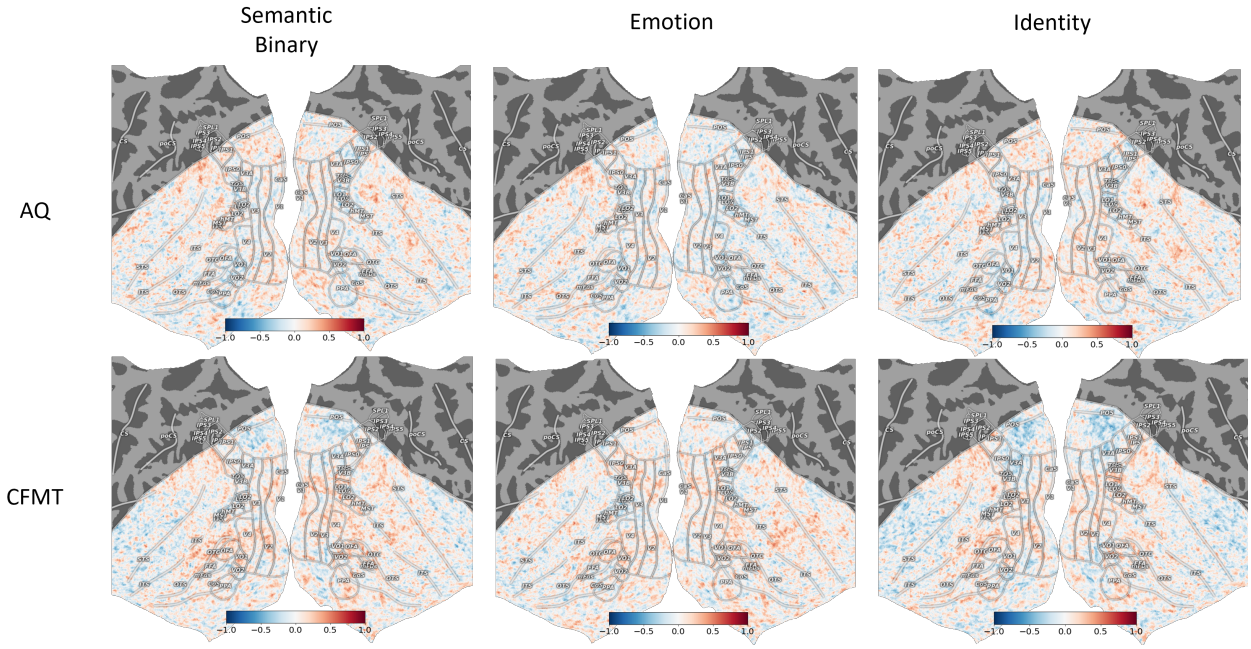


Figure 4.2: Correlation plots of model performance and AQ scores (top panel) and CFMT scores (bottom panel) in fsaverage. For each voxel, we correlated the model performance (Fisher Z transform of validation correlation) with an individual differences measure. On the top panel, positive red correlations mean that participants with higher AQ scores (more severe Autistic traits) had better model performance. Negative blue correlations indicate that participants with lower AQ scores (less Autistic traits) had better model performance. On the bottom panel, red indicates that participants with higher CFMT scores (better at face identity) had better performance whereas blue indicates that participants with lower CFMT scores (worse at face identity) had better model performance. We observed in the Semantic Binary model and Identity model that individuals who were better at faces (higher CFMT scores) had better performance in OFA.

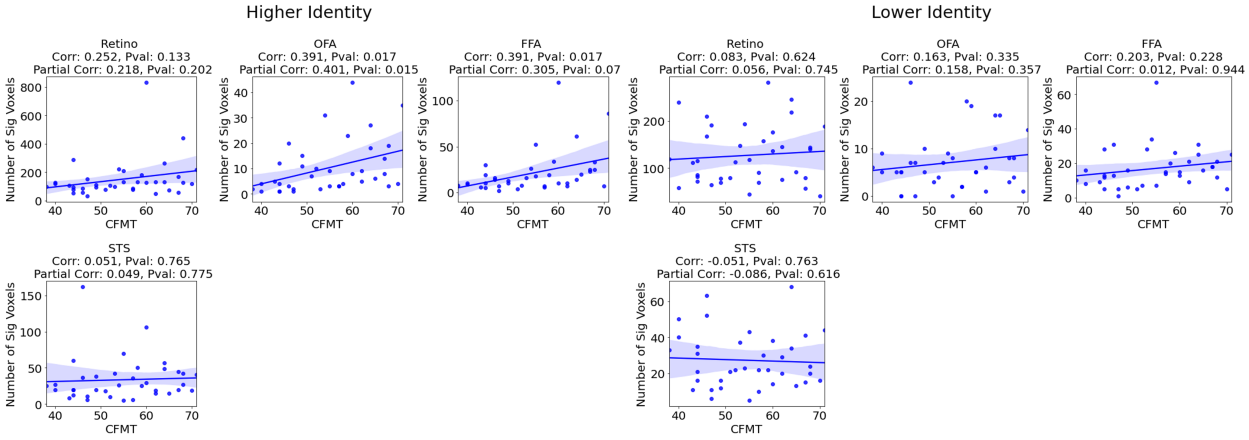


Figure 4.3: Model performance correlated with CFMT for the Identity sub-models. For each participant, the number of significant voxels in each ROI were calculated from permutation tests ($p < .05$, uncorrected) for each model (left: Higher Identity, right: Lower Identity) and correlated with CFMT scores. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation as well as the partial correlation and partial p value when controlling for ROI size. We show a significant relationship between Higher Identity model performance in OFA and CFMT.

To further determine what aspects of identity individuals with Developmental Prosopagnosia traits have deficits in, we correlated the model performance of two sub-models (Lower Identity and Higher Identity) with CFMT scores. We found that the number of significant voxels in the Higher Identity model, specifically, correlated significantly with CFMT in OFA ($R_{partial} = .40$, $p = .015$), see Figure 4.3. Similar to what we observed in the Semantic Binary and Identity models, there was a significant relationship in model performance in FFA for the Higher Identity model without controlling for size ($R = .39$, $p = .017$), but once we covaried out ROI size, the relationship was again not significant ($R_{partial} = .31$, $p = .07$). We did not observe any significant correlations with the Lower Identity model and CFMT. These results indicate that participants who have poorer face recognition abilities (lower CFMT scores) have worse model performance in OFA specifically for Higher Identity.

Variance Partitioning

In the previous section, we showed a significant relationship between model performance and CFMT scores for the Semantic Binary, Identity, and Higher Identity models in OFA and no significant relationship for the Emotion or the Lower Identity models. This might suggest that participants who have Developmental Prosopagnosia traits have worse representation of higher identity in OFA while their lower identity and emotion representation remain intact throughout OTC. However, this analysis does not inform us of two things: 1) whether our results are due to the unique contribution of higher identity or perhaps shared variance common across the Higher Identity model and other semantic models, and 2) whether this relationship is due to an issue with higher identity alone or rather competing representations

of higher identity vs lower identity. To disentangle both these points, we conducted variance partitioning.

Variance partitioning was first performed on a group basis to understand how different information was represented throughout OTC. For the Semantic Binary model, we sought to find the unique contribution of emotion and identity terms. A higher proportion of unique variance for identity in retinotopy, OFA, and FFA and a higher proportion of unique variance for emotion in STS (Figure 4.4A) was observed. While this might support the two-pathway model of emotion and identity (Haxby et al., 2000), we also see high amounts of shared variance across the ventral stream. We further broke down the emotion and identity terms into higher and lower-level features. For identity, a higher proportion of unique variance of lower identity features (Figure 4.4B) was observed throughout OTC suggesting that face processing in the ventral stream mainly extracts features that describe face attributes. For the Emotion model, we see a higher amount of unique variance of higher emotion features (Figure 4.4C) across all ROIs. This leads us to suspect that there is a preference for higher levels of emotion representation throughout cortex.

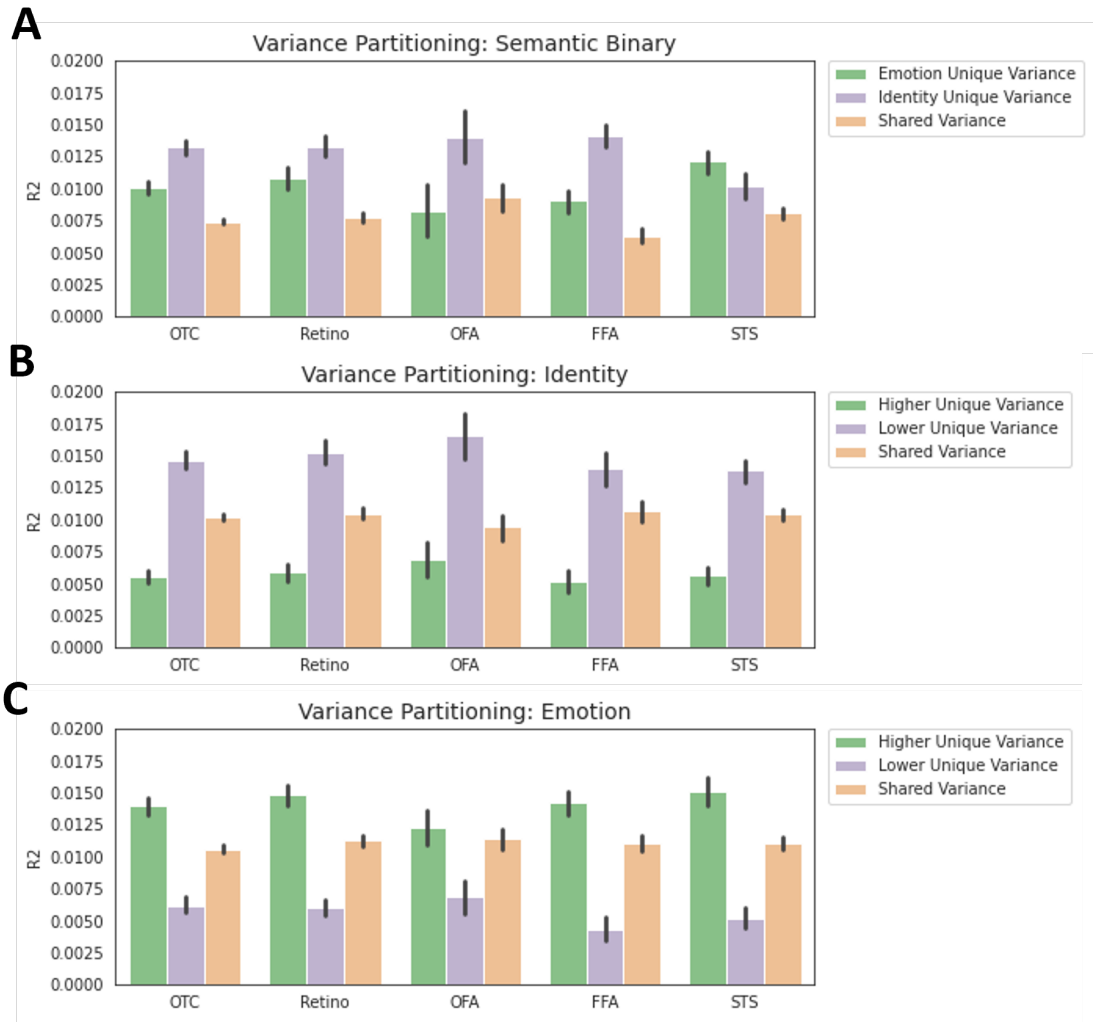


Figure 4.4: Variance partitioning for (A) Semantic Binary (emotion vs identity features) (B) Identity (higher vs lower identity features), and (C) Emotion models (higher vs lower emotion features). Analyses were limited in significant voxels ($p < .05$, uncorrected) for the union models (see Methods). For each model, we calculated the unique contribution of two features spaces as well as the shared variance between those features spaces for each voxel. We then averaged these values across voxels in a given ROI for each participant and then again across participants. Bars represent the mean variance explained across participants, and error bars indicate standard errors across participants. (A) For the Semantic Binary model, we see that identity features contribute higher unique variances in retinotopy, OFA, and FFA while emotion features have a higher unique variance in STS. (B) In the Identity model, lower identity features have higher unique variance across all ROIs. (C) In the Emotion model, higher emotion features have higher unique variance across all ROIs.

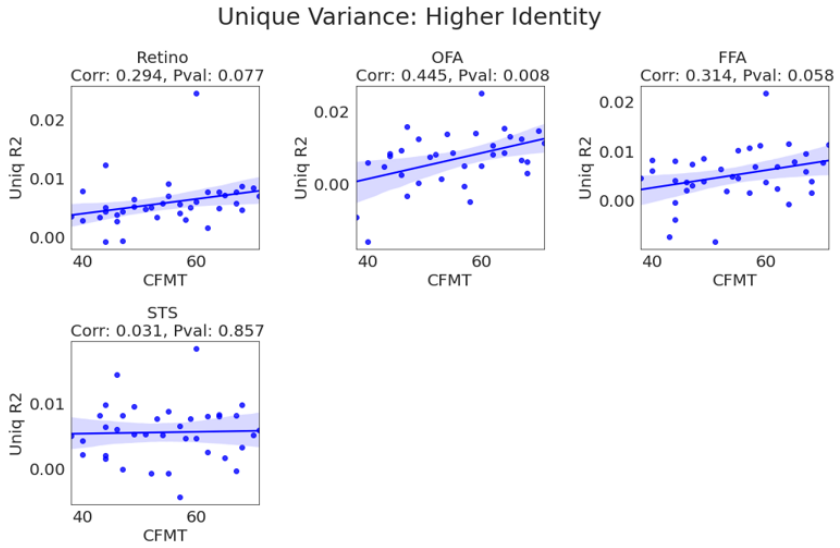


Figure 4.5: Correlation between the unique contribution of higher identity features and CFMT. We obtained the unique variance of higher identity features by subtracting the variance explained by the lower identity features from the variance explained by the Identity model (see Methods) in a given ROI for each participant. We then correlated this metric against CFMT. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation. We observe a significant relationship between the unique contribution of higher identity features and CFMT in OFA.

To test whether the relationship between Higher Identity and CFMT still holds when taking out shared variance common across other features, we correlated the unique variance of Higher Identity with CFMT in each ROI. We again observed a significant relationship ($R = .45$, $p = .008$) with perceptual face recognition performance and unique contribution of Higher Identity in OFA (Figure 4.5). This result coupled with the absence of significant correlations between CFMT and shared variance between higher and lower identity features (Figure C.7) further support the notion that individuals who have trouble recognizing face identity have worse encoding of higher identity in OFA.

To understand whether the relative unique contribution of features varies as a function of face identity recognition, we correlated the difference between the unique variances of higher and lower identity with scores on CFMT for each ROI. We observed a significant correlation between CFMT and the relative contributions of higher identity over lower identity in retinotopy ($R = .33$, $p = .048$), OFA ($R = .37$, $p = .032$), and FFA ($R = .34$, $p = .042$), see Figure 4.6. Although we didn't see any previous significant relationship between the unique contribution of Higher Identity features in Retinotopy and FFA and CFMT, this analysis suggests these regions still contribute to competing representations of higher vs lower identity. Thus, the encoding of higher identity representations for participants with better perceptual identity performance is not isolated exclusively in OFA. Rather, there are mechanisms in place throughout the ventral stream favoring a dominant representation of higher over lower identity.

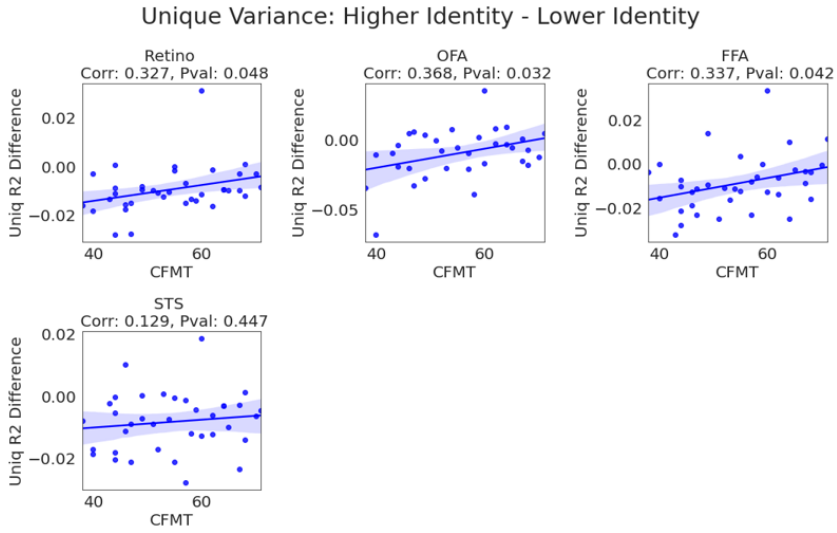


Figure 4.6: The relative contribution of higher identity features vs lower identity features and CFMT. Lower identity unique variance was subtracted from higher identity unique variance in a given ROI and then correlated with CFMT. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation. We observe a significant relationship in Retinotopic areas, OFA, and FFA suggesting that there is competing representations of higher identity over lower identity throughout the ventral stream, and this is related to overall face identity performance.

4.4 Discussion

Using a multi-feature voxel-wise modelling framework, we created three main semantic models (Semantic Binary: Emotion + Identity, Identity Only, Emotion Only) to look at group level effects. While we didn't see any clear organization in OTC when looking at group significance maps for the three main semantic models, we did observe a distinct organization of emotion and identity across the ventral stream when conducting variance partitioning across participants in significant voxels. Our study shows a higher amount of unique variance attributed to identity in OFA and FFA and a higher amount of unique variance attributed to emotion in STS. At first glance, these results support the two-pathway model (Bruce & Young, 1986; Haxby et el., 2000) and domain specific representation of face processing regions (Gauthier et al., 2000; Haxby et al., 1999; Hoffman & Haxby, 2000; Kanwisher et al., 1997). However, when looking at our variance partitioning results more closely, we still see a high amount of shared variance and unique variance attributed to both identity and emotion across all face ROIs. This thus supports literature in the field suggesting that there is co-tuning to both expression and identity in face processing regions (Baseler et al., 2014; Fox et al., 2011; Ganel et al., 2005; Said et al., 2011; Wegrzyn et al., 2015) and that these regions represent more diverse information than originally thought.

We further broke down our semantic models into higher and lower-level features (Higher Emotion, Lower Emotion, Higher Identity, Lower Identity) to investigate how different ab-

stractions of semantic features are tuned in OTC and whether this was idiosyncratic. When examining differences on an individual basis, we only observed a significant relationship between model performance for the Semantic Binary model (emotion + identity terms) in STS and measures that estimate Autistic trait severity (AQ). There were no significant relationships between AQ and model performance when breaking these models down further into emotion and identity terms or higher and lower-level features. This may suggest that those with severe Autistic traits recruit STS more when processing semantic information regardless of feature type. This falls in line with previous literature showing Autistic individuals have higher recruitment of STS during complex emotion and identity face tasks (Corbett et al., 2009; Harms et al., 2010)

In contrast to previous literature (Harms et al., 2010; Holt et al., 2014; Spencer et al., 2011), there does not seem to be worse representation of global, semantic features in face processing regions for participants with Autistic traits. In the previous chapter, we observed that participants with higher AQ had better representations of mid-level face information in retinotopy. Thus, in our studies, we show that local biases as indicated by superior mid-level model performance does not come at the expense of semantic model performance. Studies supporting the Enhanced Perceptual Functioning theory (Mottron, 2006) argue that Autistic individuals both have enhanced local processing and intact global processing (Mottron et al., 2003; Tanaka & Sung, 2016). Our results further support this framework over opposing theories such as the Weak Central Coherence theory (Happe & Frith, 2006) which claim that enhanced local processing in Autistic individuals is a result of poor integration of global information.

When investigating idiosyncrasies in model fit for participants with face identity processing deficits (Developmental Prosopagnosia traits), we demonstrate that model fit for the Semantic Binary, Identity, and Higher Identity models in OFA correlated with perceptual performance on face recognition tasks (CFMT). Thus, those who had worse face identity recognition (low CFMT score) had significantly worse model fit in OFA. Although these results lead us to believe that encoding of identity, specifically higher identity, features in OFA vary with perceptual performance on identity recognition tasks, this analysis alone does not disentangle whether our results are due to shared variance common across other features.

To understand the unique contributions of features, we conducted variance partitioning and correlated the unique variance attributed to Higher Identity with CFMT. We again saw a significant relationship between the unique contribution of higher identity features leading us to support our previous findings that those with worse face identity recognition have poorer encoding of higher identity features in OFA. We additionally examined the relative unique contribution of higher identity features over lower identity terms and observed a significant relationship between the unique variance of higher identity compared to lower identity in retinotopy, OFA, and FFA with perceptual performance on face identity recognition.

These results are in line with literature showing that there is selectivity reduction to faces in higher-level face processing regions (Jiahui et al., 2018) for Developmental Prosopagnosics, but in our study, we show that these differences stem specifically from the inability to encode higher identity features such as race, ethnicity, gender, and age in OFA. Additionally, our

study suggests that throughout the ventral processing stream, the relative contribution of higher identity vs lower identity features scales with perceptual performance in face recognition. This might indicate that face recognition abilities are related to consistent, competing perceptual representations of higher identity over lower identity features from Early Visual Cortex to FFA.

Chapter 5

General Discussion

Traditionally, studies have used controlled, restricted stimulus sets to research face perception with findings being potentially specific to the stimuli used in those studies. In this dissertation, I utilized diverse, varied naturalistic face stimuli to circumvent these biases in an attempt to understand how individuals perceive emotional faces. In Chapter 2, I detailed two behavioral experiments that investigated which features of naturalistic scenes of crowds contribute to participants' overall percept of crowd valence, arousal, and dominance. In the first of these experiments, I found that participants used different aspects of the scene depending on the emotional dimension being judged. The mean valence across faces contributed most to the overall perception of crowd valence whereas for arousal and dominance, the most arousing or most dominant face (max emotion) explained the most unique variance in crowd emotion. In the second experiment, background context was found to mitigate the relative reliance on mean emotion versus max emotion in judgements of arousal and dominance. That is, when we remove the background context of a scene, individuals tended to weigh mean emotion more heavily compared to when context was intact. However, in the case of arousal, the most arousing face continued to exert influence even when context was removed which is in line with theoretical accounts arguing for a unique bottom-up effect of emotional arousal on attentional competition and post-attentive perceptual processing. Together, the findings for Chapter 2 suggest that the reliance of mean emotion vs max emotion when judging crowd emotion differs depending on the dimension of affect. Additionally, background context modulates the relative influence of max vs mean emotion potentially directing attention to more salient parts of the scene.

In Chapters 3 and 4, I leveraged a multi-feature encoding modeling framework to understand how features of naturalistic emotional faces are encoded throughout Occipital Temporal Cortex (OTC) and whether this representation differs in individuals with face processing deficits. In Chapter 3, I found that mid-level representation of features was best characterized in Early Visual Cortex (EVC) even when controlling for low-level statistical dependencies. The strength of this representation in EVC scaled with Autistic trait severity leading us to conclude that participants with more severe Autistic traits have enhanced local representation. No relationship between model performance and face recognition ability, as indexed

by scores on the Cambridge Face Memory Test (CFMT) was observed suggesting that traits relating to Developmental Prosopagnosia are not defined by idiosyncrasies in tuning to mid-level features. We additionally observed a negative relationship between mid-level feature encoding in EVC and the encoding of higher identity features in higher-level face processing regions. This thus led me to ask in Chapter 4 how higher-level semantic features were represented in OTC and whether individuals with more severe Autistic traits had worse global, semantic representations.

In Chapter 4, I investigated how semantic features were represented in OTC and how this differed as a function of Autistic traits and face recognition ability. I found co-tuning of emotion and identity features throughout face processing regions at the level of the group. Contrary to our hypothesis in Chapter 3, individuals with more severe Autistic traits did not have worse semantic representation. The presence of enhanced local processing and intact global processing in those with more severe Autistic traits is in line with the Enhanced Perceptual Functioning model (Mottron, 2006) over other accounts of Autism such as the Weak Central Coherence model (Happe & Frith, 2006) that argue for enhanced local processing at the expense of global processing.

While the results in Chapter 4 suggest that those with Autistic traits do not have issues with higher-level semantic representation, there was an observed significant relationship demonstrated between encoding of higher identity features (race, age, gender, ethnicity) in occipital face area (OFA) and face identity recognition performance, as indexed by scores on the CFMT. Competing representations of higher vs lower identity in EVC, OFA, and fusiform face area (FFA) were also found to scale with face recognition ability. These results suggest that those who are better at face processing can abstract higher levels of identity representation in face processing regions.

Together, the findings in this dissertation give insights into how face processing manifests perceptually and cortically when using naturalistic stimuli. I believe this dissertation can provide new perspectives to existing research, and my hope is that throughout psychology, neuroscience, and other domains of research, there is an emphasis on using more naturalistic stimuli in conjunction with interpretable analyses to mitigate biases in the field. The brain is extremely intricate and complex, and as researchers, we can only estimate to the best of our ability what we think it's doing. While we still have a long way to go in understanding the mechanisms underlying visual perception, I'm excited to see how the field advances and what new insights it will bring.

Chapter 6

References

Alwis, Y., & Haberman, J. M. (2020). Emotional judgments of scenes are influenced by unintentional averaging. *Cognitive Research: Principles and Implications*, *5*(1), 28. https://doi.org/10.1186/s41235-020-00228-3

Arivazhagan, S., Ganesan, L., & Priyal, S. P. (2006). Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters*, *27*(16), 1976–1982. https://doi.org/10.1016/j.patrec.2006.05.008

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion* , *11*(6), 1406–1414. https://doi.org/10.1037/a0023578

Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, *17*, 47–54. https://doi.org/10.1016/j.copsyc.2017.06.006

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001a). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 241–251.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001b). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://www.ncbi.nlm.nih.gov/pubmed/11439754

Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science*, *21*(4), 595–599. https://doi.org/10.1177/0956797610363547

Baseler, H. A., Harris, R. J., Young, A. W., & Andrews, T. J. (2014). Neural Responses to

Expression and Gaze in the Posterior Superior Temporal Sulcus Interact with Facial Identity. In *Cerebral Cortex* (Vol. 24, Issue 3, pp. 737–744). https://doi.org/10.1093/cercor/bhs360

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, *23*(2), 137–152. https://doi.org/10.1109/TMI.2003.822821

Bradley, M. M., Keil, A., & Lang, P. J. (2012). Orienting and emotional perception: facilitation, attenuation, and interference. *Frontiers in Psychology*, *3*, 493. https://doi.org/10.3389/fpsyg.2012.00493

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. https://doi.org/10.1080/13506280500165188

Brooks, J. C. W., Beckmann, C. F., Miller, K. L., Wise, R. G., Porro, C. A., Tracey, I., & Jenkinson, M. (2008). Physiological noise modelling for spinal functional magnetic resonance imaging studies. *NeuroImage*, *39*(2), 680–692. https://doi.org/10.1016/j.neuroimage.2007.09.018

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology , 77 (Pt 3)*, 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Calbi, M., Heimann, K., Barratt, D., Siri, F., Umiltà, M. A., & Gallese, V. (2017). How Context Influences Our Perception of Emotional Faces: A Behavioral Study on the Kuleshov Effect. *Frontiers in Psychology*, *8*, 1684. https://doi.org/10.3389/fpsyg.2017.01684

Chang, L., & Tsao, D. Y. (2017). The Code for Facial Identity in the Primate Brain. *Cell*, *169*(6), 1013–1028.e14. https://doi.org/10.1016/j.cell.2017.05.011

Chen, Y., Norton, D. J., McBain, R., Gold, J., Frazier, J. A., & Coyle, J. T. (2012). Enhanced local processing of dynamic visual information in autism: evidence from speed discrimination. *Neuropsychologia*, *50*(5), 733–739. https://doi.org/10.1016/j.neuropsychologia.2012.01.007

Chen, Z., & Whitney, D. (2019). Tracking the affective state of unseen persons. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7559–7564. https://doi.org/10.1073/pnas.1812250116

Chen, Z., & Whitney, D. (2020). Inferential emotion tracking (IET) reveals the critical role of context in emotion recognition. *Emotion* . https://doi.org/10.1037/emo0000934

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*,

4(5), 170–178. https://doi.org/10.1016/s1364-6613(00)01476-5

Clausi, D. A., & Ed Jernigan, M. (2000). Designing Gabor filters for optimal texture separability. *Pattern Recognition*, *33*(11), 1835–1849. https://doi.org/10.1016/S0031-3203(99)00181-8

Corbett, B. A., Carmean, V., Ravizza, S., Wendelken, C., Henry, M. L., Carter, C., & Rivera, S. M. (2009). A functional and structural study of emotion and face processing in children with autism. *Psychiatry Research*, *173*(3), 196–205. https://doi.org/10.1016/j.pscychresns.2008.08.005

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, *48*(3), 497–507. https://doi.org/10.1016/j.neuron.2005.10.018

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

DeBruine, L. (2018). debruine/webmorph: Beta release 2 (v0.0.0.9001). *Zenodo*. https://doi.org/10.5281/zenodo.1162670

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., & Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(6), 2382–2386. https://doi.org/10.1073/pnas.93.6.2382

Dhall, A., Goecke, R., Ghosh, S., Joshi, J., Hoey, J., & Gedeon, T. (2017). From individual to group-level emotion recognition: EmotiW 5.0. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 524–528. https://doi.org/10.1145/3136755.3143004

Dhall, A., Goecke, R., Joshi, J., Hoey, J., & Gedeon, T. (2016). EmotiW 2016: video and group-level emotion recognition challenges. *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 427–432. https://doi.org/10.1145/2993148.2997638

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

Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science*, *1*, 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex* , *7*(2), 181–192. https://doi.org/10.1093/cercor/7.2.181

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, *369*(6481), 525. https://doi.org/10.1038/369525a0

Ewbank, M. P., Pell, P. J., Powell, T. E., von dem Hagen, E. A. H., Baron-Cohen, S., & Calder, A. J. (2017). Repetition Suppression and Memory for Faces is Reduced in Adults with Autism Spectrum Conditions. *Cerebral Cortex* , *27*(1), 92–103. https://doi.org/10.1093/cercor/bhw373

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284. https://doi.org/10.1002/(sici)1097-0193(1999)8:4¡272::aid-hbm10¿3.0.co;2-4

Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. In *Biological Cybernetics* (Vol. 61, Issue 2, pp. 103–113). https://doi.org/10.1007/bf00204594

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: effects of social status and speaking. *Cognition*, *117*(3), 319–331. https://doi.org/10.1016/j.cognition.2010.09.003

Fox, C. J., Hanif, H. M., Iaria, G., Duchaine, B. C., & Barton, J. J. S. (2011). Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia*, *49*(12), 3188–3200. https://doi.org/10.1016/j.neuropsychologia.2011.07.018

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–981. https://doi.org/10.1038/nn.3402

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the "fusiform face area" in processing facial expression. In *Neuropsychologia* (Vol. 43, Issue 11, pp. 1645–1654). https://doi.org/10.1016/j.neuropsychologia.2005.01.012

Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: an interactive surface visualizer for fMRI. *Frontiers in Neuroinformatics*, *9*, 23.

https://doi.org/10.3389/fninf.2015.00023

Garrido, L., Furl, N., Draganski, B., Weiskopf, N., Stevens, J., Tan, G. C.-Y., Driver, J., Dolan, R. J., & Duchaine, B. (2009). Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. *Brain: A Journal of Neurology*, *132*(Pt 12), 3443–3455. https://doi.org/10.1093/brain/awp271

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. https://doi.org/10.1162/089892900562165

Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The "reading the mind in films" task: complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience*, *1*(2), 111–123. https://doi.org/10.1080/17470910600980986

Golarai, G., Grill-Spector, K., & Reiss, A. L. (2006). Autism and the development of face processing. *Clinical Neuroscience Research*, *6*(3), 145–160. https://doi.org/10.1016/j.cnr.2006.08.001

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11), 1.1–13. https://doi.org/10.1167/9.11.1

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology: CB*, *17*(17), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, *18*(5), 855–859. https://doi.org/10.3758/s13423-011-0125-6

Han, L., Yamanashi Leib, A., Chen, Z., & Whitney, D. (2021). Holistic ensemble perception. *Attention, Perception & Psychophysics*, *83*(3), 998–1013. https://doi.org/10.3758/s13414-020-02173-1

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

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. https://www.ncbi.nlm.nih.gov/pubmed/10827445

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, *22*(1), 189–199. https://doi.org/10.1016/s0896-6273(00)80690-x

Hernandez, R. N., Feinberg, R. L., Vaurio, R., Passanante, N. M., Thompson, R. E., & Kaufmann, W. E. (2009). Autism spectrum disorder in fragile X syndrome: a longitudinal evaluation. *American Journal of Medical Genetics. Part A*, *149A*(6), 1125–1137. https://doi.org/10.1002/ajmg.a.32848

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84. https://doi.org/10.1038/71152

Holt, R. J., Chura, L. R., Lai, M.-C., Suckling, J., von dem Hagen, E., Calder, A. J., Bullmore, E. T., Baron-Cohen, S., & Spencer, M. D. (2014). "Reading the Mind in the Eyes": an fMRI study of adolescents with autism and their siblings. *Psychological Medicine*, *44*(15), 3215–3227. https://doi.org/10.1017/S0033291714000233

Hooper, J. J., Sutherland, C. A. M., Ewing, L., Langdon, R., Caruana, N., Connaughton, E., Williams, N., Greenwell-Barnden, J., & Rhodes, G. (2018). Should I trust you? Autistic traits predict reduced appearance-based trust decisions. *British Journal of Psychology* . https://doi.org/10.1111/bjop.12357

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the

human brain. *Neuron*, *76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Im, H. Y., Chong, S. C., Sun, J., Steiner, T. G., Albohn, D. N., Adams, R. B., Jr, & Kveraga, K. (2017). Cross-cultural and hemispheric laterality effects on the ensemble coding of emotion in facial crowds. *Culture and Brain*, *5*(2), 125–152. https://doi.org/10.1007/s40167-017-0054-y

Jemel, B., Mottron, L., & Dawson, M. (2006). Impaired face processing in autism: fact or artifact? *Journal of Autism and Developmental Disorders*, *36*(1), 91–106. https://doi.org/10.1007/s10803-005-0050-5

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1016/s1053-8119(02)91132-8

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), E6418–E6427. https://doi.org/10.1073/pnas.1802246115

Jolliffe, T., & Baron-Cohen, S. (1997). Are people with autism and Asperger syndrome faster than normal on the Embedded Figures Test? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(5), 527–534. https://doi.org/10.1111/j.1469-7610.1997.tb01539.x

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. https://doi.org/10.1038/77664

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kim, T., Bair, W., & Pasupathy, A. (2022). Perceptual Texture Dimensions Modulate Neuronal Response Dynamics in Visual Cortical Area V4. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(4), 631–642.

https://doi.org/10.1523/JNEUROSCI.0971-21.2021

Koldewyn, K., Jiang, Y. V., Weigelt, S., & Kanwisher, N. (2013). Global/local processing in autism: not a disability, but a disinclination. *Journal of Autism and Developmental Disorders*, *43*(10), 2329–2340. https://doi.org/10.1007/s10803-013-1777-z

Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, A. (2020). Context Based Emotion Recognition Using EMOTIC Dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(11), 2755–2766. https://doi.org/10.1109/TPAMI.2019.2916866

Kret, M. E., Roelofs, K., Stekelenburg, J. J., & de Gelder, B. (2013). Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in Human Neuroscience*, *7*, 810. https://doi.org/10.3389/fnhum.2013.00810

Lahaie, A., Mottron, L., Arguin, M., Berthiaume, C., Jemel, B., & Saumier, D. (2006). Face perception in high-functioning autistic adults: evidence for superior processing of face parts, not for a configural face-processing deficit. *Neuropsychology*, *20*(1), 30–41. https://doi.org/10.1037/0894-4105.20.1.30

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186. https://doi.org/10.1038/ncomms13186

Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of Individual Items during Ensemble Coding of Facial Expressions. *Frontiers in Psychology*, *7*, 1332. https://doi.org/10.3389/fpsyg.2016.01332

Loth, E., Garrido, L., Ahmad, J., Watson, E., Duff, A., & Duchaine, B. (2018). Facial expression recognition as a candidate marker for autism spectrum disorder: how frequent and severe are deficits? *Molecular Autism*, *9*, 7. https://doi.org/10.1186/s13229-018-0187-7

Lundqvist, D., Bruce, N., & Öhman, A. (2015). Finding an emotional face in a crowd: emotional and perceptual stimulus factors influence visual search efficiency. *Cognition & Emotion*, *29*(4), 621–633. https://doi.org/10.1080/02699931.2014.927352

Mather, M., & Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. In *Magnetic Res-*

onance in Medicine (Vol. 63, Issue 5, pp. 1144–1153). https://doi.org/10.1002/mrm.22361

Mottron, L., Belleville, S., & Ménard, E. (1999). Local bias in autistic subjects as evidenced by graphic tasks: perceptual hierarchization or working memory deficit? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *40*(5), 743–755. https://www.ncbi.nlm.nih.gov/pubmed/10433408

Mottron, L., Burack, J. A., Iarocci, G., Belleville, S., & Enns, J. T. (2003). Locally oriented perception with intact global processing among adolescents with high-functioning autism: evidence from multiple paradigms. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *44*(6), 904–913. https://doi.org/10.1111/1469-7610.00174

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Mou, W., Celiktutan, O., & Gunes, H. (2015). Group-level arousal and valence recognition in static images: Face, body and context. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, *05*, 1–6. https://doi.org/10.1109/FG.2015.7284862

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902–915. https://doi.org/10.1016/j.neuron.2009.09.006

Nimon, K., & Reio, T. G. (2011). Regression Commonality Analysis: A Technique for Quantitative Theory Building. *Human Resource Development Review*, *10*(3), 329–340. https://doi.org/10.1177/1534484311411077

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology: CB*, *21*(19), 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031

Ohlsen, G., van Zoest, W., & van Vugt, M. (2013). Gender and facial dominance in gaze cuing: emotional context matters in the eyes that we follow. *PloS One*, *8*(4), e59471. https://doi.org/10.1371/journal.pone.0059471

Ossenfort, K. L., & Isaacowitz, D. M. (2021). Spatial attention to arousing emotional stimuli in younger and older adults. *Motivation and Emotion*, *45*(6), 790–797. https://doi.org/10.1007/s11031-021-09899-x

Philip, R. C. M., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., Atkinson, A. P., Calder, A. J., Johnstone, E. C., Lawrie, S. M., & Hall, J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine*, *40*(11), 1919–1929. https://doi.org/10.1017/S0033291709992364

Pierce, K., Haist, F., Sedaghat, F., & Courchesne, E. (2004). The brain response to personally familiar faces in autism: findings of fusiform activity and beyond. Brain: A Journal of Neurology, 127(Pt 12), 2703–2716. https://doi.org/10.1093/brain/awh289

Pinkham, A. E., Griffin, M., Baron, R., Sasson, N. J., & Gur, R. C. (2010). The face in the crowd effect: anger superiority when using real faces and multiple identities. *Emotion* , *10*(1), 141–146. https://doi.org/10.1037/a0017387

Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and FMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology: CB*, *24*(17), 2066–2070. https://doi.org/10.1016/j.cub.2014.07.060

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rassadin, A., Gruzdev, A., & Savchenko, A. (2017). Group-level emotion recognition using transfer learning from face identification. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 544–548. https://doi.org/10.1145/3136755.3143007

Ratcliff, N. J., Hugenberg, K., Shriver, E. R., & Bernstein, M. J. (2011). The allure of status: high-status targets are privileged in face processing and memory. *Personality & Social Psychology Bulletin*, *37*(8), 1003–1015. https://doi.org/10.1177/0146167211407210

Righart, R., & de Gelder, B. (2008a). Rapid influence of emotional scenes on encoding of facial expressions: an ERP study. *Social Cognitive and Affective Neuroscience*, *3*(3), 270–278. https://doi.org/10.1093/scan/nsn021

Righart, R., & de Gelder, B. (2008b). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective & Behavioral Neuroscience*, *8*(3), 264–272. https://doi.org/10.3758/cabn.8.3.264

Robertson, C. E., Kravitz, D. J., Freyberg, J., Baron-Cohen, S., & Baker, C. I. (2013). Tunnel vision: sharper gradient of spatial attention in autism. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(16), 6776–6781. https://doi.org/10.1523/JNEUROSCI.5120-12.2013

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron*, *74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Rossion, B. (2008). Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. *NeuroImage*, *40*(2), 423–426.
https://doi.org/10.1016/j.neuroimage.2007.10.047

Said, C. P., Haxby, J. V., & Todorov, A. (2011). Brain systems for assessing the affective value of faces. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 366, Issue 1571, pp. 1660–1670). https://doi.org/10.1098/rstb.2010.0351

Sasson, N. J., Pinkham, A. E., Weittenhiller, L. P., Faso, D. J., & Simpson, C. (2016). Context Effects on Facial Affect Recognition in Schizophrenia and Autism: Behavioral and Eye-Tracking Evidence. *Schizophrenia Bulletin*, *42*(3), 675–683.
https://doi.org/10.1093/schbul/sbv176

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893.
https://doi.org/10.1126/science.7754376

Sheth, B. R., & Pham, T. (2008). How emotional arousal and valence influence access to awareness. *Vision Research*, *48*(23-24), 2415–2424.
https://doi.org/10.1016/j.visres.2008.07.013

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Spencer, M. D., Holt, R. J., Chura, L. R., Suckling, J., Calder, A. J., Bullmore, E. T., & Baron-Cohen, S. (2011). A novel functional brain imaging endophenotype of autism: the neural response to facial expression of emotion. *Translational Psychiatry*, *1*, e19.
https://doi.org/10.1038/tp.2011.18

Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron*, *79*(5), 1025–1034.
https://doi.org/10.1016/j.neuron.2013.06.034

Sun, J., & Chong, S. C. (2020). Power of averaging: Noise reduction by ensemble coding of multiple faces. *Journal of Experimental Psychology. General*, *149*(3), 550–563.
https://doi.org/10.1037/xge0000667

Tanaka, J. W., & Sung, A. (2016). The "Eye Avoidance" Hypothesis of Autism Face Processing. *Journal of Autism and Developmental Disorders*, *46*(5), 1538–1552. https://doi.org/10.1007/s10803-013-1976-7

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: a formal meta-analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(32), E3353–E3361. https://doi.org/10.1073/pnas.1409860111

Vogt, J., De Houwer, J., Koster, E. H. W., Van Damme, S., & Crombez, G. (2008). Allocation of spatial attention to emotional stimuli depends upon arousal and not valence. *Emotion* , *8*(6), 880–885. https://doi.org/10.1037/a0013981

Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., Bien, C. G., & Kissler, J. (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. In *Cortex* (Vol. 69, pp. 131–140). https://doi.org/10.1016/j.cortex.2015.05.003

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neuroscience and Biobehavioral Reviews*, *36*(3), 1060–1084. https://doi.org/10.1016/j.neubiorev.2011.12.008

Weiner, K. S., Jonas, J., Gomez, J., Maillard, L., Brissart, H., Hossu, G., Jacques, C., Loftus, D., Colnat-Coulbois, S., Stigliani, A., Barnett, M. A., Grill-Spector, K., & Rossion, B. (2016). The Face-Processing Network Is Resilient to Focal Resection of Human Visual Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(32), 8425–8440. https://doi.org/10.1523/JNEUROSCI.4509-15.2016

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience*, *8*(6), 820–827. https://doi.org/10.1038/nn1461

Westfall, J., Nichols, T. E., & Yarkoni, T. (2016). Fixing the stimulus-as-fixed-effect fallacy in task fMRI. *Wellcome Open Research*, *1*, 23.

https://doi.org/10.12688/wellcomeopenres.10298.2

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology*. https://doi.org/10.1146/annurev-psych-010416-044232

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*(4), 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Wright, B., Clarke, N., Jordan, J., Young, A. W., Clarke, P., Miles, J., Nation, K., Clarke, L., & Williams, C. (2008). Emotion recognition in faces and the use of visual context in young people with high-functioning autism spectrum disorders. *Autism: The International Journal of Research and Practice*, *12*(6), 607–626. https://doi.org/10.1177/1362361308097118

Zhou, E., Cao, Z., & Yin, Q. (2015). Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not? In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1501.04690

Appendix A

Supplemental Materials for Chapter 2: Judging Emotion in Natural Image

A.1 Instructions for Task

Ratings of the Full Image

On each trial you will briefly see a picture with faces in it. You will then be asked to judge the average emotion of the faces in the picture using three ratings. The first rating you will be asked to make is an average VALENCE rating. By average VALENCE, we mean how negative or positive the average emotion of the faces is. The second type of rating you will be asked to make is an average AROUSAL rating. By average AROUSAL, we mean how calm (low arousal) or exciting (high arousal) the average emotion of the faces is. The third and final type of rating you will be asked to make is an average DOMINANCE rating. High dominance is associated with the actual assumed or projected possession of power. Low dominance is associated with submissiveness.

Ratings of Context

On each trial you will briefly see a picture of a scene without faces. You will then be asked to judge the emotion of the scene using three ratings. The first rating you will be asked to make is a VALENCE rating. By VALENCE, we mean how negative or positive the emotion of the scene is. The second type of rating you will be asked to make is an AROUSAL rating. By AROUSAL, we mean how calm (low arousal) or exciting (high arousal) the emotion of the scene is. The third and final type of rating you will be asked to make is a DOMINANCE rating. High dominance is associated with the actual assumed or projected possession of power. Low dominance is associated with submissiveness.

Ratings of Individual Faces

On each trial you will briefly see a picture of a face. You will then be asked to judge the emotion of the face in the picture using three ratings. The first rating you will be asked to make is a VALENCE rating. By VALENCE, we mean how negative or positive the emotion of the face is. The second type of rating you will be asked to make is an AROUSAL rating. By AROUSAL, we mean how calm (low arousal) or exciting (high arousal) the emotion of the face is. The third and final type of rating you will be asked to make is a DOMINANCE rating. High dominance is associated with the actual assumed or projected possession of power. Low dominance is associated with submissiveness.

Faces Without Context and Scrambled Faces

On each trial you will briefly see a picture with faces in it. You will then be asked to judge the average emotion of the faces in the picture using three ratings. The first rating you will be asked to make is an average VALENCE rating. By average VALENCE, we mean how negative or positive the average emotion of the faces is. The second type of rating you will be asked to make is an average AROUSAL rating. By average AROUSAL, we mean how calm (low arousal) or exciting (high arousal) the average emotion of the faces is. The third and final type of rating you will be asked to make is an average DOMINANCE rating. High dominance is associated with the actual assumed or projected possession of power. Low dominance is associated with submissiveness.

A.2 Tables & Figures

Self-identified racial group	**White**	**Black**	**Asian**	**Native American, Alaska Native, Native Hawaiian, Pacific Islander**	**More than 1 race**	**Other**
Full Images						
Male	34	2	2	0	3	0
Female	14	5	0	0	1	1
Context Condition						
Male	30	3	3	1	0	1
Female	16	2	1	0	2	0
Individual Faces Condition: Set 1						
Male	28	6	1	1	2	0
Female	19	2	0	0	0	0
Individual Faces Condition: Set 2						
Male	33	5	3	0	0	1
Female	25	3	2	0	0	0
Individual Faces Condition: Set 3						
Male	22	5	1	0	2	0
Female	27	4	2	0	2	0

Age group	**18-25**	**26-30**	**31-40** (years)
Full Images			
Male	6	14	21
Female	1	9	11
Context Condition			
Male	7	19	12
Female	0	9	12
Individual Faces Condition: Set 1			
Male	2	15	21
Female	3	4	14
Individual Faces Condition: Set 2			
Male	5	23	14
Female	3	16	11
Individual Faces Condition: Set 3			
Male	4	10	16
Female	6	12	17

Table A.1: Demographics for Experiment 1. Participant numbers for Experiment 1 broken down by rating condition, gender and race (top panel) and by rating condition, gender and age group (bottom panel).

Self-identified racial group	**White**	**Black**	**Asian**	**Native American, Alaska Native, Native Hawaiian, Pacific Islander**	**More than 1 race**	**Other**
Faces without Context						
Male	41	4	3	0	4	0
Female	13	2	3	0	2	0

Age group	**18-25**	**26-30**	**31-40** (years)
Faces without Context			
Male	3	30	19
Female	4	6	10

Table A.2: Demographics for Experiment 2a. Participant numbers for Experiment 2a broken down by gender by race (top panel) and gender by age group (bottom panel).

Self-identified racial group	**White**	**Black**	**Asian**	**Native American, Alaska Native, Native Hawaiian, Pacific Islander**	**More than 1 race**	**Other**
Scrambled Faces						
Male	22	6	2	0	1	1
Female	18	8	1	0	1	2

Age group	**18-25**	**26-30**	**31-40** (years)
Scrambled Faces			
Male	3	16	13
Female	3	6	21

Table A.3: Demographics for Experiment 2b. Participant numbers for Experiment 2b broken down by gender by race (top panel) and gender by age group (bottom panel).

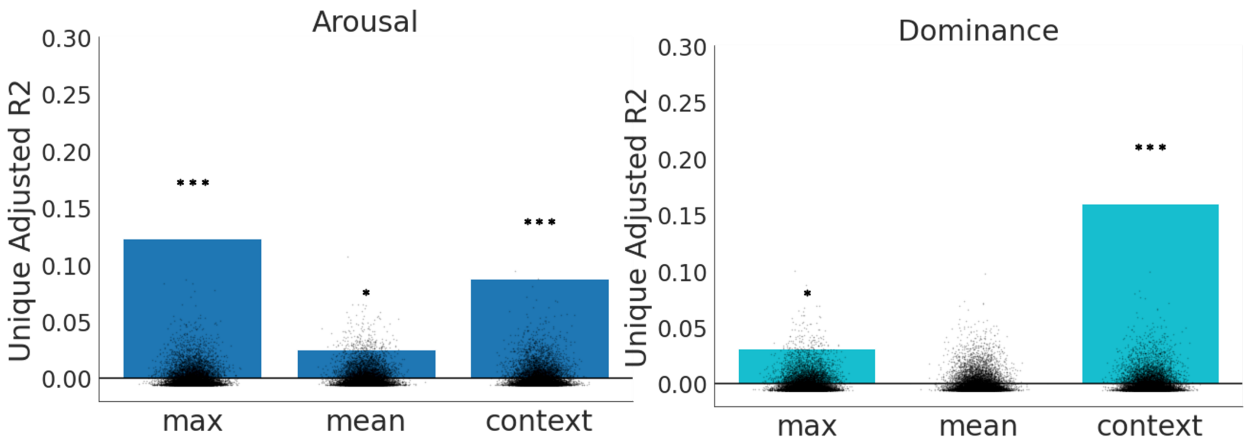


Figure A.1: Is the effect of max face emotion explainable by face size? A control analysis for Experiment 1. One possibility is that the maximally emotional face is often also the largest and that this might be driving the apparent influence of the most arousing or most dominant face upon ratings of the full image. To control for this, we used the ratings of arousal or dominance for the largest face in each image to predict the rating of the full image on the corresponding dimension of affect. We calculated the residual and used this as the dependent variable in a subsequent drop column analysis. This effectively controls for the ratings of the largest face on the dimension of interest. In the plots shown here, bars represent the unique variance explained by each feature as obtained from this drop column analysis. Scattered dots overlaying the plots show the permuted null distribution of the adjusted R^2 values. It can be seen that the maximally arousing ($R^2_{adj} = .12$, $p < .001$) and maximally dominant faces ($R^2_{adj} = .03$, $p = .012$) contribute significantly to the prediction of full image ratings, even when the variance explained by the arousal and dominance ratings of the largest face have been residualized out of these full image ratings.*$p < .05$, **$p < .01$, ***$p < .001$

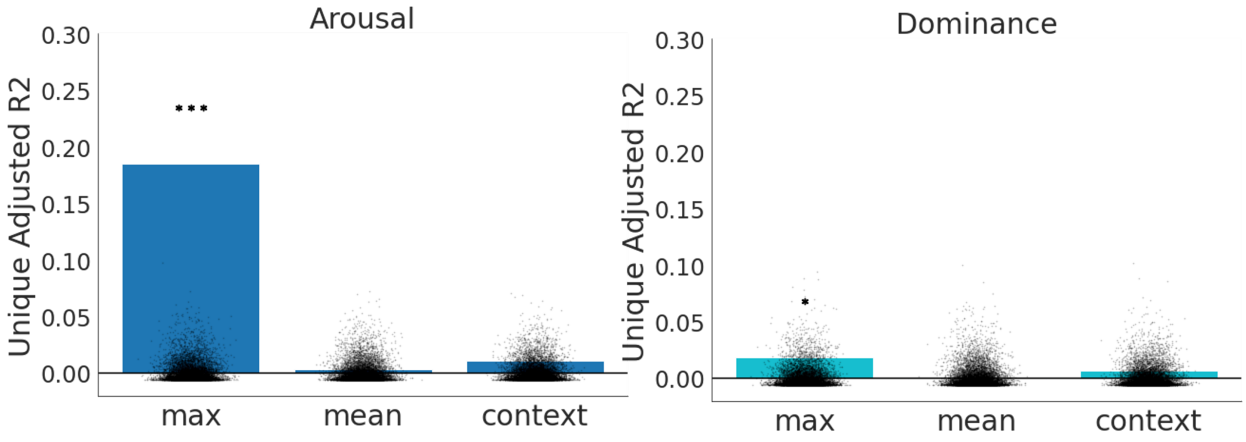


Figure A.2: Is the effect of max face arousal driven by max dominance or vice versa? A control analysis for Experiment 1. Another possible confound in Experiment 1 is that the MAX arousal face and the MAX dominance face might be the same face. If this is the case, then attentional orienting towards the most arousing face might have a knock-on influence on dominance ratings. To control for this, we used dominance ratings of the maximally arousing face to predict full image dominance ratings and arousal rating of the maximally dominant face to predict full image ratings of arousal. We then calculated the residuals and used these as the dependent variables in two drop column feature importance analyses. To the extent that the maximally arousing face is also the maximally dominant face, it will be prevented from contributing to the unique variance explained by the MAX feature in either drop column analysis. In the plots shown here, bars represent the unique variance in the residual full image arousal ratings (left panel) and dominance ratings (right panel) explained by each feature. Dots overlaying the bars show the permuted null distribution of the adjusted R^2 values. It can be seen that the maximally arousing ($R^2_{adj} = .18$, $p < .001$) and maximally dominant faces ($R^2_{adj} = .02$, $p = .046$) still contribute significantly to the prediction of full image ratings. *$p < .05$, **$p < .01$, ***$p < .001$

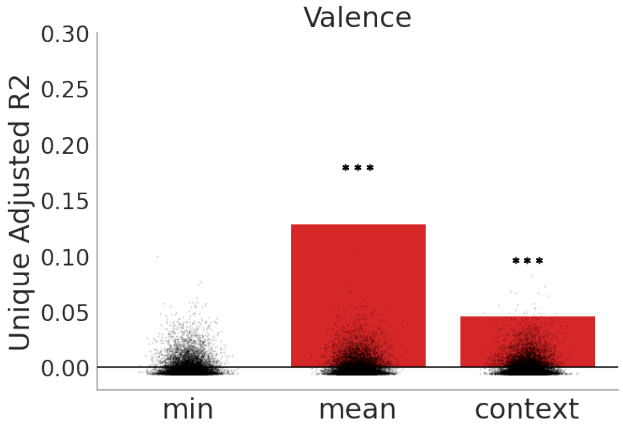
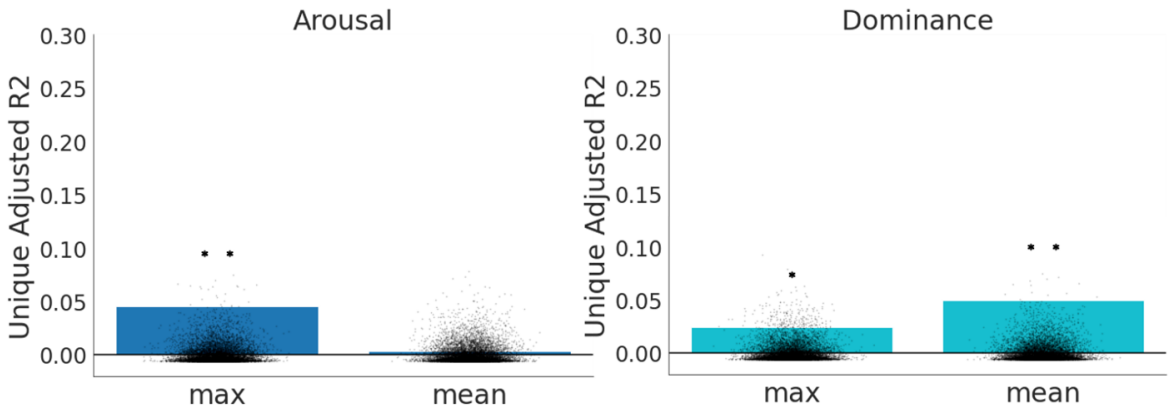


Figure A.3: Does the min face for valence contribute significantly to full image ratings for Experiment 1? A control analysis. For valence, the scale's ends are defined as negative (rating of 1) and positive (rating of 9). Thus, the minimally positive face (MIN) is also the maximally negative face. We hence conducted a drop column feature analysis with the MIN (positive) valence face regressor replacing the MAX (positive) valence face regressor to ensure that there was no unique effect of this regressor. The MIN (positive) valence face regressor did not explain a significant level of unique variance for valence ($R^2_{adj} < .01$, $p = .30$). $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

A. Experiment 2a



B. Experiment 2b

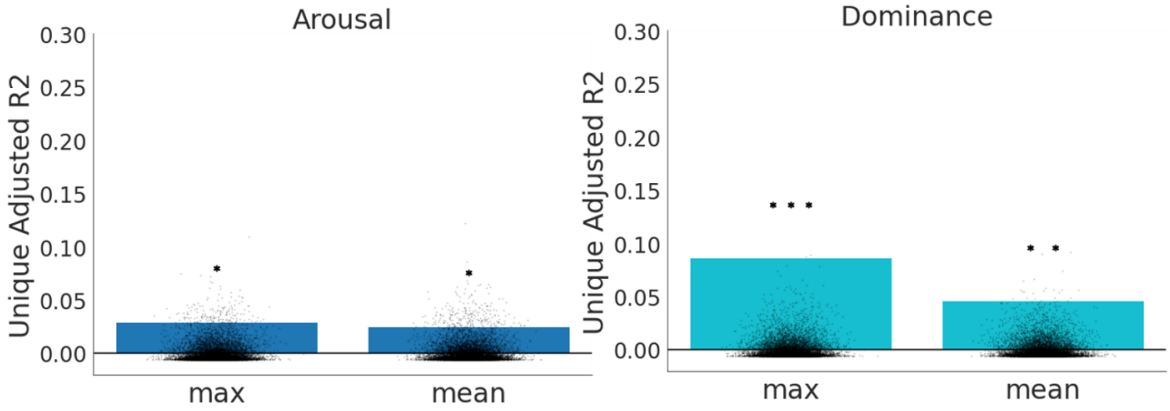


Figure A.4: Is the effect of max face emotion driven by face size? A control analysis for Experiment 2. Ratings of arousal and dominance for the largest face in each image were used to predict the rating of group facial arousal and dominance, respectively, for the Faces without Context stimuli in Experiment 2a (panel A) and for the Scrambled Faces stimuli in Experiment 2b (Panel B). The residuals were used as dependent variables in drop column feature importance analyses for the dimensions of arousal (left) and dominance (right), respectively. This controls for any confounding influence of the maximally arousing or dominant face also being the largest in the image. In the plots shown here, bars represent the unique variance in these residual ratings explained by the maximally arousing (left) or dominant (right) face and by mean facial arousal (left) and dominance (right). Scattered dots overlaying the bars show the permuted null distribution of the adjusted R^2 values. In both Experiment 2a (Panel A) and Experiment 2b (Panel B), the maximally arousing face continued to explain unique variance in the ratings of group arousal after variance explained by arousal ratings for the largest face were regressed out ($R^2_{adj} = .05$, $p = .002$; $R^2_{adj} = .03$, $p = .016$, respectively). With the dominance ratings of the largest face regressed out, the maximally dominant face also explained a significant amount of unique variance in group facial dominance ratings across both Experiments 2a (Faces without Context) and Experiment 2b (Scrambled Faces) ($R^2_{adj} = .02$, $p = .024$; $R^2_{adj} = .09$, $p < .001$, respectively). $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

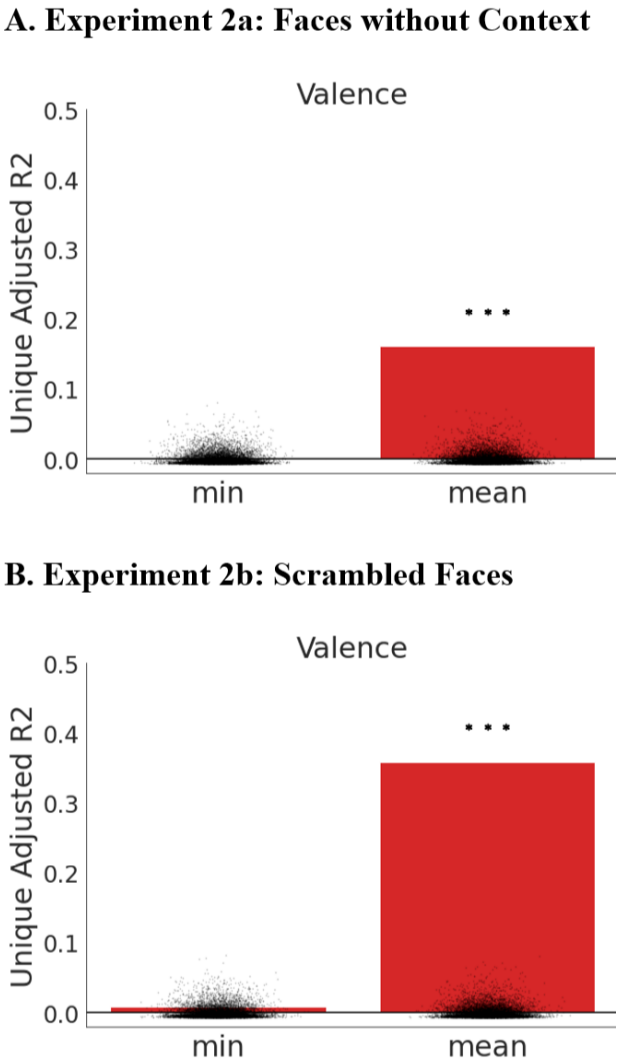


Figure A.5: Does the MIN (negative) valence regressor (i.e., the maximally negative face) contribute significantly to Faces without Context and Scrambled Faces ratings in Experiment 2? For valence, the scale's ends are defined as negative (rating of 1) and positive (rating of 9). Thus, the minimally positive face (MIN) is also the maximally negative face. We hence conducted drop column feature analyses with the MIN (negative) valence face regressor replacing the MAX (positive) valence face regressor for both Experiment 2a and Experiment 2b. Panel A shows the results of the Experiment 2a drop column analysis with ratings of the Faces Without Context stimuli as the dependent variable. Panel B shows the results of the Experiment 2b drop column feature analysis with ratings of the Scrambled Faces stimuli as the dependent variable. In each case, ratings of the minimally positive valence face (MIN) (i.e., the most negative face) did not account for significant unique variance ($R^2_{adj} < .01$, $p = .347$; $R^2_{adj} = .01$, $p = .120$, respectively) in group valence ratings. The mean facial valence regressor continued to explain significant variance in the group facial valence ratings across both analyses. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Appendix B

Supplemental Materials for Chapter 3: Cortical Representation of Mid-Level Face Features

B.1 Methods

Eyetracking

Individual Differences

Summary statistic measures for eyetracking were obtained for each participant. Measures such as fixation inside stimulus box (in_box), mean number of fixations (nfix_mean), standard deviation of number of fixations (nfix_std), the mean x position of fixations (xpos_mean), the standard deviation across of x position across fixations (xpos_std), the mean y position of fixations (ypos_mean), the standard deviation of y position across fixations (ypos_std), the mean eccentricity of fixations (ecc_mean), and the standard deviation of the eccentricity across fixations (ecc_std) were calculated across all trials. For all summary statistic metrics except number of fixations and standard deviation of number of fixations were calculated using the first fixation in each trial. These measures were then correlated with our measures of individual differences (AQ and CFMT), see Figure B.1. Significance of correlations were FDR corrected within each individual difference measure with a threshold of $q[FDR] < .05$.

Relationship to Model Performance

Using eyetracking summary statistics, we correlated these metrics with the number of significant voxels ($p < .05$) for each ROI while controlling for ROI size across participants. P-values were FDR corrected within each ROI with a signifcance threshold of $q[FDR] < .05$. This analysis was performed for each mid-level model (Figure B.2).

Independent Component Analysis

PCA

Principal components analysis (PCA) was used initially on model beta weights across voxels and across participants to understand the underlying organization mid-level face features across OTC. To select the voxels to use for PCA, we had the criteria as follows: 1) voxels must be in OTC, 2) validation correlation as assessed by an image on/off model (whether an image was present or absent) must be greater than 0.05, and 3) estimated beta weights for voxels must not all be 0. We then looked at two broad regions of interest (1. retinotopic areas: V1-V4 and 2. higher-level regions: areas outside of retinotopic regions but within OTC) and did PCA analysis on each region of interest separately for each model. When conducting PCA on higher-level regions, we took the top 2000 best performing voxels using the cross validated correlation on the estimation set when selecting for the lambda. Then to match the statistical properties of the higher-level regions, we obtained the number of voxels in retinotopic areas that would match the average minimum correlation of the top 2000 voxels we selected in the higher-level regions for each model (Fiducial Coordinate: 1249 voxels, min $R = 0.03$; Fiducial Attribute: 1227 voxels; min $R = 0.03$; Fiducial Texture: 917 voxels, min $R = 0.02$).

When performing group PCA, we concatenated all beta weights for selected voxels across participants for each model and for each region of interest. For each participant, we demeaned the weights before concatenating. Some structure in feature tuning across voxels may be capturing covariance in the stimulus features. Thus, for the Fiducial Coordinate and Fiducial Attribute model, we identified the top PCs where the variance explained was more than the stimulus features alone. Variance explained was calculated by using an out of sample approach where we conducted PCA on voxel beta weights for all participants except one the participant of interest and then used the right singular matrix to predict the beta weights for the participant of interest. We did this for each participant and averaged the out of sample variance explained across participants for each PC to obtain the variance explained for that PC. To calculate the significant top PCs to use, we permuted the feature matrix for each model and where beta weights were fit on this permuted feature matrix. We then used the true out of sample right singular matrix to predict the participant's permuted beta weights. This was done 500 times, and only PCs where the true explained variance was greater than 95% of the permuted out of sample explained variance ($p < 0.05$) were kept (Figure B.12).

For the Fiducial Texture model, since the feature matrix are comprised of PCs obtained from doing PCA on pixel values on morphed faces, the features are inherently independent. Thus, we cannot use the same method to determine the top PCs since there does not exist stimulus feature covariance. We then used the "elbow" method to select top PCs by looking at where the out of sample explained variance starts flattening out (Figure B.12).

ICA of Fiducial Coordinate Model Feature Weights

The distribution of voxel loadings onto PCs were non-Gaussian so we decided to use ICA to control for higher-level statistical dependencies. We selected the number of ICs to use based off of how many PCs we decided to keep in the analyses above.

Interpreting Top Group ICs

To understand what each of the ICs represents, we used hypothetical dimensions across images to find the hypothetical dimension loadings onto each IC. Hypothetical dimensions included ratings of semantic binary emotion (anger, sad, annoyed, etc.), ratings of semantic dimensional emotion (valence, arousal, dominance, and approachability), ratings of low-level identity features (mouth wide open, oval face, etc.), ratings of high-level identity features (Caucasian, black, Asian, etc.), and pose information (pitch, roll, yaw, head areas). To get the hypothetical dimension loading onto the ICs, we took the dot product of *images x mid-level features* and the *mid-level features x IC* matrices. The resulting matrix was *images x IC*. Then, we took the dot product of emphhypothetical dimensions x images and *images x IC* to give us our *hypothetical dimensions x IC* matrix (Figure B.13).

B.2 Tables & Figures

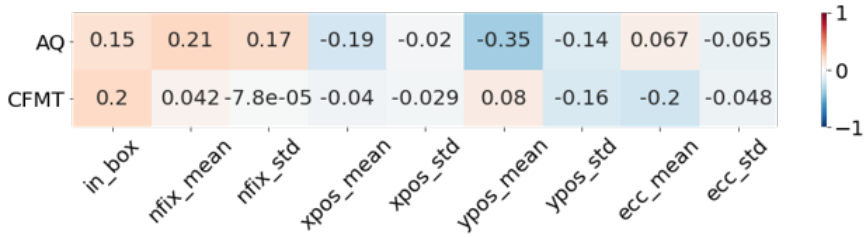


Figure B.1: Correlation matrix of eyetracking summary statistics correlated with AQ and CFMT. Summary statistics for eyetracking were calculated for each participant across fMRI trials. Each cell shows the correlation value between a given individual difference metric with an eyetracking summary statistic measure. Significance was calculated as the FDR corrected p value (corrected within each individual difference measure) with a threshold of $q[FDR] < .05$. There were no significant correlations between the individual differences measures and eyetracking summary statistics.

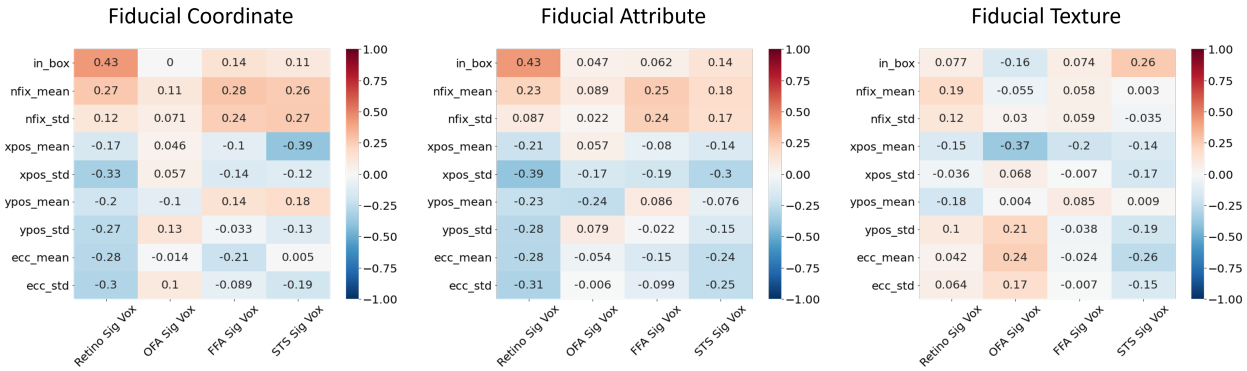


Figure B.2: Correlation matrix of eyetracking summary statistics correlated with model performance. Each cell shows the correlation value between a given eyetracking statistic with the number of significant voxels in a ROI (controlling for ROI size). Significance was calculated as the FDR corrected p value (corrected within each ROI) with a threshold of $q[FDR] < .05$. There were no significant correlations between the the number of significant voxels in a ROI and eyetracking summary statistics.

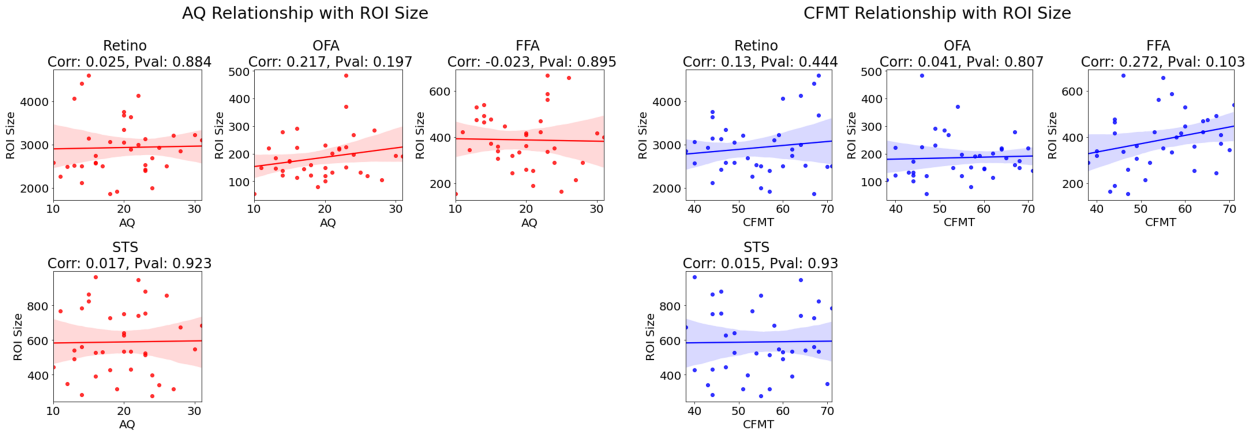


Figure B.3: ROI size correlated with AQ and CFMT. On the left panel, we correlated size (number of voxels) of functionally defined ROIs: Retino, OFA, FFA, and STS with AQ. We found that there were no significant correlations. On the right panel, we correlated the ROI size with CFMT and did not observe any significant correlations.

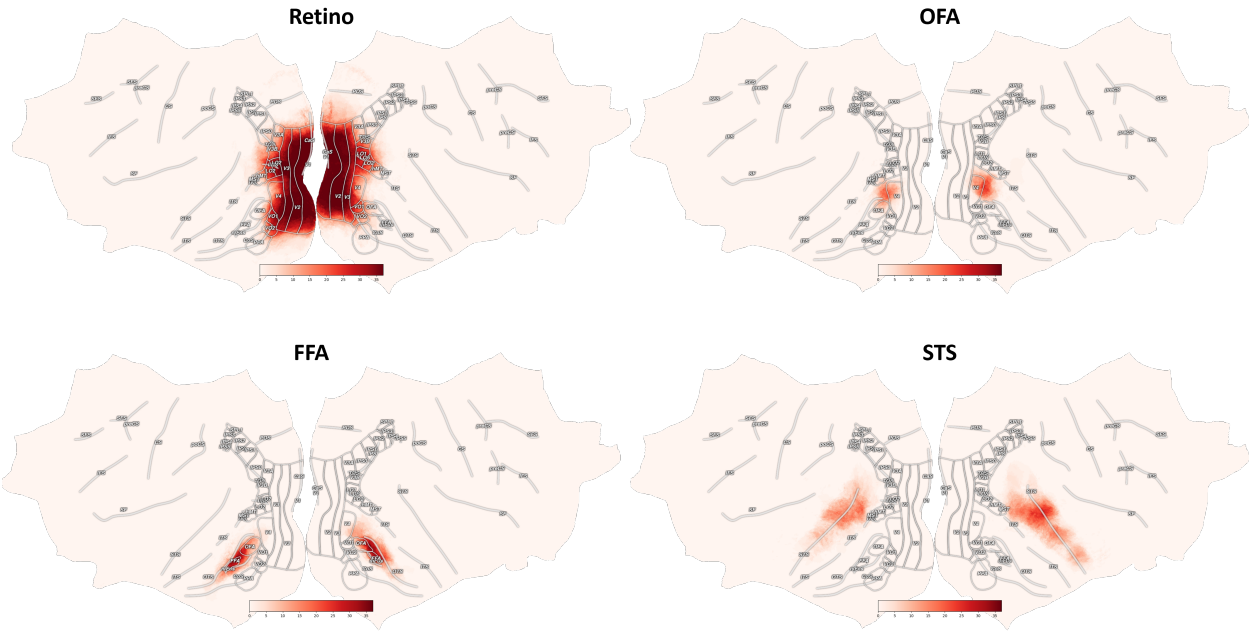


Figure B.4: Functional ROIs across participants plotted in fsaverage space. Individual cortex maps with functionally defined ROI masks were transformed into fsaverage space. Intensity of red indicates the number of participants whose functional ROIs occurred at that vertex with darker red corresponding to a larger number of participants whose ROI overlaps within that vertex.

Higher Identity	Lower Identity	Higher Emotion	Lower Emotion
asian	5 o'clock shadow	adoring	arched eyebrows
black african	bags under eyes	afraid	eyes closed
child	bald	amazed	mouth closed
female	bangs	amused	mouth slightly open
hispanic or latino	beard	angry	mouth wide open
indian	big eyes	annoyed	squinting
middle aged	big lips	anticipating	teeth not visible
middle eastern	black hair	cautious	teeth very prominent
senior	blond hair	confident	
teenager	blue eyes	confused	
white caucasian	brown eyes	contented	
young adult	brown hair	curious	
	bushy eyebrows	dejected	
	curly hair	delighted	
	double chin	disapproving	
	facial hair	disgusted	
	goatee	distressed	
	gray hair	excited	
	green eyes	frustrated	
	mustache	grief-struck	
	oval face	happy	
	receding hairline	joyful	
	red hair	neutral	
	round face	offended	
	short hair	pained	
	small eyes	peaceful	
	square face	playful	
	straight hair	pleased	
	thin eyebrows	sad	
	very skinny	shocked	
	wavy hair	smiling	
	waxed eyebrows	stern	
	white hair	surprised	
		suspicious	
		thoughtful	
		worried	

Table B.1: Four semantic sub-models created to understand the relationship between mid-level models and high-level models. Binary features were labeled by 3 in-lab raters and agreement of two raters was used as the final value (0/1). Raters were asked whether an image contained a certain feature (1) or did not (0). Feature matrix construction and model fit procedure were identical to that described in the Methods section.

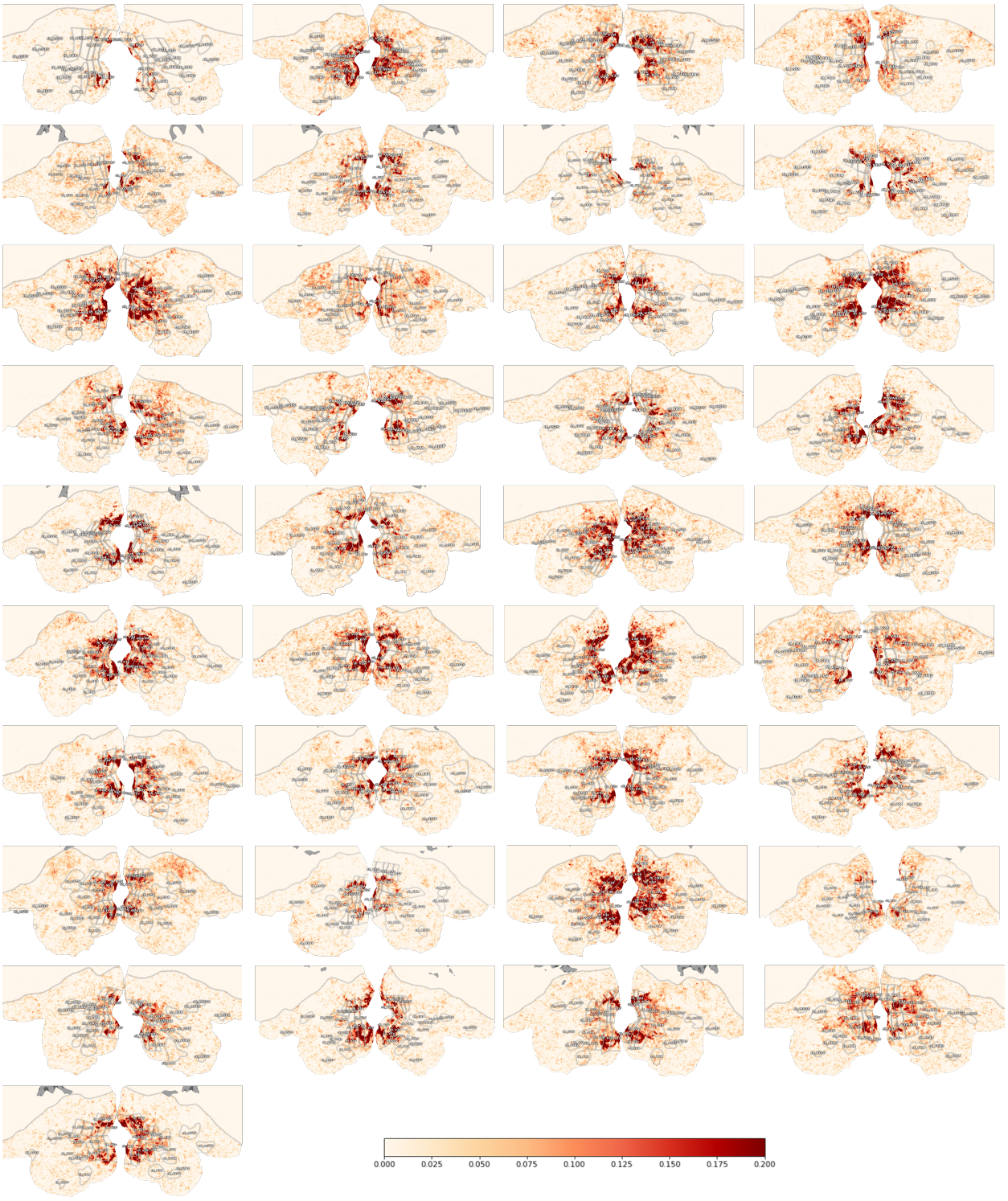


Figure B.5: Fiducial Coordinate model performance (validation correlation) on individual 37 participant cortex maps zoomed in on OTC. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction.

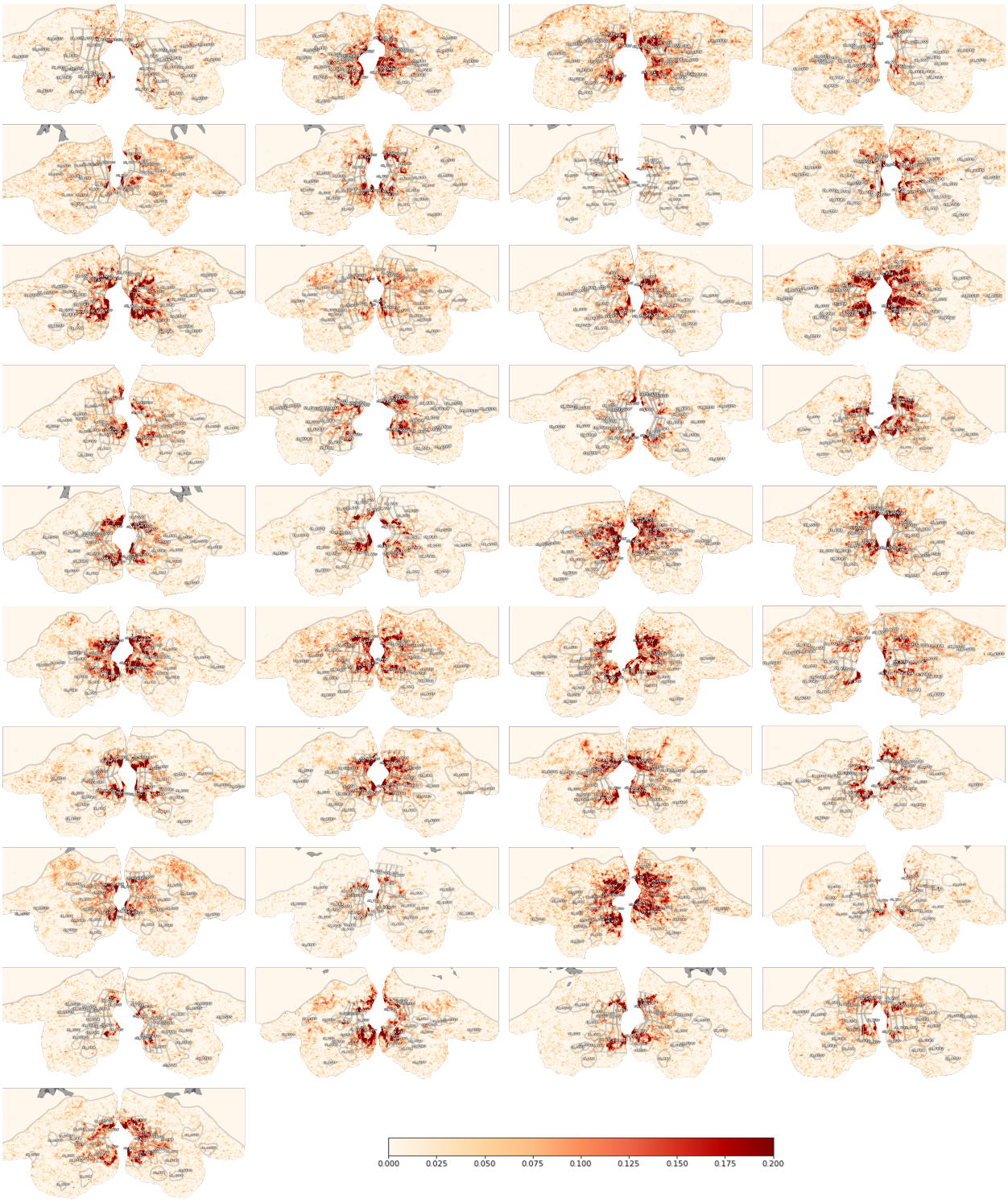


Figure B.6: Fiducial Attribute model performance (validation correlation) on individual 37 participant cortex maps zoomed in on OTC. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction.

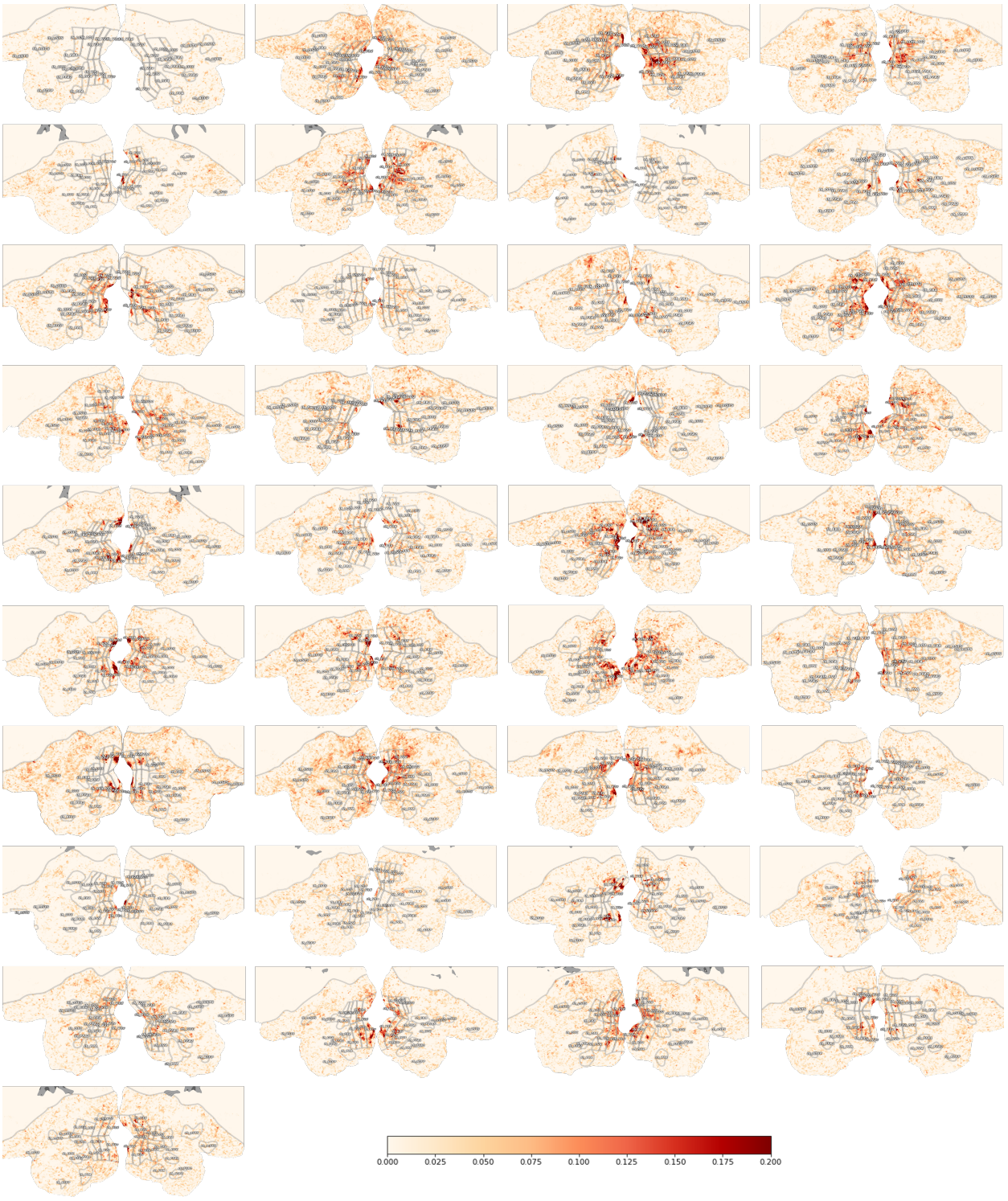


Figure B.7: Fiducial Texture model performance (validation correlation) on individual 37 participant cortex maps zoomed in on OTC. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction.

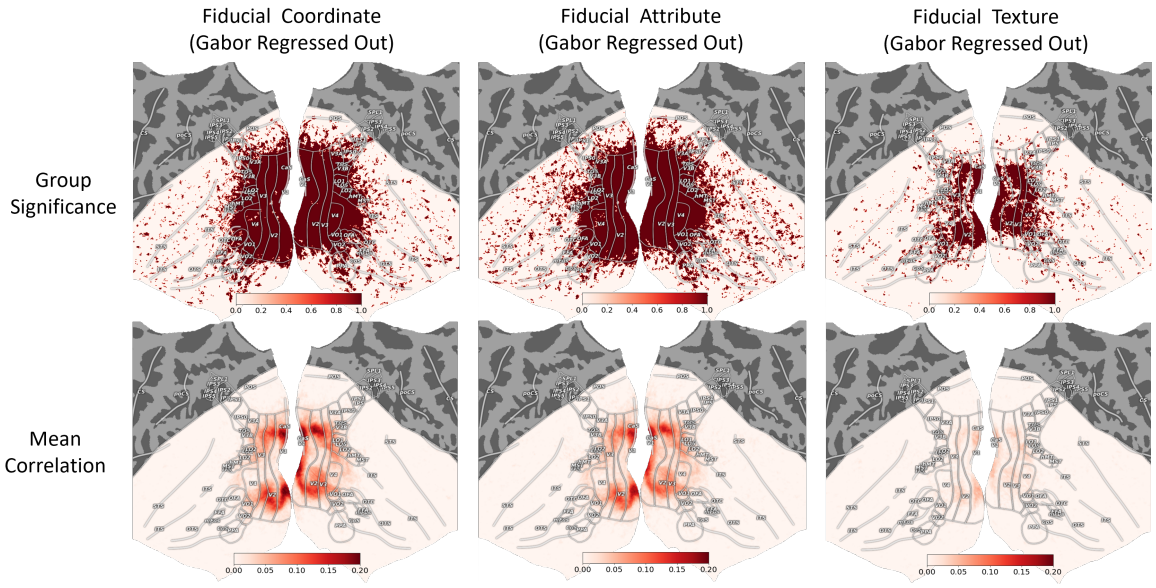


Figure B.8: Mid-level model performance in fsaverage with low level features regressed out. A low-level Gabor model was fit in OTC and residuals were obtained. The mid-level models were then fit on those residuals to control for low level statistical dependencies. We found that even when regressing the Gabor model out, there was still high model performance in retinotopic areas. In the top panel, we have corrected ($q[FDR] < .05$) group significance maps that were calculated using binomial t-tests. The values for each vertex were an array of binary numbers indicating whether that vertex was significant ($p < .05$, uncorrected) for an individual participant as calculated from permutation tests. Red indicates that the vertex was significant after FDR correction (across OTC) for the group.

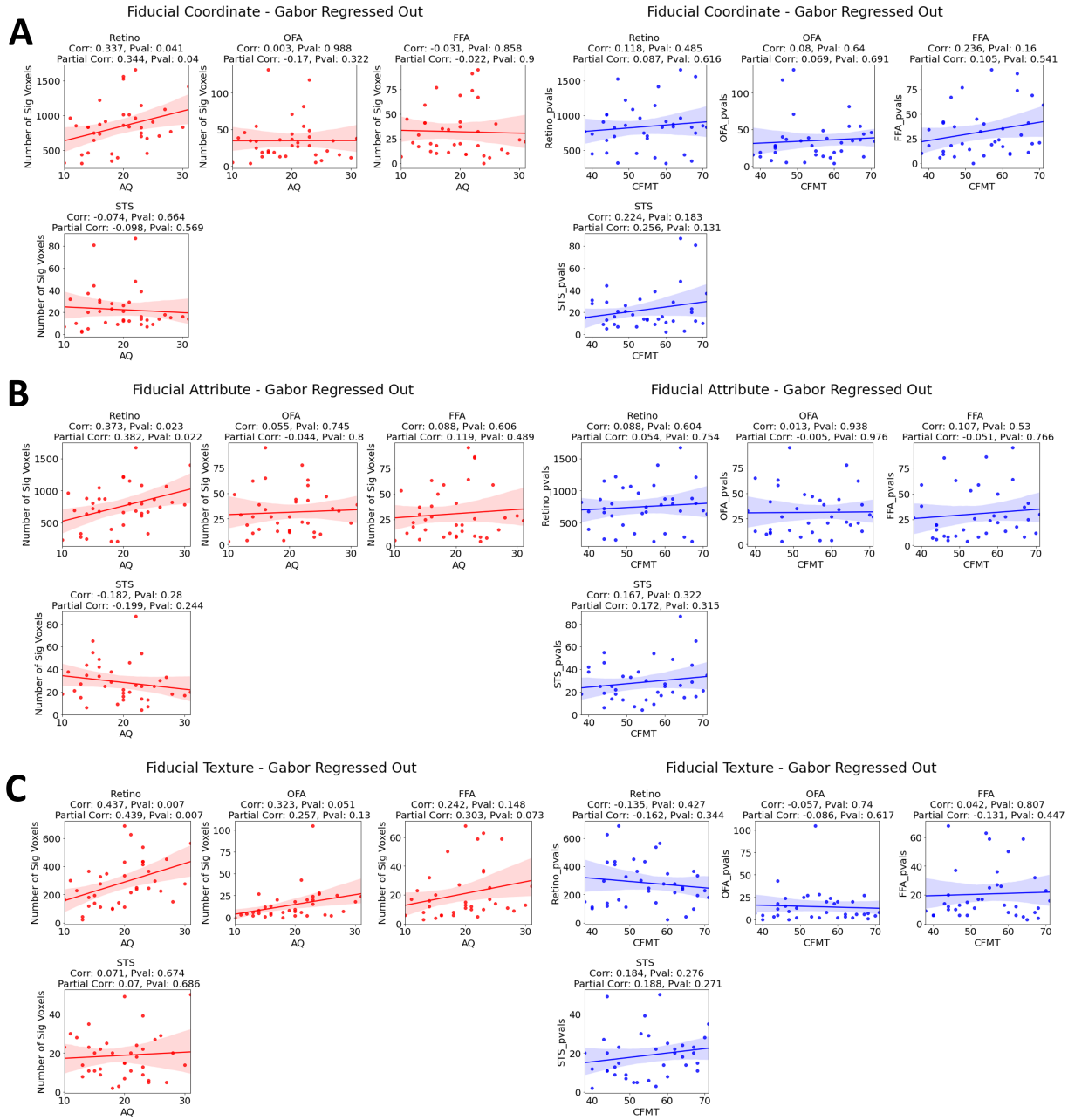


Figure B.9: Mid-Level model performance fit on Gabor residuals (low-level information regressed out) correlated with AQ (red) and CFMT (blue). For each participant, the number of significant voxels in each ROI were calculated from permutation tests ($p < .05$, uncorrected) for each model (top panel A: Fiducial Coordinate, middle panel B: Fiducial Attribute, bottom panel C: Fiducial Texture) and correlated with AQ and CFMT. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation as well as the partial correlation and partial p value when controlling for ROI size. We show that across all models that the number of significant voxels in retinotopy is significantly correlated with AQ when controlling out for ROI size.

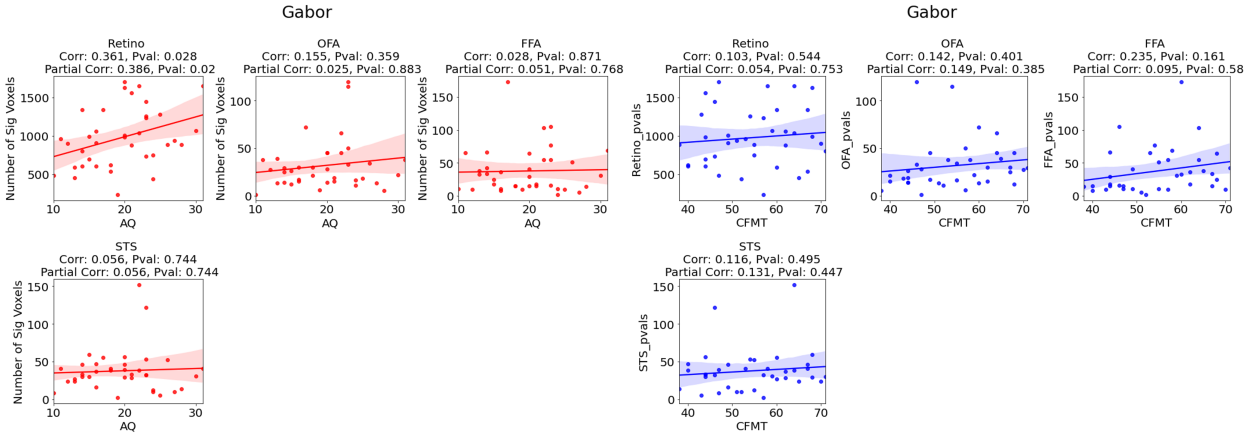


Figure B.10: Gabor model fit correlated with AQ (red) and CFMT (blue). For each participant, the number of significant voxels in each ROI were calculated from permutation tests ($p < .05$, uncorrected) for the Gabor model and correlated with AQ and CFMT. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation as well as the partial correlation and partial p value when controlling for ROI size. We show there is a significant relationship between Gabor model performance in retinotopy and AQ.

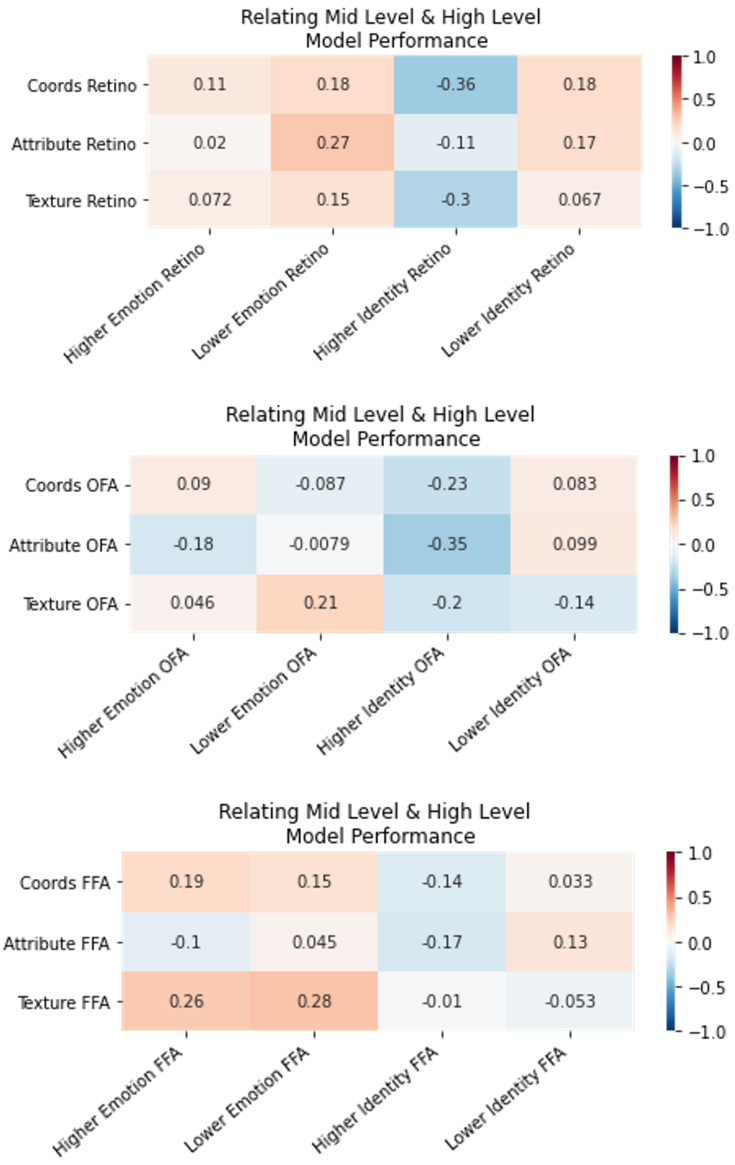


Figure B.11: Correlation matrix of model performance in each ROI (retinotopy: top, OFA: middle, FFA: bottom) for mid-level models with high-level model. Model performance was calculated as the percentage of voxels that were significant ($q[FDR] < .05$). Each cell shows the correlation value between a given mid-level model with a given high-level model. Black boxes denote significant correlation values ($q[FDR] < .05$). There were no significant correlations observed in any of the ROIs suggesting that there are not competing representations of mid-level vs high-level information within an ROI.

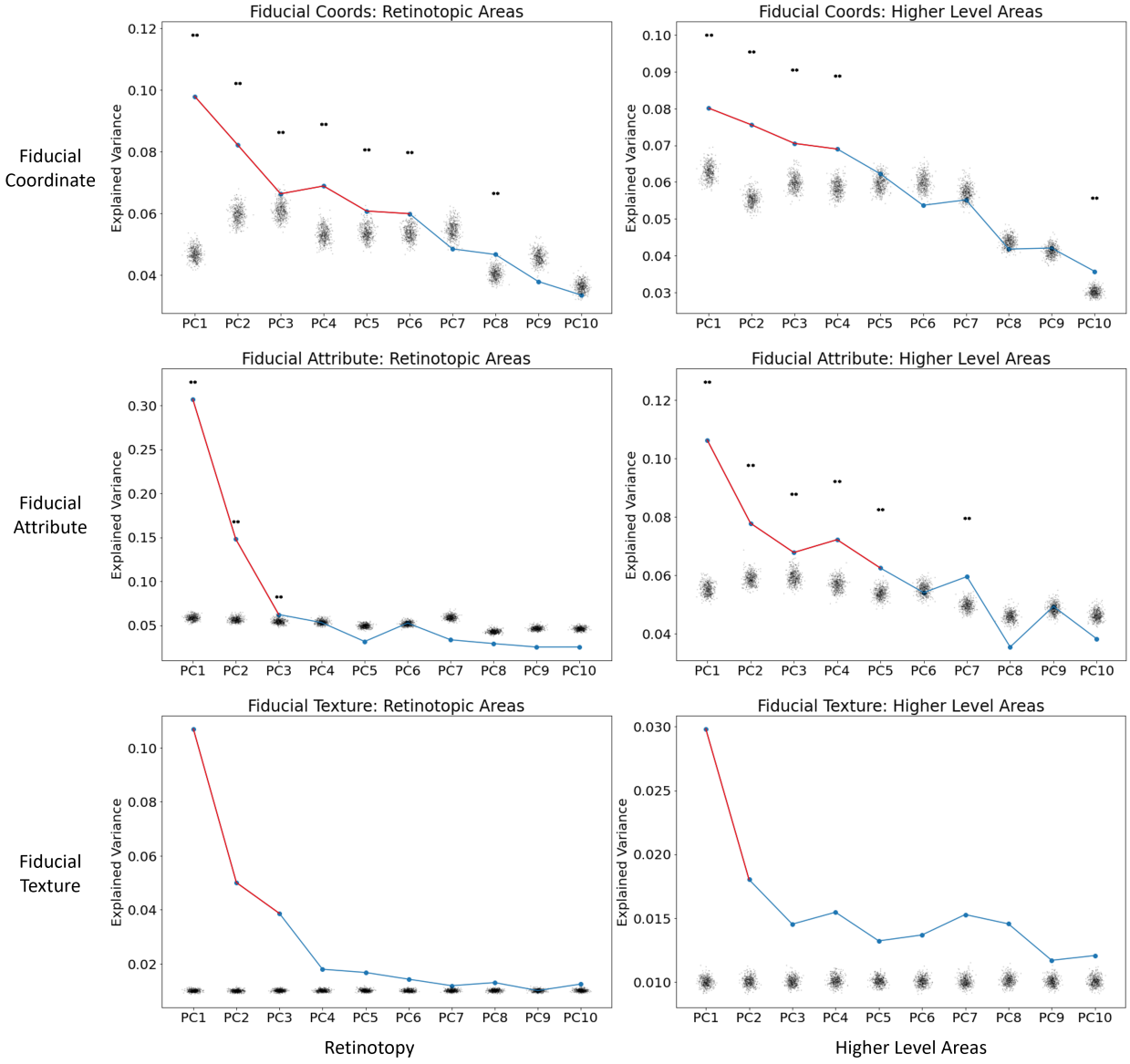


Figure B.12: PC Explained Variance for the Fiducial Coordinate (top), Fiducial Attribute (middle), and Fiducial Texture (bottom) models in retinotopic areas (left panel) and higher-level areas (right panel). Solid lines represent out of sample variance explained calculations for each PC. Scattered dots in the Fiducial Coordinate and Fiducial Attribute graphs show the permuted variance explained. For the Fiducial Texture model, we did not use permutation tests to select PCs but rather the elbow method. Red denotes the PCs that were kept for ICA analyses (Fiducial Coordinate: 6 for Retinotopic, 4 for Higher-level Areas; Fiducial Attribute: 3 for Retinotopic, 5 for Higher-level Areas; Fiducial Texture: 3 for Retinotopic, 2 for Higher-level Areas).

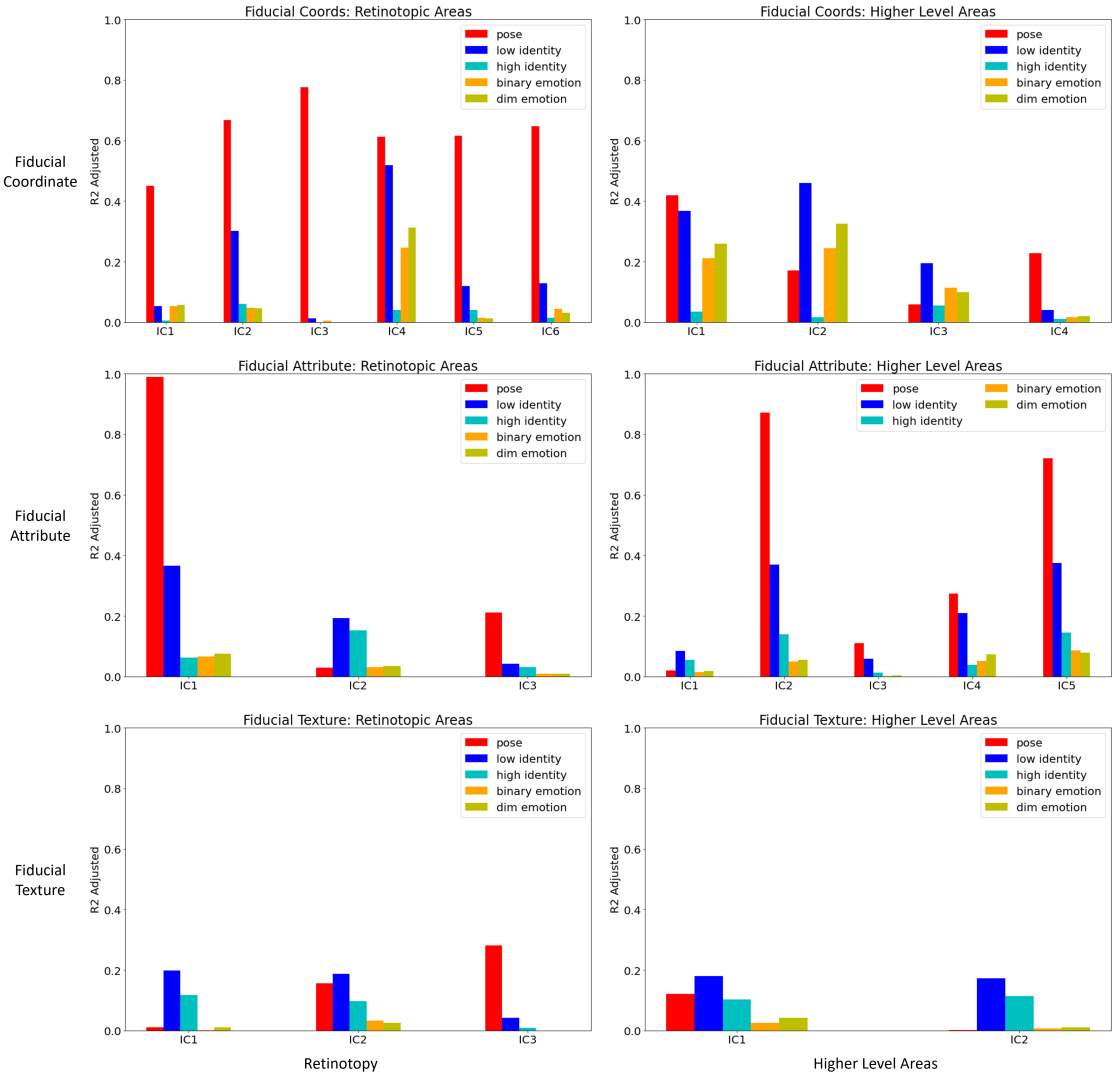


Figure B.13: Hypothetical Dimension Group Regression Analyses. We grouped hypothetical dimension ratings on images into 5 groups: pose (red bars), low-level identity (blue bars), high-level identity (cyan bars), binary emotion (orange bars), and dimensional emotion (yellow bars). We then did separate regression analyses for each group where the hypothetical dimensions were the independent variables and the image loadings onto the ICs were the dependent variables. Adjusted R2 values are represented by the bar height. We see that pose, low-level identity, and emotion information is represented in both retinotopic and higher-level areas for the Fiducial Coordinate model (with pose having lower variance explained in higher level regions). For the Attribute model, we see less representation of emotion, but more identity features explaining higher variance. For the Texture model, we only see pose and identity information explaining high variance.

Appendix C

Supplemental Materials for Chapter 4: Cortical Representation of High-Level Semantic Face Features

C.1 Figures

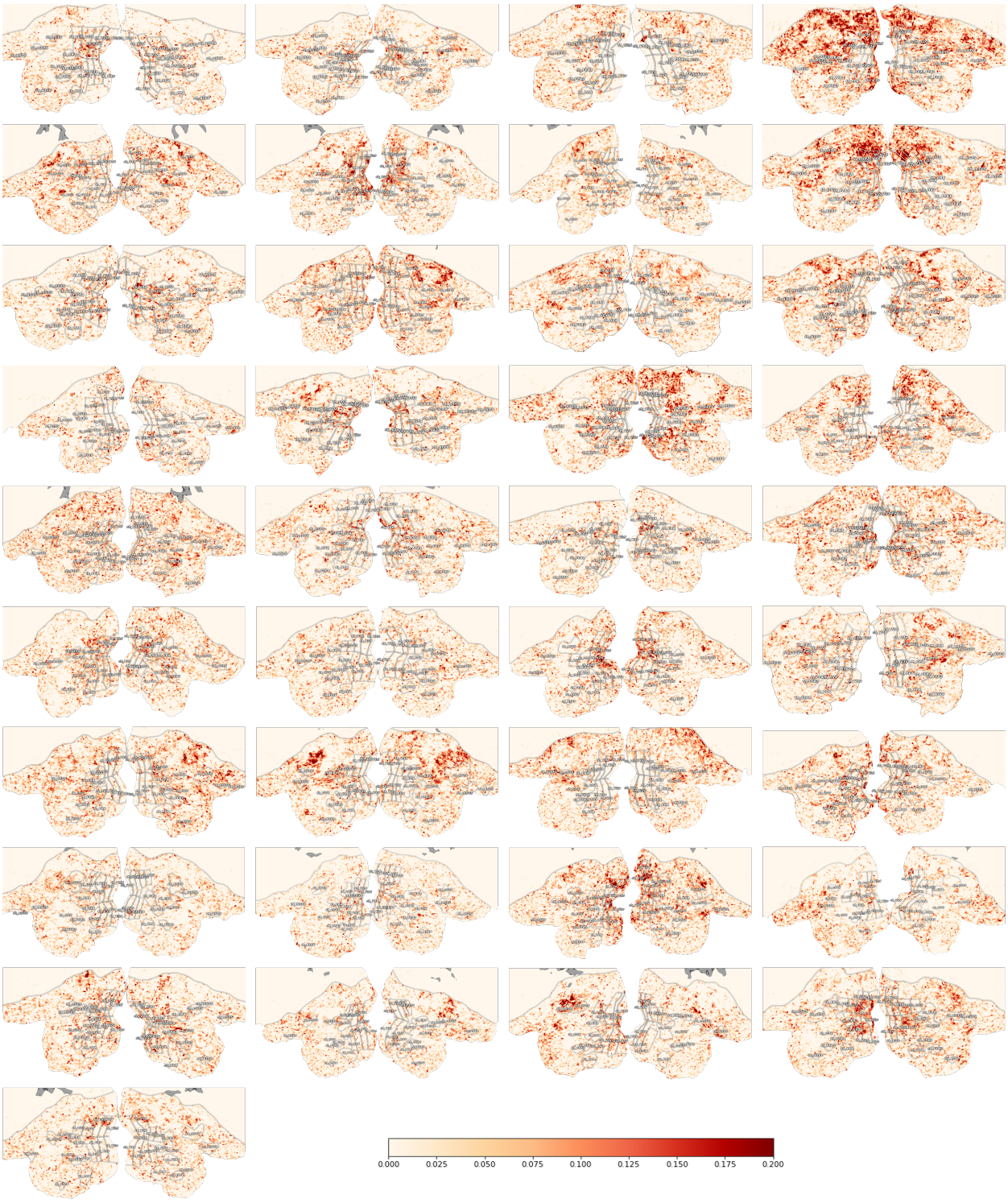


Figure C.1: Semantic Binary model performance (validation correlation) on individual 37 participant cortex maps zoomed in on OTC. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction.

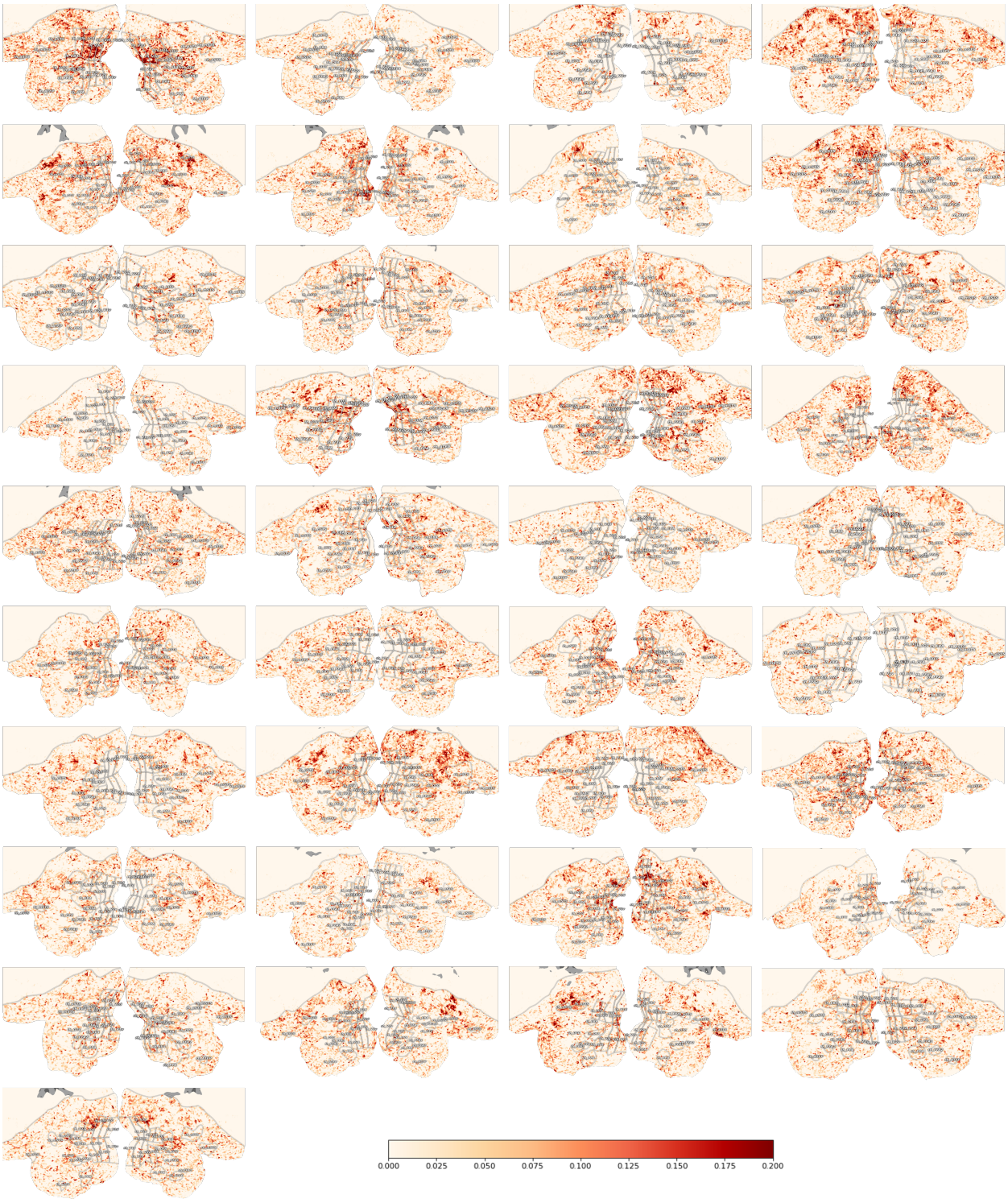


Figure C.2: Identity model performance (validation correlation) on individual 37 participant cortex maps zoomed in on OTC. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction.

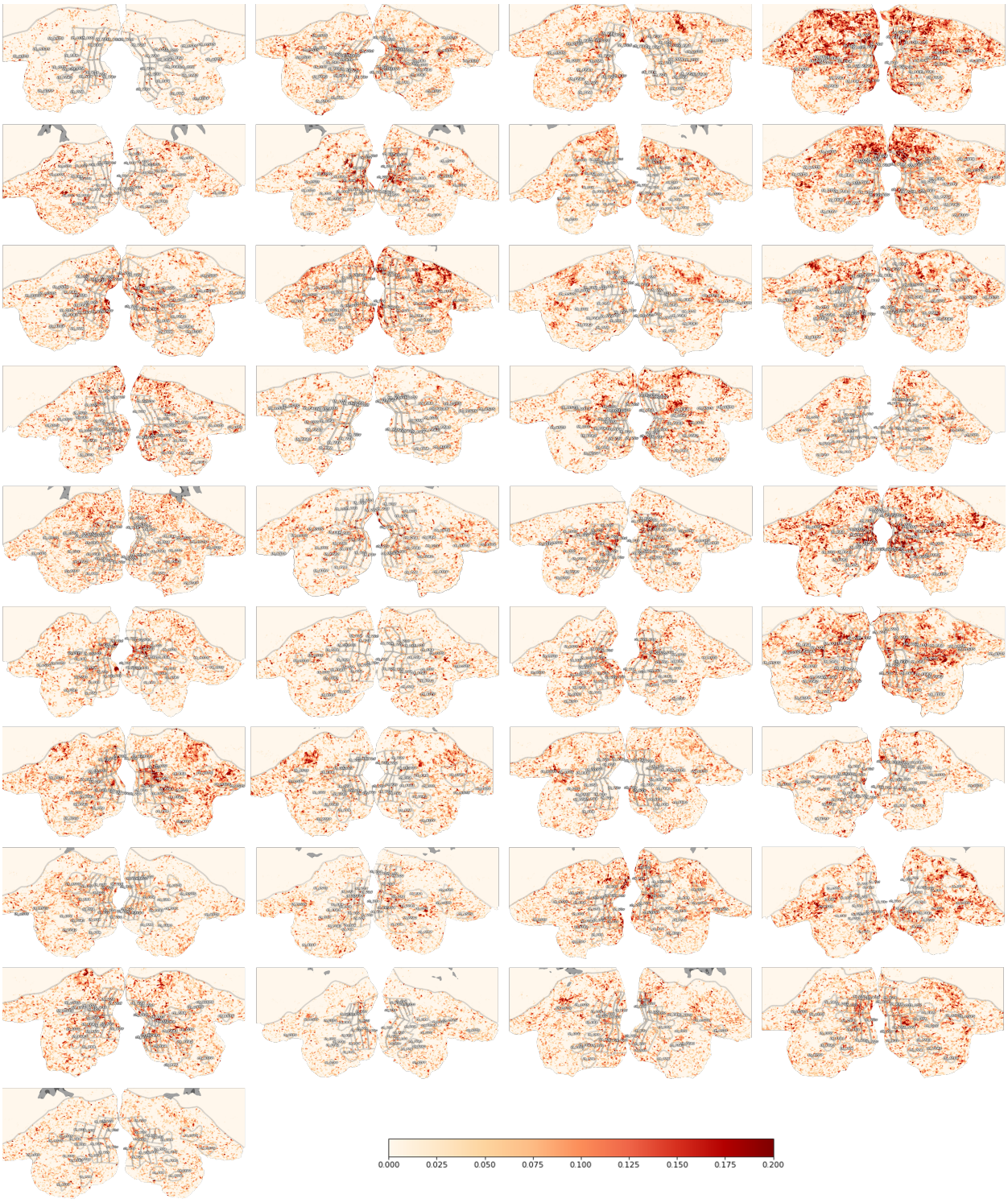


Figure C.3: Emotion model performance (validation correlation) on individual 37 participant cortex maps zoomed in on OTC. Red shows a maximum validation correlation of 0.2 and white showing the minimum correlation of 0. The intensity of the color corresponds to the strength of model prediction.

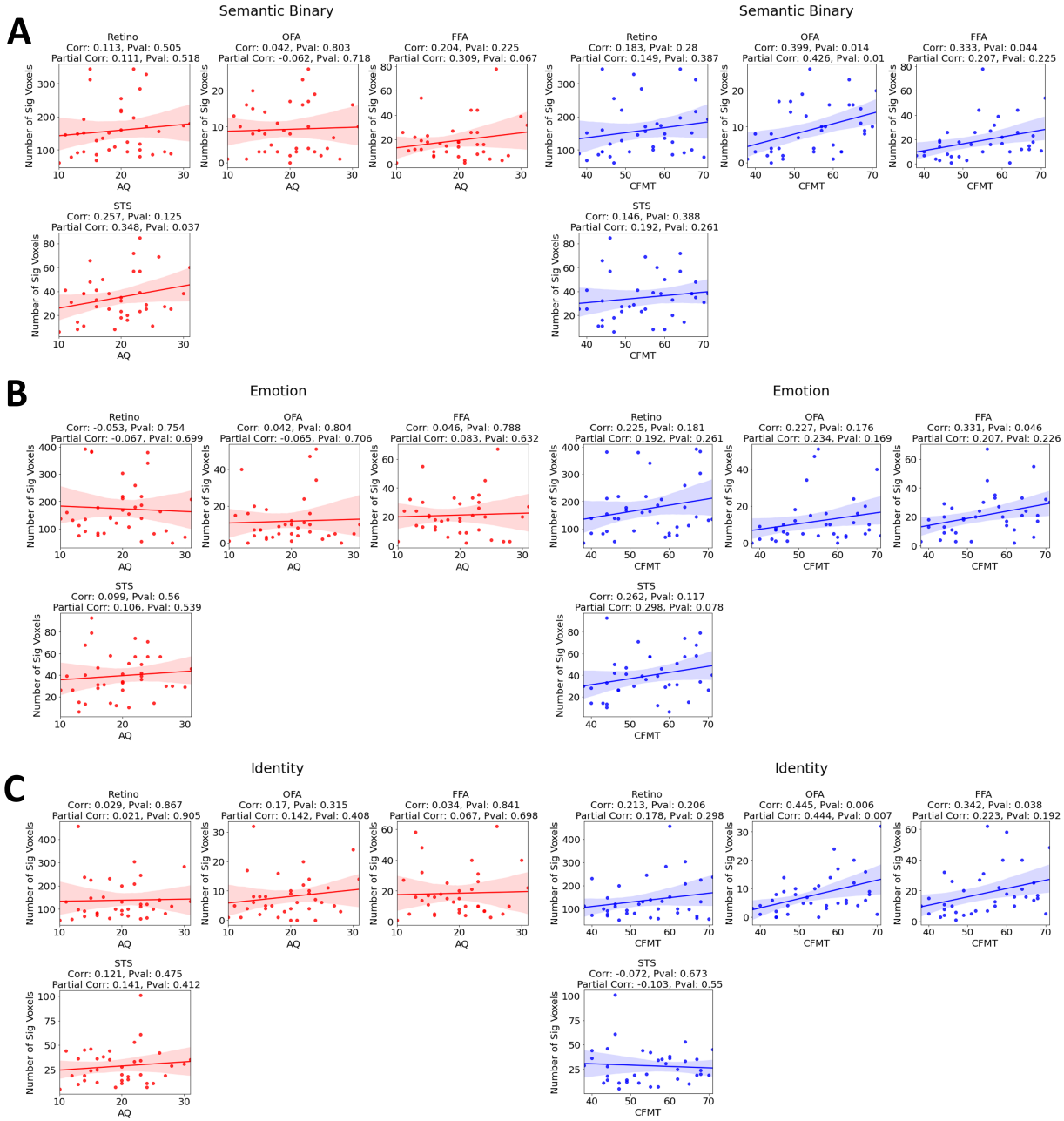


Figure C.4: Model performance correlated with AQ (red) and CFMT (blue) for the three main semantic models. For each participant, the number of significant voxels in each ROI were calculated from permutation tests ($p < .05$, uncorrected) for each model (top panel A: Semantic Binary, middle panel B: Identity, bottom panel C: Emotion) and correlated with AQ and CFMT. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation as well as the partial correlation and partial p value when controlling for ROI size. We see a significant relationship between AQ and the Semantic Binary model in STS as well as significant correlations between all the models in OFA and CFMT.

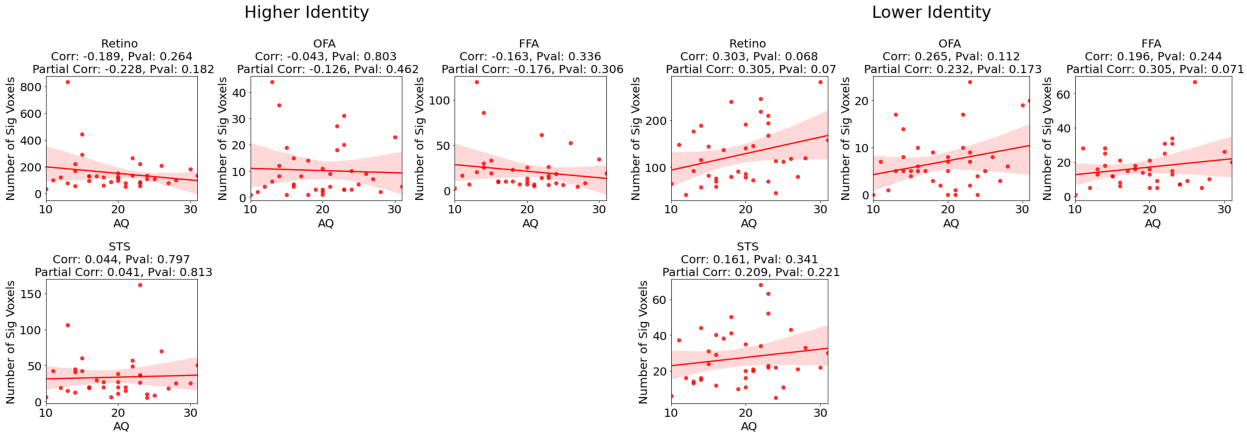


Figure C.5: Model performance correlated with AQ for the Identity sub-models. For each participant, the number of significant voxels in each ROI were calculated from permutation tests ($p < .05$, uncorrected) for each model (Left: Higher Identity, Right: Lower Identity) and correlated with AQ. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation as well as the partial correlation and partial p value when controlling for ROI size. We do not observe a significant correlation for Higher or Lower Identity models and AQ.

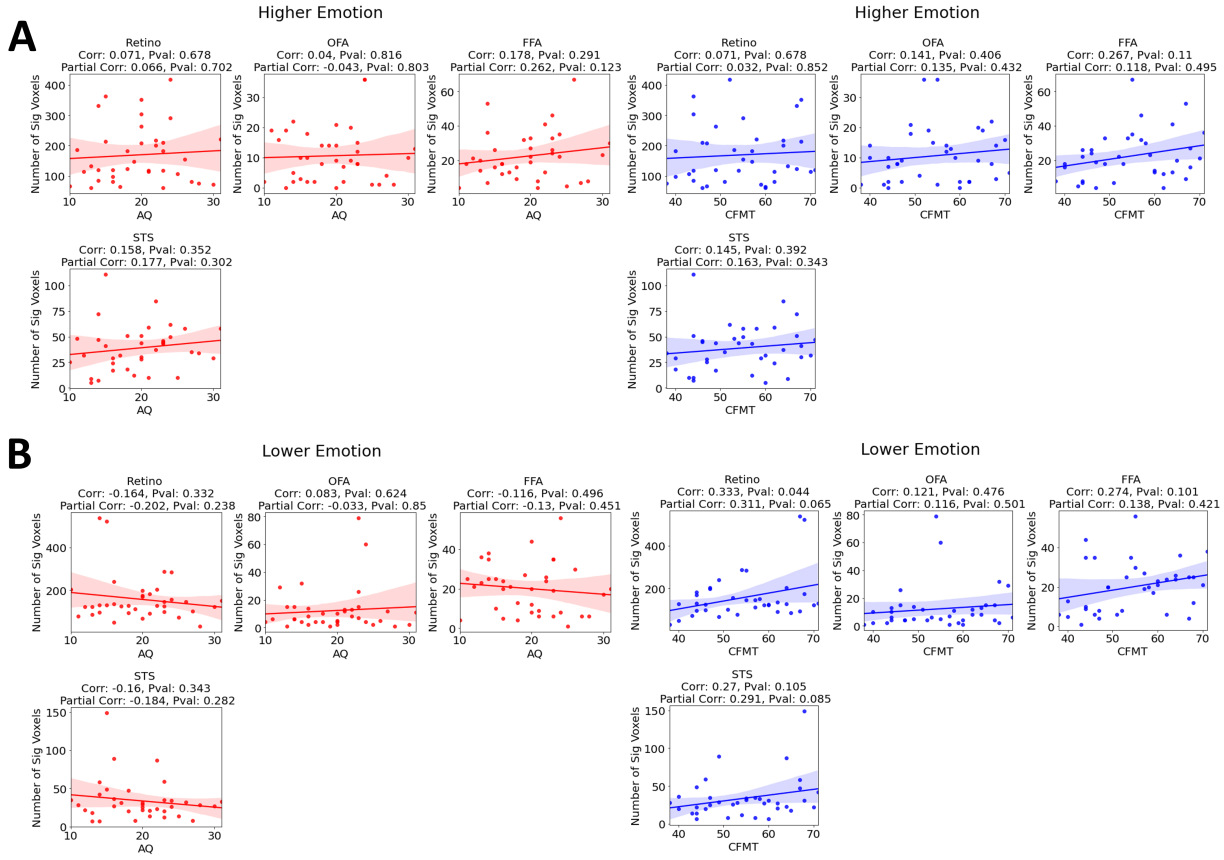


Figure C.6: Model performance correlated with AQ (red) and CFMT (blue) for the Emotion sub-models. For each participant, the number of significant voxels in each ROI were calculated from permutation tests ($p < .05$, uncorrected) for each model (top panel A: Higher Emotion, bottom panel B: Lower Emotion) and correlated with AQ and CFMT. Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation as well as the partial correlation and partial p value when controlling for ROI size.

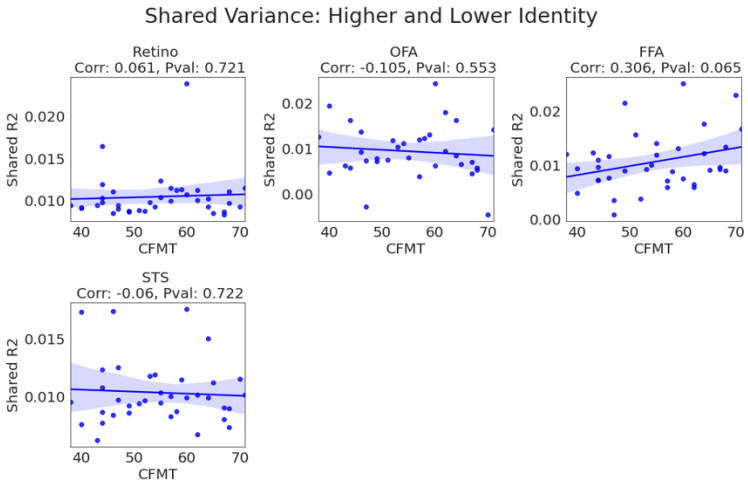


Figure C.7: The shared variance between higher identity features and lower identity features correlated with CFMT. Shared variance was calculated as the intersection of explained variance between the Lower Identity model and Higher Identity model (see Methods). Line denotes the regression line with shaded region as bootstrapped confidence interval. Each dot represents a participant. The title of each subplot shows the correlation and p value for each correlation. We do not observe any significant correlations.