UNIVERSITY OF CALIFORNIA

Santa Barbara

Long-term visual experience shapes neural representations of faces

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychological and Brain Sciences

by

Puneeth Nallan Chakravarthula

Committee in charge:

Professor Miguel Eckstein, Chair

Professor Craig Abbey

Professor Barry Giesbrecht

Professor Mary Hegarty

March 2022

The dissertation of Puneeth Nallan Chakravarthula is approved.

Barry Giesbrecht

Craig Abbey

Mary Hegarty

Miguel Eckstein, Committee Chair

December 2021

Long-term visual experience shapes neural representations of faces

by

Puneeth Nallan Chakravarthula

ACKNOWLEDGEMENTS

This manuscript represents an important milestone in my quest to better understand human object recognition mechanisms, a journey that I embarked on 9 years ago. The contents of this manuscript are a culmination of the concerted efforts of a large cast of characters that mentored me, kept me motivated, and most importantly, believed in me. I am grateful to everyone that has been a part of this journey.

Firstly, I would like to thank my mother, N. C. Sudha, for her unwavering support at every step in my academic career, despite all the uncertainties along the path. Whenever things did not go as expected, she pushed me to review what went wrong and followed up with me to make sure I did not fall prey to self-doubt or buckle under the pressure to perform. Over time, these discussions built in me the values of perseverance, open-mindedness, a strong sense of self-worth, and fortitude. If these qualities reflect in this piece of work, it is undoubtedly due to her legacy. Second, I would like to thank my father, N. C. Raghavacharyulu, for his guidance and critical but well-meaning feedback. Third, I thank my wife, Niveditha Srirambhatla, for always being there through the ups and downs of my Ph. D. and beyond.

Next, I thank my advisor Miguel Eckstein, whose mentorship pushed me to refine and clarify my ideas till they could be supported by hard research results. He let me play by my strengths but also supported me when I struggled with my weaknesses. I would also like to thank my other committee members Craig Abbey, Barry Giesbrecht, and Mary Hegarty for their feedback throughout the program.

Next, I would like to thank my lab mates Matt Peterson, Mordechai Juni, Arturo Deza, Yuliy Tsank, Katie Koehler, Aditya Jonnalagadda, Nicole Han, Devi Klein, Sudhanshu

Srivastava, Shravan Muralidaran, Byron Johnson, and Weimin Zhou for the stimulating conversations and all the good times spent in and away from the lab.

Next, I would like to thank my cohort mates Evan Layher, Diya Das, Youngki Hong, and Spencer Mermelstein. We undertook this journey together, and I could not have asked for a more fun group that made me feel at home and comfortable right off the bat.

Next, I want to thank my friends Yogesh Goyal, Neel Nadkarni, Soorya Gopal, Bharat Monga, Anant Gupta, Ekta Prashnani, and Anchal Agrawal for the comradery that kept me sane through my Ph. D.

Finally, the research in this manuscript would not be possible without the help of several motivated undergraduate research assistants that spent a lot of time conducting the experiments. While there are too many to name, if you worked on the research in this manuscript, you know who you are. Thank you for helping out!

VITA OF PUNEETH NALLAN CHAKRAVARTHULA
December 2021

EDUCATION

Degree	University	Year
Ph. D. (expected)	University of California, Santa Barbara	2021
B. Tech (Mech. Eng.)	Indian Institute of Technology Gandhinagar	2012

WORK EXPERIENCE:

1. Graduate Research/Teaching Assistant, University of California, Santa Barbara (August 2015 - present)
2. Research Assistant, Indian Institute of Science, Bangalore (August 2012 - July 2015)

TEACHING EXPERIENCE:

As a Teaching Assistant:
Psychology - 1 Introduction to Psychology (x2)
Psychology - 7, Research Methods in Psychology (x1)
Psychology 169L, Neuroanatomy Lab (x1)
Psychology 129L, Perception Lab (x4)
Psychology 120L, Advanced Research Methods Lab (x2)
Psychology 118L, Attention Lab (x1)

PUBLICATIONS:

1. Chakravarthula, P. N., Tsank, Y., & Eckstein, M. P. (2021). Eye movement strategies in face ethnicity categorization vs. face identification tasks. *Vision Research*, *186*, 59–70. https://doi.org/10.1016/j.visres.2021.05.007

2. Han, N. X., Chakravarthula, P. N., & Eckstein, M. P. (2021). Peripheral facial features guiding eye movements and reducing fixational variability. Journal of Vision, 21(8), 7–7. https://doi.org/10.1167/JOV.21.8.7

3. Puneeth, N. C., & Arun, S. P. (2016). A neural substrate for object permanence in monkey inferotemporal cortex*.* Scientific Reports, 6*.*

MANUSCRIPTS UNDER REVIEW/IN PREPARATION

1. Chakravarthula, P.N. and Eckstein, M. P. (2021). The Composite Face Effect is tuned to an individual's preferred gaze position on faces. Manuscript in preparation
2. Peterson, M.F., Tsank, Y., Chakravarthula, P.N, Or, C. and Eckstein M.P. (2021). The emergence of fixation tuned representations from repetitive eye movement programs. Manuscript in preparation

CONFERENCE PRESENTATIONS:

1. Chakravarthula, P.N., Young, A., Chow, M., Eckstein, M.P. (2021). The preferred fixation location on the face modulates the locus of the Composite Face Effect., *Vision Sciences Society*, 2021 (virtual).
2. Chakravarthula, P.N., Shrivastava, A., Eckstein, M.P (2019). Retino-specificity of the face adaptation effect at the preferred point of fixation. *Gordon Research in Eye Movements*, (2019). Lewiston, ME, United States.
3. Chakravarthula, P. N., Ghazaryan, A., & Eckstein, M. P. (2019). Looking at the preferred point of fixation mediates the composite face effect. *Journal of Vision*, 19(10), 216-216.

4. Han, X. N., Chakravarthula, P. and & Eckstein, M. (2018). Peripheral cues guiding the first eye movement to faces. *Journal of Vision, 18*(10), 233-233
5. Chakravarthula, P.N., & Eckstein, M. (2016). Eye movements during challenging cultural group identification of faces. *Journal of Vision*, 16(12), 70-70*.*
6. H. Katti, N.C. Puneeth, and S.P. Arun, *Competitive Interactions between rule and association learning in face categorization,* 824.11/JJ6, Neuroscience 2014, Washington D.C., *Society for Neuroscience*, 2014.

INVITED TALKS/PRESENTATIONS:

1. "The Cognitive Science of Parts and Wholes", Department of Physics, Harvard University, 5/11/2021.

ABSTRACT

Long-term visual experience shapes neural representations of faces

by

Puneeth Nallan Chakravarthula

Face recognition is a ubiquitous task that humans are adept at. Recent research has estimated that adult humans typically view faces for about 20% of every waking hour (Oruc et al., 2019). Although the behavioral and neural signatures of face processing have been well studied, little is known about how long-term visual experience shapes them. The incoming visual experience of faces is thought to have a changeable component (such as expression and gaze direction) and an invariant component (such as the configuration of features, see Haxby et al., 2000). One claim that is often made is that the aspects of the visual experience that are invariant influence the underlying neural computations and, consequently, the perception of faces(Diamond & Carey, 1986). However, we do not fully understand the scope of this influence and the computational mechanisms by which this influence is exerted. This thesis aims to address this knowledge gap. Specifically, we examined two well-established kinds of invariance in our visual experience of faces: one arising from the prolonged use of a consistent oculomotor strategy to view faces (Peterson & Eckstein, 2012), and the second arising from a preponderance of viewing faces in an upright configuration (Yin, 1969).

Earlier research has shown that humans land their first fixation around a consistent point on the face, known as the preferred fixation location (PFL). Human performance in various common perceptual tasks like person-identification, gender categorization, and emotion recognition are shown to be tuned to the PFL (Peterson & Eckstein, 2012). However, not known is whether other perceptual effects in faces are also tuned to the PFL. Chapter I

studied how the long-term oculomotor strategy of moving the eyes to the PFL shapes two well-studied perceptual effects: the Composite Face effect (CFE, Young, et al., 1987) and the Face Adaptation Effect (FAE, Webster, et al., 2004). For this, we compared the strengths of these effects in two groups of observers: one with a PFL close to the eyes (upper lookers) and the other with a PFL lower on the face around the nose region (lower lookers). We found that the PFL modulates both effects. On the one hand, the CFE was smaller for the lower lookers. On the other hand, the FAE was more fixation position-specific for observers whose PFL was farther away from the eyes. These findings extend the scope of the influence of visual experience on face perception to the CFE and the FAE.

Humans view upright faces disproportionately more often than inverted faces. The perceptual consequence of this predilection is the well-known face inversion effect(FIE, Yin, 1969). Tsank (2019) related the FIE to visual experience using a Convolutional Neural Network (CNN). However, the mechanism by which the tuning to upright faces as a result of visual experience to upright but not inverted faces develops is unknown. Recent research has suggested that the face inversion effect arises due to spatial summation in higher-level visual regions in the brain (Poltoratski et al., 2021). Therefore, we hypothesized that visual experience influences the spatial summation ability for upright and inverted faces. To study this, we used a spatial summation task where observers had to identify upright or faces covered by apertures of different sizes. Our results revealed that as long as the apertures are not too small, humans' spatial summation efficiency increases with aperture size for upright faces but reduces for inverted faces. We then used a convolutional neural network (CNN) that was trained on full upright or inverted faces to show that the network developed superior spatial integration abilities only for the orientation of faces it was trained on. Moreover, the

divergence in spatial integration efficiency for upright vs. inverted faces only appears when a network with a large receptive field is used, supporting the hypothesis that higher visual areas in the brain mediate the FIE.

Our results extend our understanding of the role of prolonged use of a consistent oculomotor strategy to view faces in shaping perceptual effects in faces. We also illustrate the utility of model observers in constraining hypotheses about computational mechanisms driving perceptual effects. Together, this research furthers our understanding of the unique perceptual and computational consequences of visual experience to faces.

TABLE OF CONTENTS

I. Contents

LIST OF FIGURES

I. Introduction

A. Background

Are faces *special*? Most people would perhaps answer yes, while some may ask an additional clarification question, such as "*in what way?"*. While there is no doubt that faces are *special* historically, functionally, and socially (Pernet, 2006; Rivers, 1994; Zebrowitz, 2018), many cognitive scientists have been embroiled in a debate of *how* faces are special for the past 50 years. Spurred on by modern measurement techniques fMRI, EEG, and single-cell recordings, this debate has triggered a "golden age" of face recognition research (Bruce & Young, 2013). To better situate the purpose and contributions of this thesis, it is helpful to understand the landscape of the existing literature on this topic.

What has been said about how face processing is *special* in the past fifty years? Figure 1 shows a 5-year moving window average of the number of publications that featured some common keywords used in conjunction with face processing today. The figure shows us that research on facial expression dominated the study of face perception fifty years ago and still continues to be an important research direction in face research. However, the graph also reveals that several other research terms entered the discussion of face perception in recent years, gaining popularity rapidly. For example, *domain specificity* and *expertise* have been associated with face recognition in the literature since the 70s. Likewise, the term *fusiform* started being increasingly associated with faces starting in the mid-90s. While our understanding of the mechanisms of face perceptions has evolved significantly in the past 50 years, the issue of how and why our ability to process faces is special still remains an active area of debate.

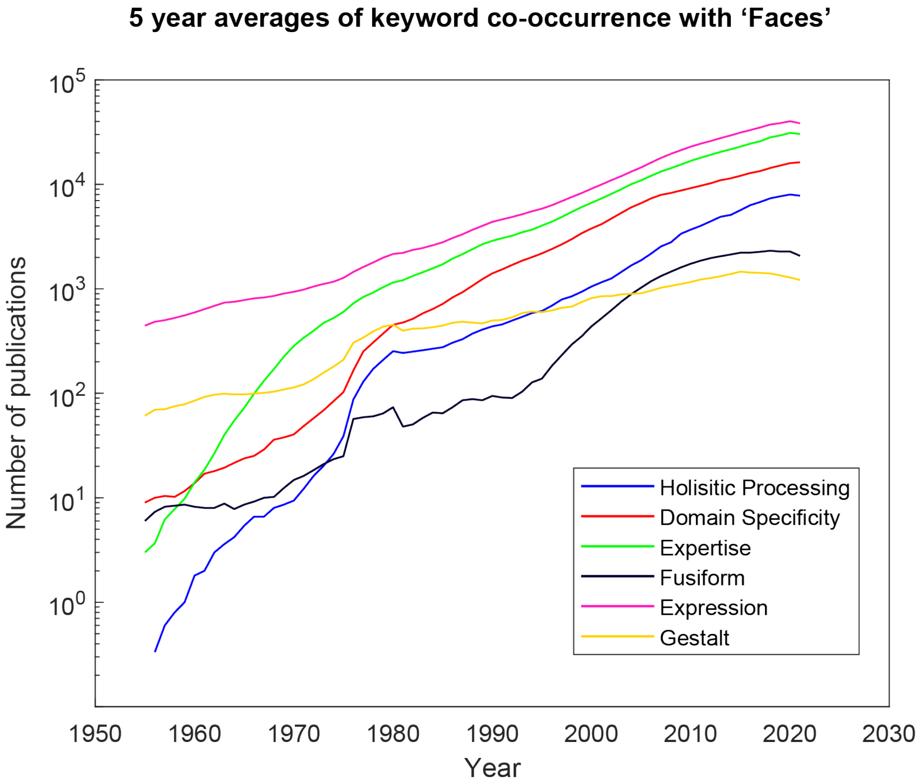


Figure 1. 5-year moving window average of the number of publications having co-occurrence of various keywords with the word 'faces' in journals in the fields of Psychology, Cognitive Science, Neuroscience, Artificial Intelligence/Computer vision journals. The chart shows that several terms have been added to the discussion in the past 50 or so years.

Taking a birds-eye view of the debate, we can place all the points that have been made into a space with the following four broad categories: (A) the holistic vs. parts-based processing, (B) domain-specificity vs. domain generality, (C) configural vs. featural coding, and (D) innate vs. learned representations. Figure 2 depicts this landscape as a mindmap. It shows each of these axes and sample references that provide empirical evidence for or against a hypothesis along the axis. Note that the references provided in this mindmap are a mix of reviews and original research papers. The intention is not to provide an exhaustive list of references on the topic of face processing but to show that each category has been explored in detail both individually and in relation to the other categories.

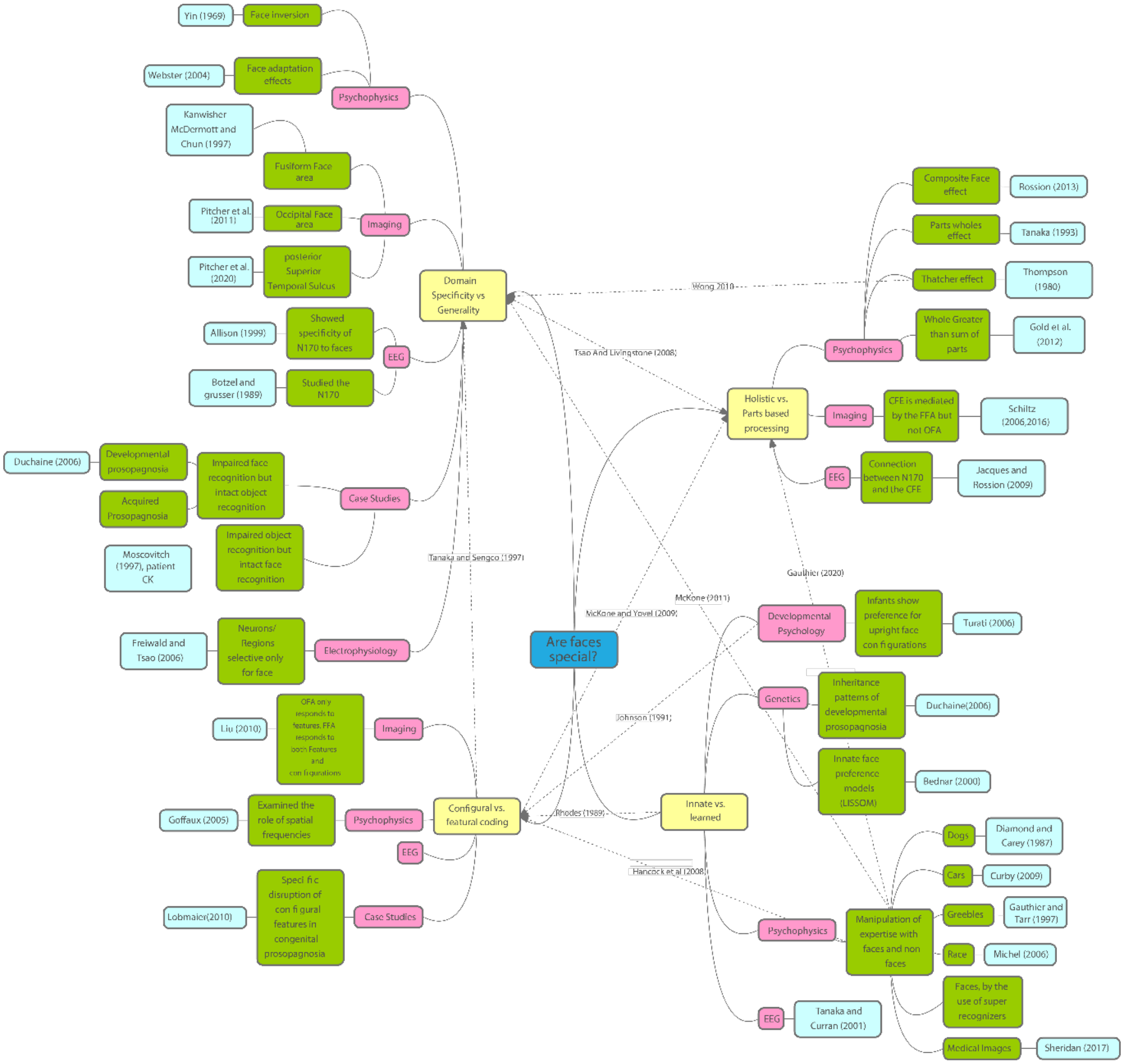

Yin (1969)
Face inversion
Webster (2004)
Face adaptation effects
Psychophysics
Kanwisher McDermott and Chun (1997)
Fusiform Face area
Pitcher et al. (2011)
Occipital Face area
Imaging
Pitcher et al. (2020)
posterior Superior Temporal Sulcus
Domain Specificity vs Generality
Allison (1999)
Showed specificity of N170 to faces
EEG
Botzel and grusser (1989)
Studied the N170
Duchaine (2006)
Developmental prosopagnosia
Acquired Prosopagnosia
Impaired face recognition but intact object recognition
Case Studies
Moscovitch (1997), patient CK
Impaired object recognition but intact face recognition
Tanaka and Sengco (1997)
Freiwald and Tsao (2006)
Neurons/ Regions selective only for face
Electrophysiology
Liu (2010)
OFA only responds to features. FFA responds to both Features and configurations
Imaging
Goffaux (2005)
Examined the role of spatial frequencies
Psychophysics
Configural vs. featural coding
EEG
Lobmaier(2010)
Specific disruption of configural features in congenital prosopagnosia
Case Studies
Are faces special?
Wong 2010
Tsao And Livingstone (2008)
Composite Face effect
Rossion (2013)
Parts wholes effect
Tanaka (1993)
Thatcher effect
Thompson (1980)
Whole Greater than sum of parts
Gold et al. (2012)
Psychophysics
Holistic vs. Parts based processing
Imaging
CFE is mediated by the FFA but not OFA
Schiltz (2006,2016)
EEG
Connection between N170 and the CFE
Jacques and Rossion (2009)
Gauthier (2020)
McKone (2011)
McKone and Yovel (2009)
Developmental Psychology
Infants show preference for upright face configurations
Turati (2006)
Genetics
Inheritance patterns of developmental prosopagnosia
Duchaine(2006)
Johnson (1991)
Innate face preference models (LISSOM)
Bednar (2000)
Rhodes (1989)
Innate vs. learned
Dogs
Diamond and Carey (1987)
Cars
Curby (2009)
Hancock et al (2008)
Greebles
Gauthier and Tarr (1997)
Psychophysics
Manipulation of expertise with faces and non faces
Race
Michel (2006)
Faces, by the use of super recognizers
EEG
Tanaka and Curran (2001)
Medical Images
Sheridan (2017)

Figure 2. A mind map depicts the landscape of the research on face processing. The four yellow boxes depict central themes that the research efforts on face processing usually align with. The pink boxes denote different techniques that have been used to collect evidence for or against a particular theory. The green boxes are short comments explaining the nature of the empirical study. The light blue boxes contain references. The references in this mind map are not exhaustive but rather provided as examples of research aiming to characterize various relationships in this space. The solid lines depict connections of ideas related to one theory, whereas the dotted lines represent connections between theories.

The holistic processing hypothesis states that faces (but not other stimuli) are processed as units, such that the perception of one part of the face influences the perception of the other part (Richler et al., 2012). In behavioral experiments, holistic processing is operationalized in one of two ways: a reduction in performance when the task is to judge a face part, or an enhancement of performance when the task is to judge a part within the whole face rather than by itself. The domain specificity hypothesis (Kanwisher, 2000) states that separate neural mechanisms are involved in upright face and object processing. These mechanisms are affected differently by stimulus manipulations, like inversion or breaking the stimulus into parts. In experiments, this is usually operationalized as a region in the brain or a neuron showing stronger neuronal activation for upright faces but not for inverted faces or other objects. The opposing domain generality hypothesis states that the same neural mechanisms process faces and objects. The concept of configural coding refers to the priority given to distances between features (like eyes, nose, and mouth) rather than their shape (Diamond & Carey, 1986). It is hypothesized that faces rather than other objects show configural coding. Finally, the innate vs. learned representations hypothesis states that infants have an innate preference for face-like configurations but not for inverted faces or other configurations (Morton & Johnson, 1991). The opposing theory is that face processing is acquired through visual experience.

B. Contributions of this thesis

Despite the large volume of research on face recognition, we still do not have a complete picture of face recognition mechanisms in humans. Collating ideas from various reviews from scientists that have studied this problem for a long time, we have identified three main challenges. The first one is a technical challenge. Significant advances in our understanding of face recognition have come in the early 2000s due to the introduction of fMRI to study face perception (Kanwisher, 2017). However, several researchers have noted the limitations of fMRI in resolving the activities of small subpopulations of neurons mediating specific face recognition functions (Dubois et al., 2015). As Miller describes in his reflective piece, progress in cognitive science often depends on discoveries made in other fields (Miller, 2003). This highlights the importance of bringing in state-of-the-art tools from various disciplines to aid research on face recognition. The second factor is a semantic one. There has been a lack of consensus on various meanings and measures of perceptual effects in face recognition (Richler et al., 2012). This has resulted in elaborate debates on methodological factors (Rossion, 2008), the validity of measurements (Richler & Gauthier, 2014), disagreements on the exact wording of working definitions (Piepers & Robbins, 2012), and ultimately divergent results from different paradigms trying to measure the same effect (Rezlescu et al., 2017). Such disagreements can be avoided by the use of computational models (Heinke & Mavritsaki, 2009). The use of computational modeling forces us to create testable theories that can be proven or disproven based on the outcomes of simulations (Guest & Martin, 2021). Finally, researchers have acknowledged the lack of a reliable way to manipulate the visual experience of faces (Gauthier, 2020). While there is a consensus that visual experience shapes visual object perception, considerable challenges (Op de Beeck & Baker, 2010), we don’t understand how it does so due to the dearth of studies that systematically manipulate the visual experience of faces (Gauthier et al., 2010). Together, these three challenges have significantly impeded progress in the field of face recognition.

A novel technique to manipulate visual experience to faces has emerged recently due to progress in understanding oculomotor strategies to faces. The human visual system is foveated with a high fidelity region at the center and a progressively lower resolution with increasing eccentricity (Curcio et al., 1987). It was shown that human eye movements to faces land at a consistent point on the face called the preferred fixation location (PFL). Fixation at this location modulates performance in many face tasks like person-identification, gender categorization, and emotion recognition (Peterson & Eckstein, 2012). Additionally, stable individual differences in the PFL (Peterson & Eckstein, 2013) that extend to the real world (Peterson et al., 2016) have also been demonstrated. More recently, Tsank (2019) showed using a Convolutional neural network (CNN) trained on faces foveated at a specific location on the face develops tuning to that position, similar to humans. This suggests that visual experience mediated by long-term consistent oculomotor strategies shapes the tuning of human face recognition to the PFL. Taken together, these findings provide the basis for the premise that visual experience to faces can be manipulated by varying the position of the PFL, provided the faces are large enough such that varying the fixation location on the face leads to sufficiently different visual input. Oruc et al. (2019) found that the typical face size encountered in natural settings is 6-10° wide. This size is comparable to the size of faces used in Tsank (2019), increasing the validity of the claim that the PFL is a proxy for manipulating the visual experience of faces. This thesis extends the influence of the PFL to other well-studied face tasks like the Composite Face Effect (CFE) and the Face Adaptation Aftereffect (FAE).

Additionally, we demonstrated the role of spatial integration in the face inversion effect. For this, we calculated the efficiency of spatial integration for upright and inverted faces as viewed through apertures of various sizes. Our experimental design is distinct from earlier research comparing efficiencies in processing upright and inverted faces. For example, Yang et al. (2014) calculated the efficiency of full upright and inverted faces with varying overall face sizes. Gold et al. (2012) computed the relative efficiency of the whole face compared to face parts. Our experiment design was inspired by earlier research demonstrating spatial summation in an upright vs. inverted face detection task (Tyler & Chen, 2006). However, we used a face identification task instead. Further, we also trained CNNs on this task and compared their spatial integration efficiency for each aperture size to that in humans. Using different aperture sizes, we were able to control the spatial extent of the task-relevant information, which allowed us to assess the role of increasing receptive field sizes on the face inversion effect.

C. Significance

As discussed earlier, one significant challenge to progress in face recognition is to find a reliable way to experimentally manipulate the visual experience of faces. The gold-standard experiment one can do to investigate this would involve a regulated rearing study with strict control of the amount of visual experience to faces. Such studies are impossible in humans and impractical in animals. In the absence of such evidence, researchers have turned to indirect ways to manipulate visual experience to faces: to study within-subject variability for familiar vs. unfamiliar faces or to study a stimulus set for which there exist individual differences in visual experience. The former has been advocated for by Burton (2013). While human face recognition performance for faces is superior for familiar than unfamiliar faces, there is a saturation effect for familiar faces. The interstimulus variability of experience for familiar faces does not modulate measurable factors enough to measure correlations(Megreya & Burton, 2006). While on the one hand, some researchers have tried to use sets of novel objects where they can train participants on one set and leave the other set unfamiliar (Gauthier & Tarr, 1997), on the other hand, some others have pointed to the stimuli being too different or too similar to faces (Kanwisher, 2000). This arbitrary distinction prevents us from answering questions like "*is processing is domain-specific*?" with conviction.

The other method is to look for individual differences in the visual experience of faces. Many researchers have used individual differences research to inform theory in cognitive science (Yovel et al., 2014). Perhaps the most striking example of this is the Other Race Effect (ORE). Early research on the ORE suggested that holistic processing is exclusive to own-race faces (Michel, Rossion, et al., 2006). However, more recent research has found the opposite. (Harrison et al., 2014; Horry et al., 2015). Gauthier (2020) suggests that race manipulation may be too weak, with the average population already being experts at all kinds of faces. Thus, manipulating the race may not affect measures like holistic processing much. Given these challenges, a reliable way of manipulating the visual experience of faces is the need of the hour.

We propose a novel method to quasi-experimentally manipulate the visual experience of faces based on a recently-established measure that shows reliable individual differences: the landing point of the first saccade to faces (Peterson & Eckstein, 2013; Peterson et al., 2016). Our research dovetails with the recent research showing fixation-specific tuning for faces in both neural responses (Issa & DiCarlo, 2012; Stacchi et al., 2019) and behavior (Tsank, 2019). Importantly, this thesis aims to relate this body of research to the mainstream conversation of the neural and computational basis of face recognition in humans.

We also demonstrate the utility of the cutting-edge tool of deep learning to aid arguments and constrain the computational theories of face recognition. Our use of CNNs for face recognition is by no means novel. At the time of this writing, several sophisticated and powerful deep learning architectures exist that are capable of detection (Sun et al., 2018), alignment (Kowalski et al., 2017), and recognition (Ghazi & Ekenel, 2016) exist. However, the goal of these networks is to achieve or surpass human face recognition abilities in a real-world setting. By contrast, we use small and simple CNNs to test hypotheses arising from a theoretical framework computationally. Following the deep learning boom in the last decade, its value has only recently been recognized in forcing theory-building in psychological science (Guest & Martin, 2021). To our best knowledge, this is the first time deep learning has been used in conjunction with spatial integration efficiency measurements to test theories of holistic processing.

D. Organization of the thesis

Chapter II demonstrates a novel psychophysical paradigm that exploits inter-subject variability in oculomotor strategies to manipulate the PFL in lieu of visual experience. This experimental design is based on the premise that humans have a foveated visual system where visual information gradually decays in quality with eccentricity. Therefore, in the long term, different eye movement strategies on faces could ultimately culminate differences in the visual experiences of faces. While this is not guaranteed, several recent results strengthen this hypothesis (see section B). These are discussed in chapter II. The role of visual experience as mediated by the PFL has been established for face identification and gender recognition but not for other face tasks. We used modified versions of two classic paradigms to demonstrate the efficacy of this paradigm: the composite face effect (CFE) task and the face adaptation effect (FAE) task.

Chapter III discusses another approach to establish the role of long-term visual experience in the face inversion effect: by comparing the efficiency of humans with that of model observers on a spatial summation task that used upright and inverted faces covered by an aperture of varying sizes. We used an adaptive staircase procedure for both agents to estimate the contrast thresholds for each condition. To calculate the efficiency, we compared the performance of a human or the model observer to that of an ideal observer. This method has been used often in vision science to benchmark an observer's performance relative to the theoretical optimum performance given the underlying statistics of the stimulus (Geisler, 2011). The model observer was a Convolutional Neural Network, which is a deep learning algorithm. The last decade has seen a massive boom in deep learning algorithms (Alom et al., 2018). Deep learning algorithms are a class of machine learning algorithms that can categorize visual stimuli by learning the statistical regularities underlying images belonging to different categories. Here, we use them to test whether holistic processing for upright rather than inverted faces can be explained by more efficient use of information distributed spatially across the extent of upright faces. Bearing in mind various limitations in treating CNNs as a model for the visual system (Borowski et al., 2019; Lindsay, 2021), we offer possible mechanisms by which visual experience may shape the perception of upright and inverted faces.

II. The prolonged visual experience is a common factor modulating holistic processing and face identification

A. Abstract.

Recent research has shown that humans face processing is tuned to a unique preferred fixation location (PFL) on the face. Across various face tasks like face recognition, gender recognition, and emotion identification, humans perform best when they fixate their PFL, and this PFL varies moderately across subjects. However, not known is the role of the PFL in other face tasks. Here we examine how the Composite face effect (CFE) and the Face adaptation aftereffect (FAE) are modulated by the PFL. Experiment 1 compared the strength of the CFE for two groups of observers that were screened to have their preferred first fixation location either close to the eyes or closer to the nose tip. The premise for this manipulation was that consistent differences in the visual experience of faces over a prolonged period would shape neural circuits responsible for face processing differently in these two groups. We found that the location of the PFL modulated the magnitude of the CFE such that observers whose PFL was lower on the face had a weaker CFE. In experiment 2, we again compared two groups of observers screened similarly as in Experiment 1 with PFLs either closer to the eyes or to the noise tip on a gaze-contingent Face adaptation aftereffect (FAE) task. Observers adapted to a face while fixating either their own group's mean PFL or of the other group. The test stimulus, a morph between the adapter and another face, flashed either at the same location or the alternate fixation location. We measured the strength of the adaptation aftereffect in each condition. We found that the location of the PFL modulated the FAE such that observers whose PFL was higher on the face showed a higher reduction in the FAE when fixating farther

away from their PFL. Taken together, the results from these two experiments suggest that prolonged visual experience to faces through the use of a consistent oculomotor strategy of landing the first fixation on a face at the PFL modulates both the CFE and the FAE.

B. Introduction

A quick look at a face allows us to glean a lot of socially relevant information like identity, gender, mood, attractiveness, and even trustworthiness. In the past half-century, there has been a lot of interest in understanding the mechanisms by which humans are able to achieve this feat (Bruce & Young, 2013). This interest in the mechanism of face processing has led to the discovery of many peculiar perceptual effects that occur in face judgments on stimuli that have been manipulated or doctored in ways that are uncommon or non-existent in nature. Notable examples of such effects are the face inversion effect (Yin, 1969), the Thatcher illusion (Thompson, 1980), the composite face effect (Young et al., 1987), the parts vs. wholes effect (Tanaka & Farah, 1993), and the face adaptation after effect (Webster et al., 2004). While these effects are not ecologically valid in the sense that they don’t have a functional role in real-life face recognition settings, their study has been central in our theoretical understanding of how humans process faces. Therefore, these visual tasks are an important testbed for the study of face recognition.

Recently, it was shown that across a variety of common face tasks like person, gender, and emotion recognition, humans consistently land their first saccade at a below the eyes (Peterson & Eckstein, 2012). This location, hereby referred to as the Preferred fixation location (PFL), varies moderately across individuals (Peterson & Eckstein, 2013), and these variations generalize to the real world (Peterson et al., 2016). This point of fixation plays a functional role in the aforementioned tasks, such that when observers fixate away from the PFL, there is a reduction in the task performance. This variation has been explained in terms of an interaction between the foveated nature of the visual system and the distribution of task-relevant features (Peterson & Eckstein, 2012). More recently, it was shown that the internal representations of faces in humans are tuned to the PFL in these tasks (Tsank, 2019). These observations are consistent with research showing that human face processing is optimized for the *face diet* that we encounter during the course of our life (Oruc et al., 2019; Yang et al., 2014). This is because the act of consistently moving one's eyes to a preferred location on the face can bias their internal representations to stimulus features accessible when fixating at the PFL, given peripheral information loss. While this body of research has largely focused on common face tasks, we do not understand how the PFL modulates perceptual effects in tasks that are doctored to produce perceptual effects. As noted earlier, such research has the potential to elucidate constraints and mechanisms by which the long-term oculomotor strategy of consistently moving one's eyes to a PFL on the face can shape face perception. In this chapter, we consider how the PFL modulates two popular face effects: the composite face effect (CFE), and the face adaptation aftereffect (FAE).

The discovery of the CFE by Young et al. (1987) is an important landmark in the evolution of an influential hypothesis that upright faces are processed as units (Piepers & Robbins, 2012). In the CFE, the top (or bottom) halves of two faces, although identical, are perceived as being different when the bottom (or top) halves of the faces are different (Young et al., 1987). Over the years, the CFE has been studied extensively (see Murphy et al., 2017; Rossion, 2013, for a review). Figure 3 shows the general stimulus design for testing the CFE. Note that the reader may not experience the CFE equally strongly for both halves.

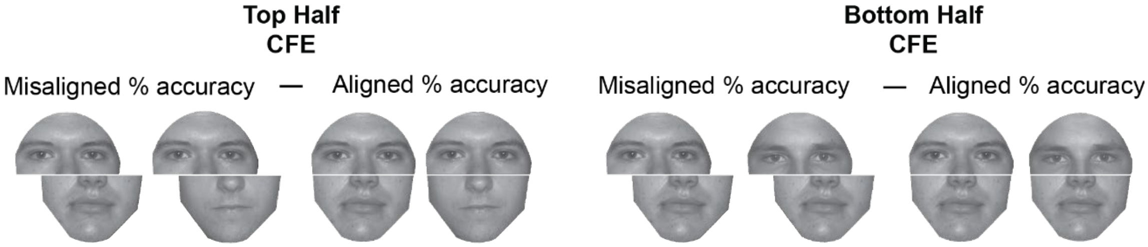

Top Half
CFE
Misaligned % accuracy — Aligned % accuracy
Bottom Half
CFE
Misaligned % accuracy — Aligned % accuracy

Figure 3. The top and bottom half CFEs. In a sequential face-half matching task, the irrelevant half affects the performance of the relevant half more when the faces are aligned as compared to when they are misaligned.

At its core, the CFE represents an inflexibility of the visual system (presumed to be expert at face processing) to discount an irrelevant part of a face, while it does comparatively better for other non-face objects (Cassia et al., 2009; Robbins & McKone, 2007). Typically, this effect is attributed to *holistic* face-processing, i.e., a mode of computation that considers all the parts at once(Richler et al., 2012). There are several other examples of tasks where observers also show such inflexibility in adjusting for unfamiliar manipulations to features or viewing conditions. These include the inability to recognize inverted faces (Yin, 1969), diminished recognition performance with other-races (Cross, Cross, & Daly, 1971), atypical illumination directions (Braje, Kersten, Tarr, & Troje, 1998), diminished recognition for too small or too large faces (Yang, Shafai, & Oruc, 2014), and diminished recognition when fixating away from the preferred first fixation location (Peterson & Eckstein, 2012). Thus, the CFE can be interpreted as an inability to flexibly use learned internal representations of full faces on a task that demands judgments based on partial use of incoming face information. In this general framework, we examine the role of long-term learned representations of faces as mediated by the PFL on the CFE.

The FAE is related to a well-known property of the visual system: visual adaptation (Webster, 2015). The FAE has been used extensively to characterize the relationship between the physical features of faces and the internal representations of faces(Webster et al., 2004). The FAE is demonstrated by making an observer fixate a face for a certain period (typically 500 msec – 4000 msec) and then presenting an ambiguous test face that looks midway between the person that saw earlier (say *face A*) or a different face (say *face B*). In the context of having seen face A, observers tend to report that the ambiguous face as *face B* more often. Thus, even though the task is a standard face recognition task, the FAE captures a temporal dependence of the current perception on a previously viewed face. This temporal influence is usually attributed to neuronal fatigue arising from strong activation of neurons coding for face A following prolonged exposure, resulting in a weaker representation of *face A* than *face B* in the ambiguous face. Given the close relationship of the FAE with acquired internal representations, we picked this effect to investigate the role of visual experience mediated by the PFL in face processing.

For this, we implemented gaze-contingent versions of the classic CFE and FAE tasks, where we quasi-experimentally manipulated the PFL on the face. We screened observers whose PFL was either high up on the face close to the eyes or lower on the face close to the nose tip. We hypothesized that if the PFL modulates these effects, there would be a main effect or an interaction effect of the observer group on the strengths of these effects.

C. Experiment 1: The Composite face effects task

1. Participants*.*

a) Prescreening Task.

A total of 126 observers participated in a short free eye movement face identification screening task. These participants were students at the University of California, Santa Barbara, and participated in the study either for course credit or for a small monetary reward. After consent for participation in the study and for future contact were obtained, each participant was given instructions on how to perform the task. The experimental paradigm was identical to the standard free eye movements face identification task described in the Experiment Design section, except that it was a 1-in-5 face identification task using an in-house dataset of Caucasian faces instead of the composite faces from figure 3. The actual stimuli used are shown in figure 4a. The dimensions and alignment of the Caucasian faces were matched to the composite faces. We measured the mean first fixation location on the face across trials when initiating a saccade from a peripheral location.

b) Composite Face Tasks.

To select upper-lookers and lower-lookers, we invited 15% of the participants of the prescreening experiment from the top and bottom ends of the distribution (about 18 participants from each group) for the full study. Three upper lookers and one lower looker were unavailable to continue with the study. Thus, we had 15 upper lookers and 17 lower lookers. The upper looker group consisted of 14 females and one male with ages ranging from 18-24. The lower looker group consisted of 12 females and six males with ages ranging from 19-25. All participants were students at the University of California, Santa Barbara, and participated in the experiment in exchange for an hourly monetary compensation for a total of 7-10 hours. All participants had a normal or corrected-to-normal vision.

2. Stimuli.

a) Prescreening task.

We used a set of 5 Caucasian male frontally photographed faces for the prescreening task. These faces were a part of an in-house dataset. These faces were standardized by rotating, cropping, and resizing to align the eyes and the chin across the stimuli. We also matched the luminance and contrast energy to ensure that low-level features don't drive eye movements. The faces were about 12.6° x 9.2° in size. The images were presented in full original contrast (see Figure 4a).

b) Composite Face tasks.

We selected two Caucasian male faces photographed frontally for generating composite faces. These faces were part of an in-house dataset of faces. These faces were first standardized by rotating, cropping, resizing them such that the eyes and chin were centered and aligned. These standardized faces were then converted to an 8-bit grayscale format and embedded in a mask that only revealed frontal facial features. Both the luminance and contrast energy were matched so that variations in skin color or texture cannot be used to judge the identity. The faces were then split into two halves along the vertical dimension, and these halves were placed at a gap of 0.11°, as shown in Figure 4b. Different combinations of the top and bottom halves of these faces were assembled to create composite faces. They will be referred to as AA, AB, BA, and BB, with the first letter denoting the top half and the second letter denoting the bottom half (see figure 4a). The faces were about 12.1° x 9.9° in size (see Figure 4b).

Another corresponding set of four misaligned faces was formed by displacing the bottom half to the right by 2.4°. The stimuli were presented at 30% of the original contrast to avoid ceiling effects in task performance. These settings are based on piloting to have an average identification performance of around 80%. Various other measurements between features in the face are shown in figure 4b.

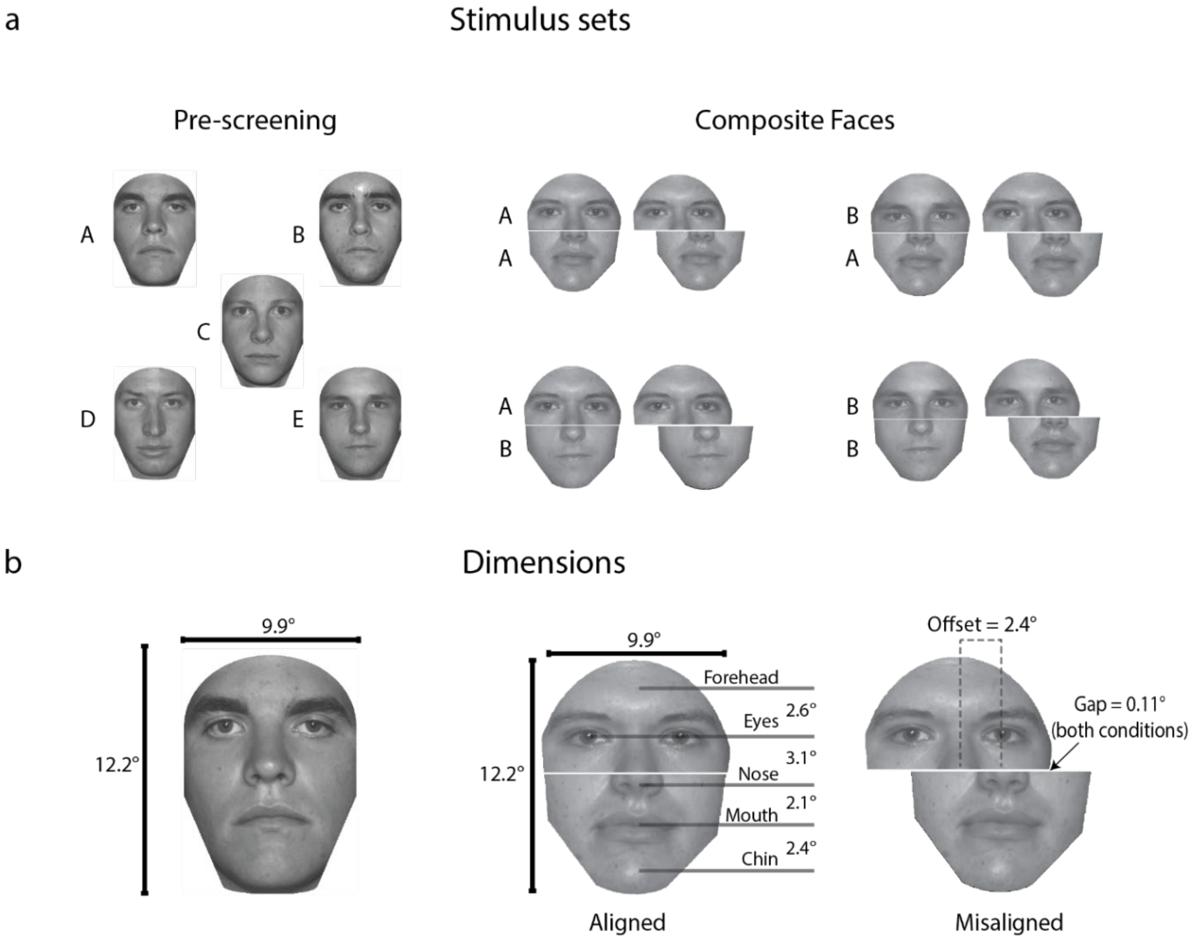


Figure 4. (a) The left panel shows the five faces used in the 1-in-5 free fixation face-matching task used for pre-screening. The right panel shows the four composite faces used in the main study (represented by AA, AB, BA, and BB). (b) Stimulus dimensions for the stimuli used in the prescreening task (left panel) and the composite face tasks

3. Apparatus

The stimuli were presented on a Barco monitor with a resolution of 1280 x 1024 and a refresh rate of 60 Hz. The monitor was calibrated linearly with a maximum luminance of 114.7 cd/m^2. The screen was placed at 75 cm from the participant's eyes, such that each pixel on the monitor subtended a visual angle of 0.021° on their eyes. The stimulus display was controlled by software written using PsychToolbox-3 (Brainard, 1997) running on MATLAB 2018.

An Eye Link 1000 Plus desktop portable eye tracker was used for tracking the left eye of each participant. The sampling rate of the tracker was 250 Hz. A 9-point calibration procedure was used at the beginning and repeated periodically to ensure accurate gaze data recording. Standard algorithms from the EyeLink toolbox were used to identify saccade and fixation events from the gaze data.

4. Experiment Design.

We present data from 2 kinds of experimental paradigms in this chapter: the free fixation face identification task and the enforced fixation sequential face-part matching task. The basic trial design for these two paradigms was the same across different experimental conditions.

a) Free fixation face identification task.

This task was used to measure an observer's preferred first fixation location on a face during face identification. To get an accurate measurement of the preferred first fixation location, each observer performed a total of 320 trials split across four blocks of 80 trials each. Within each block, the initial peripheral fixation location was varied across eight different peripheral locations around the screen. All participants completed the free fixation face identification task before we started running them on the enforced fixation task, which is described next.

b) *Enforced-fixation sequential face part matching task.*

This task was used to measure the composite face effect (CFE) for the top half and the bottom half of faces while observers maintained their gaze at a specific location on the face (which was manipulated). We measured the CFE by comparing the performance on the matching task in the misaligned condition with that in the aligned condition. Each observer completed 2 (top half/bottom half) x 2 (aligned/misaligned) x 2(fixation positions) x 256(repeats) = 2048 trials. The trials were distributed into 16 blocks with separate blocks for top half/bottom half judgments and aligned/misaligned conditions. Thus, there were four unique blocks, and each block was repeated four times. Within each block, there were 128 trials, and observer fixation was enforced in two different locations with equal probability. This task was completed in several sessions, usually over 7-10 days.

Within each block, half the trials were *same* trials, and the remaining half were *different* trials. Figure 4b shows the design of composite faces for these two trial types. Note that the stimuli for *same* and *different* trials depend on whether the judgment is of the top half or of the bottom half.

5. Trial design

a) *Free fixation face identification.*

In this task, the initial fixation-cross appeared in one of 8 possible peripheral fixation locations (either 25° or 19° away from the center of the screen). Participants were instructed to maintain their gaze on the fixation location and hit a key to indicate readiness. After the keypress, the program checked on the fly whether the participant maintained their fixation on the cross for a variable period of 500 msec. – 1500 msec. If the observer's fixation drifted more than 1° from the center of the fixation cross, the trial was aborted and restarted. If the participant successfully maintained fixation for the variable delay period, the cross disappeared, and a noisy contrast-reduced face appeared at the center of the screen and stayed on for 1500 msec. This face was chosen randomly from the set of faces used for the experiment. During this period, participants could freely move their eyes and examine the face. After 1500 msec., the face disappeared and was replaced by a Gaussian white noise mask of matched mean luminance and a standard deviation of ~6.4 cd/m^2 for 500 msec. This was followed by a response screen containing all faces from the dataset so the participant could select the face that was presented. The screen stayed on until the observer indicated which face they saw through a mouse click. After the response, the experiment progressed to the next trial. No feedback was given. See figure 5a for a schematic of the events within a trial.

b) Enforced-fixation sequential face part matching task.

In this task, the initial fixation-cross appeared at two possible locations along the centerline of the face. The locations were chosen to be the average PFLs of the two groups (upper and lower lookers). Like in the free fixation face identification condition, observers were instructed to maintain their fixation on the initial fixation location and hit a key to indicate readiness at the beginning of the trial. After the keypress, the program verified if the observer maintained their fixation on the cross for a variable delay period of 500 msec to 1500 msec. The trial was aborted if there was an instance where the gaze drifted more than 1° from the center of the fixation cross. If the observer successfully maintained fixation on the cross through the delay period, a contrast reduced composite face embedded in white noise with a standard deviation of ~6.3 cd/m^2 was presented for 200 msec. Note that we added white Gaussian noise to the images during the presentation for the purposes of modeling, the results of which are not discussed in this chapter. After the presentation of the first face, a white Gaussian noise mask with mean luminance matched to the face and standard deviation of ~6.3 cd/m^2 was presented for 500 msec. This was followed by the second noisy, contrast-reduced composite face (chosen based on the current trial) for 200 msec. After this, a response prompt was shown that asked the observer to indicate with a keypress whether the half of the face being tested is the same or different between the two faces. This screen stayed on until a response was made. After the response, no feedback was given, and the next trial was initiated. In half of the blocks, faces were misaligned. After the initial delay period, the fixation-cross persisted throughout the trial (overlaid on the faces and mask) but with reduced contrast, helping the observer maintain fixation. The observer was instructed to maintain their gaze on the fixation cross throughout the trial after indicating readiness with a keypress. If their gaze drifted away by more than 1° from the fixation cross at any point during the trial, the trial was aborted and repeated. Figure 5b shows the trial schematic for this task.

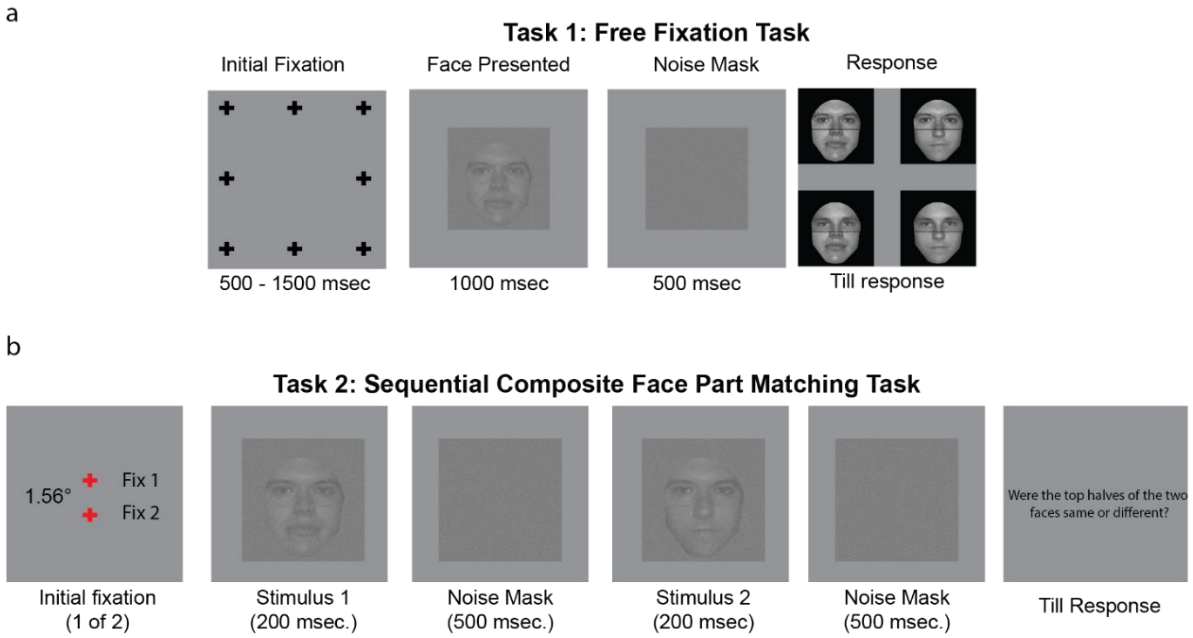


Figure 5. This figure shows the schematics of the two tasks. (a) depicts the free fixation face identification task. Observers initiated the trial fixating one of the eight possible peripheral locations. A face is then presented in the center of the screen, and observers can freely move their eyes while studying the face. In the next screen, they are required to indicate which face was shown using a mouse click. (b) depicts the enforced fixation sequential face part matching task. On each trial, observers initiate the trial at one of the two possible fixation locations that differed by 1.56°. Then two faces are flashed briefly, separated by a noise mask to wash out lingering percepts. While the faces are flashed, the gaze position observers were prevented from drifting their gaze from the fixation location by more than 1°. After viewing the two faces, observers were required to respond to a question asking them to match a given half of the face (depending on the block).

6. Procedure.

The experiments were administered by trained graduate or undergraduate researchers in accordance with protocols approved by the Institutional Review Board (IRB) of The University of California, Santa Barbara. Participants were first briefed about the nature of the study and compensation agreements. After obtaining consent to be a part of the study, they were given instructions about the task. The main experiment consisted of 2 tasks: the free fixation face identification task and the enforced fixation sequential part matching task. The two tasks were always performed in the same order, i.e., the free fixation task followed by the enforced fixation task. This was to verify that the participant had a consistent PFL for the composite faces that matched their PFL as measured in the prescreening task and to make the participants familiar with the four composite faces used in this experiment. Note that these tasks were conducted in parallel – as soon as a participant finished the free fixation task, they could start with the forced fixation task, and this timeline was separate for each participant. Each task was further divided into blocks that took 15-20 minutes to complete. Participants were encouraged to take breaks between sessions and were not allowed to spend more than 1.5 hours per session to avoid the effects of fatigue. We recalibrated the eye tracker between blocks and whenever a participant took a break to maintain eye-tracking quality throughout each session.

7. Analysis

a) Preferred fixation location.

The preferred first fixation location is defined as the first location inside the face that an observer's foveal region lands on when they make an eye movement to a face from a peripheral fixation location. For each observer, following the completion of all blocks of the free fixation task, the preferred first fixation location was estimated as the mean fixation location across all the first fixations on the face across trials. The vertical coordinates of the first fixation location on the face were used to analyze the dependence of CFE on the gaze position.

b) Power Analysis.

The choice of the number of subjects per group was made based on apriori power analysis done using data from a pilot experiment where we measured the effect of manipulating the enforced fixation position by 5.4° on the CFE. Effect sizes of 1.19 and 0.05 were obtained for the top and bottom-half CFE, respectively. It was thus reasonable for us to focus our power analysis on the top-half CFE. For a $(1 - \beta)$ rate of 0.95, we would require a sample size of 12 (assuming a fixation manipulation of 5.4°). We also conducted a power analysis to check how many participants would need to be screened to have a reasonable chance of finding two groups of participants that have the required distance between the mean vertical coordinates of the PFLs of each group. For this, we used an in-house database of PFLs measured on 186 participants. The PFLs were normally distributed, as evidenced by a Kolmogorov-Smirnov test ($p = 0.09$). The estimated mean and standard deviation were 7.27 and 0.85, respectively. For different selected values of the mean difference between PFLs, we simulated samples of mean PFLs for different sample sizes drawn from the fit normal distribution 10,000 times and counted the fraction of cases where the difference between the top and bottom 20 participants in the distribution of PFLs was at least equal to the selected mean difference. This analysis gave us the probability of finding two groups of 20 participants each as a function of the required distance between PFLs and the number of participants to be prescreened (see Figure 6). We estimated that ~120 participants would have to be screened for the mutual distance of the PFLs of the groups to be 2.1°.

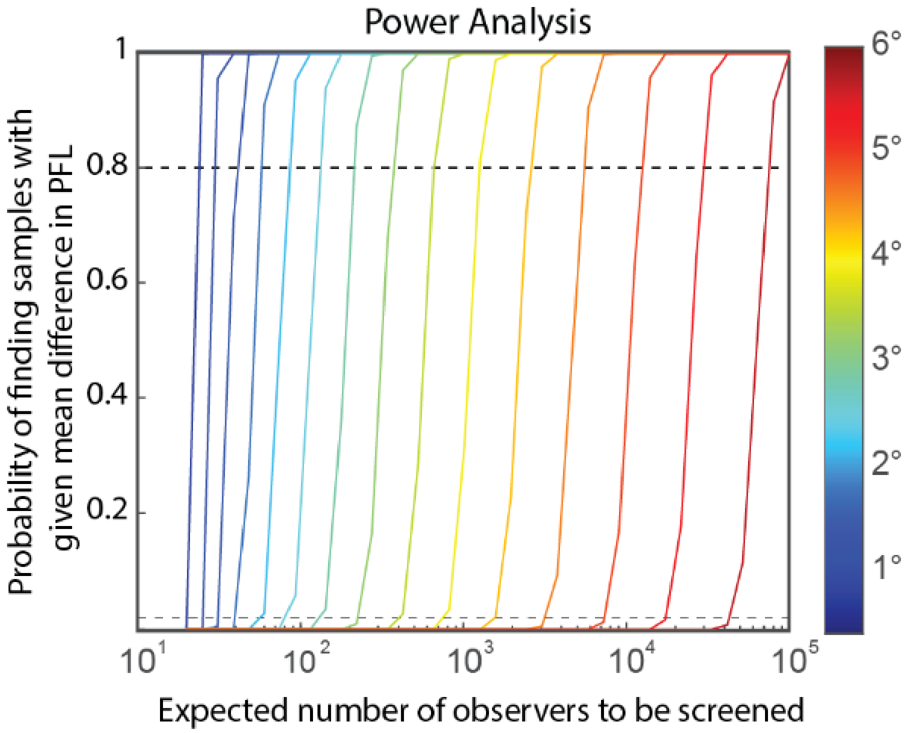


Figure 6. This plot shows the results of the power analysis conducted using an in-house database of PFLs to estimate the number of participants required to be screened to find two groups of observers that differ in their mean vertical coordinate of their PFLs by a given distance. The X-axis shows the expected number of observers to be screened. The Y-axis shows the probability of finding samples with the mean difference indicated by the colormap shown to the right of the plot. The upper and lower dotted lines represent 80% and 1% chance, respectively. The chart suggests that screening, we could expect to find groups with their PFLs separated by about 2° with >80% chance if we screened about 120 participants. (b) The panel shows the actual distribution of the vertical coordinates of the PFLs obtained from 126 screening participants. We selected the top and bottom 15% of the participants and invited them for further experiments. The upper lookers are depicted in green, and the lower lookers are depicted in pink. Those that were unavailable or not selected (due to poor quality data) are depicted in brown.

c) *Quantifying the CFE.*

The *strength of the CFE* was measured as the difference in match task performance between the misaligned and the aligned version of the composite faces for a given condition. Thus, the strength of the CFE will be reported in % points. For example, if the strength of the CFE is 10%, that means the performance of the participant in the misaligned condition was 10% higher than that in the aligned condition. While this is the classic method to test the CFE, some recent studies (Richler et al., 2008) have shown that a signal-detection approach that yields sensitivity (*d'*) and bias (λ) metrics by considering the hits and false alarms in the reporting a match on *same* and *different* trials is sometimes superior. We repeated our analyses using these metrics and found no qualitative differences.

8. Results

a) *Prescreening Task.*

The distribution of the vertical coordinate of the PFLs of the 126 observers in the prescreening task is shown in figure 7. A Kolmogorov-Smirnov test revealed that these values were normally distributed (p= 0.1). It was slightly skewed towards locations lower on the face (sample skewness = 0.7). The upper and lower lookers were selected from the 15-percentile tails of the distribution shown in purple and green, respectively. Thus, we selected about 18 participants from each group. This was done to maximize the chances of observing a significant effect of manipulating the PFL across groups. Some of the selected participants did not continue with the study, resulting in a final tally of 15 upper lookers and 17 lower lookers.

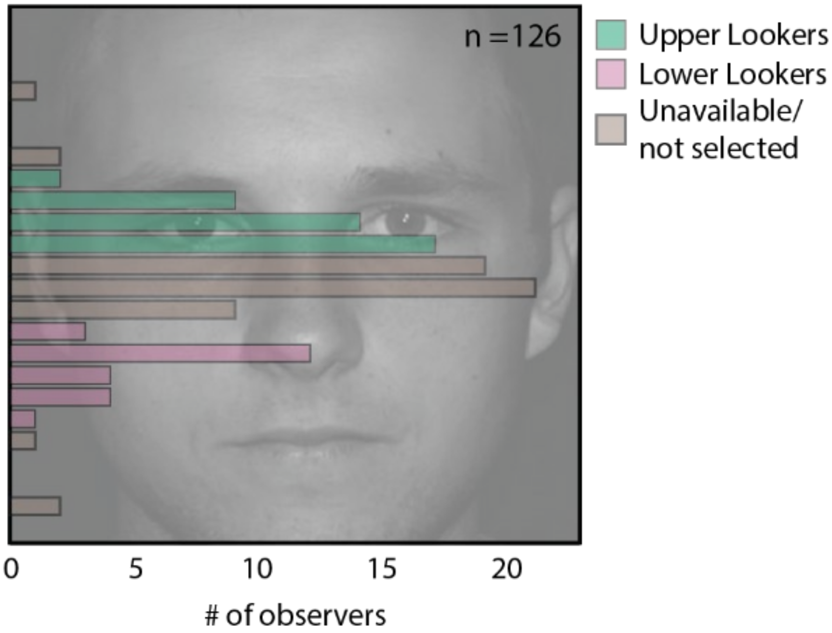


Figure 7. The panel shows the actual distribution of the vertical coordinates of the PFLs obtained from 126 screening participants. We selected the top and bottom 15% of the participants and invited them for further experiments. The upper lookers are depicted in green, and the lower lookers are depicted in pink. Those that were unavailable or not selected (due to poor quality data) are depicted in brown.

b) Free Eye Movements face ID task.

The first fixations on each trial of one upper looker and one lower looker are shown in the left panel of Figure 4a. The PFLs of the two groups are shown in the right panel of Figure 4a. The PFLs of the upper and lower looker groups were significantly different (t (30) = 12.8, $p << 0.05$). The average PFL for the upper lookers was 0.54° below the eye level, and that of the lower lookers was 2.1° below the eye level (see figure 8). There was no significant difference in face identification performance across the two groups (t (31) = 0.73, $p = 0.47$, $PC_{\text{upper-lookers}} = 87.04\%$ $PC_{\text{lower-lookers}}=83.98\%$). The two fixation positions for the enforced fixation task were chosen to be the average PFLs of the two groups (see right panel of figure 8).

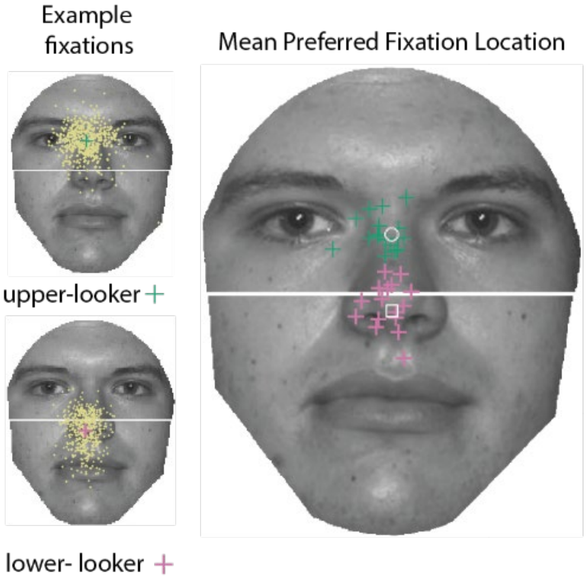


Figure 8. The left part of this panel depicts the landing positions of the first eye movements of an upper and lower looker across 320 trials. Their PFLs are shown with green and pink crosses, respectively. The right half of the panel shows the PFLs of the upper and lower lookers that participated in this study in green and pink crosses, respectively. The mean PFLs of these groups are shown with a white circle and square, respectively. The fixation position while viewing the faces in the enforced fixation sequential face-part matching task was varied between these two spots across trials for both groups.

c) Enforced Fixation sequential face part matching task.

The strength of the CFE was defined as the difference in face part matching task performance between the misaligned and aligned conditions. To characterize the variation of CFE in our experiment, we conducted a 3-way mixed factor ANOVA with *looker type* (upper vs. lower), *half being judged* (top vs. bottom), and *fixation position* (average PFL of upper lookers vs. that of the lower lookers) as factors. There was a significant main effect of the half being judged (F (1,29) = 21.13, p << 0.05), and a significant interaction effect of the *half being judged* and the *fixation position* (F (1,29) = 5.07, p = 0.03) The interaction between *looker type* and *half being judged* approached significance (F (1,29) = 4.0, p = .054). No other main or interaction effects were significant. The strengths of the CFE in various experimental conditions are plotted in Figure 9. Significant posthoc contrasts using Tukey's HSD comparisons across all possible pairs of conditions are also indicated for reference.

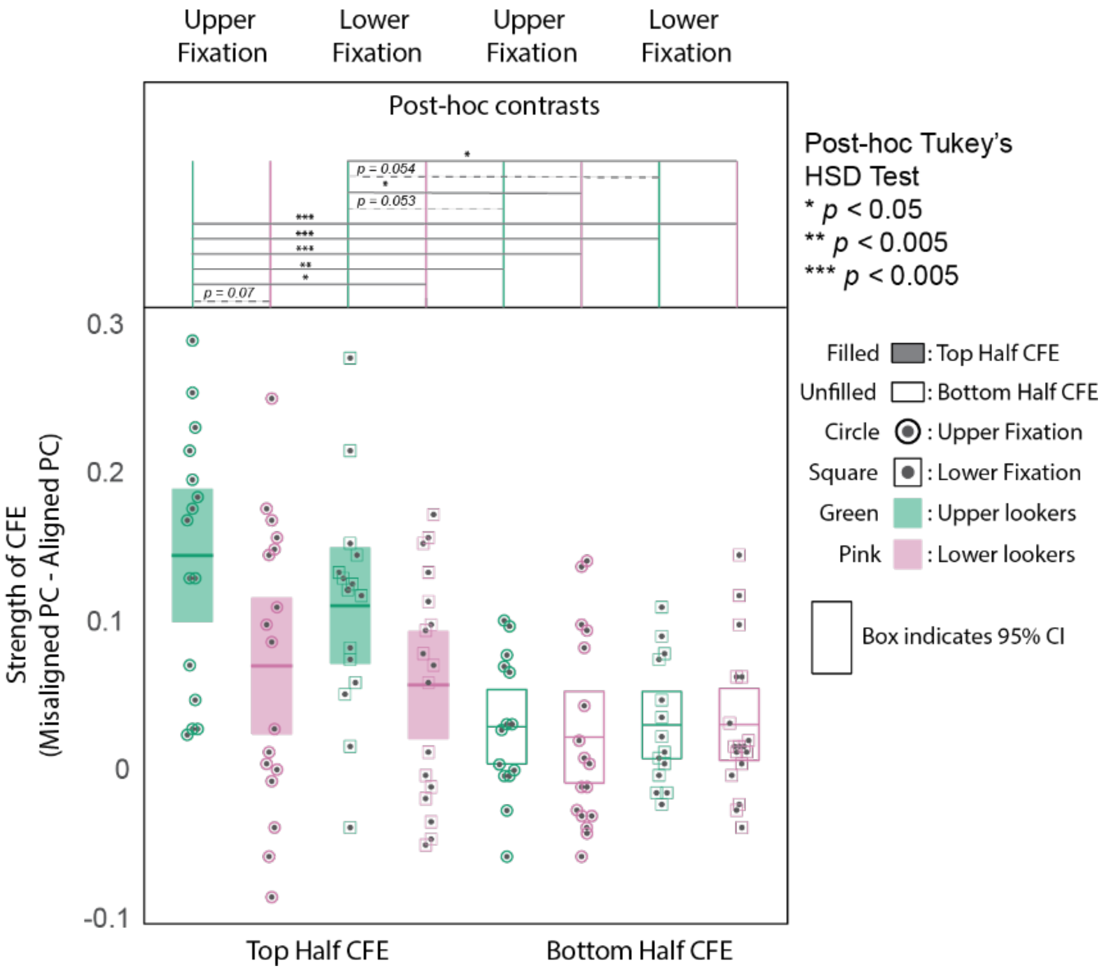


Figure 9. This figure shows the results of the face part matching task. The Y-axis shows the strength of the CFE, which was calculated as the difference between the accuracy in the misaligned and aligned conditions. The box plot shows the strengths of the top and bottom CFEs for each participant at the two fixation locations. The filled and unfilled boxes represent the top and bottom half CFEs, respectively. The boxes themselves depict the 95% confidence interval. Green and pink colors are used to represent upper and lower lookers, respectively. Finally, circles and squares represent conditions where observers fixated the mean PFL of the upper lookers and that of the lower lookers, respectively. The results of various post hoc Tukey's HSD contrasts performed after the 3-way ANOVA discussed in the text are depicted in the panel above the box plot.

The ANOVA revealed that the CFE for the top half is significantly stronger than for the bottom half ($\mu_{TopHalf}$ = 9.9%, $\mu_{BottomHalf}$= 3.1%). This is in agreement with several earlier reports of the CFE (Rossion, 2013). In his review, Rossion argues that the CFE should just be assessed for the top half since that is the measure of holistic processing that gives a higher signal-to-noise ratio, improving the likelihood of finding significant effects. Given the precedent for treating the CFE for these two effects as qualitatively different, we implemented the above ANOVA on the top and bottom half CFEs separately. Each time, we ran a mixed factor 2-way ANOVA with *looker type* (upper vs. lower) and *fixation position* (upper vs. lower) as factors. From the ANOVA on the top half CFE, found a significant main effect of *looker type* ($F(1,30)= 5.00$, $p = 0.032$) and a significant main effect of *fixation position* ($F (1,30) = 6.38$, $p = 0.017$). There was no significant interaction effect. ANOVA on the bottom half CFE yielded no significant main or interaction effects. A post-hoc Tukey test to further characterize the effect of *looker type* revealed that upper lookers had a significantly higher top half CFE compared to lower lookers ($\mu_{\text{upper}} = 13.1\%$, $\mu_{\text{lower}} = 6.7\%$, $p = 0.032$). This suggests that the proximity of the PFL to features on the upper half of the face results in a larger CFE. For the bottom half CFE, no significant difference was found between the two groups.

The ANOVA also revealed a main effect of fixation position. A post hoc Tukey test revealed that the top half CFE was stronger at the mean PFL of the upper lookers compared to that of the lower lookers ($\mu_{\text{upper}} = 11.1\%$, $\mu_{\text{lower}} = 8.7\%$, $p = 0.017$). This suggests that the CFE is stronger when the point of fixation is closer to the features in the top half, for both upper and lower lookers.

9. Discussion

Our goal in this study was to test whether individual differences in learned internal representations (mediated by different oculomotor strategies) modulate holistic processing. For this, we compared the strength of the CFE for two groups of observers with distinct

Preferred Fixation Locations (PFLs) on the face while they fixated either close to their own PFL or close to the other group's PFL. While many measures of eye movement preference have been used successfully, like fixation density maps (see Mehoudar et al., 2014), scan path analysis, and Hidden Markov Models (see Hsiao et al., 2021), we picked the PFL because it has been shown to play a functional role in a variety of face tasks when tested at different fixation locations on the face (Peterson & Eckstein, 2012). Besides, variations in PFL have been linked to variations in learned representations of faces (Tsank, 2019). Our data show that the PFL relative to the features of the face modulates the CFE for the top half but not the bottom half.

We found that the strength of the top half CFE for faces was influenced by the PFL in two ways. Firstly, observers with a PFL farther away from the top half showed a lower top half CFE, irrespective of the actual fixation location. Secondly, members of both upper and lower looker groups showed a stronger CFE when fixating the mean PFL of the upper lookers, although this trend was stronger for the upper lookers. These results indicate that both the location of the PFL and the actual gaze position relative to the features of the face modulate the top half CFE. However, these factors don't modulate the bottom half CFE.

The lack of variation of the bottom half CFE may be caused by the low overall strength of the CFE (9.9% on average for the top half compared to 3.1 % for the bottom half). The asymmetry in the strength of the CFE between the two halves agrees with previous findings (Heering et al., 2008; Rossion, 2013; Young et al., 1987). Some researchers have noted that the reduced CFE for the bottom half can be remedied by matching the difficulty (of identification) of the two halves (C-W Shyi & Wang, 2016). We did not match the difficulty of the halves in our study. The bottom half judgments were significantly easier than the top half. It is possible that the modulation of the CFE might extend to bottom half judgments if we had matched the difficulty. However, matching the difficulties using artificial means can make the faces look unnatural, which we wished to avoid when designing the experiment. Recently, Kurbel et al. (2021) reported that the CFE is equally strong whether the two halves appear natural and homogenous or unnatural and strongly segregated. This finding opens up the possibility of using composite faces with difficulty-matched halves to study if the PFL also modulates the CFE for the bottom half.

In our stimuli, the midpoint of the top half coincided with the position of the eyes on the face. The finding that the top half CFE reduces as the PFL away from the midpoint of the top half corroborates well with research showing the importance of the eye region in holistic processing. Itier et al. (2007) showed that the N170 EEG signal, which is commonly associated with holistic processing, is abolished when the eyes are removed from the face. Similarly, it has been noted that acquired prosopagnosia strongly affects the usage of eye information on faces (Caldara et al., 2005). Interestingly, prosopagnosia selectively impairs holistic perception (Ramon et al., 2010), but not other judgments on faces (Jiang et al., 2011; Quadflieg et al., 2012; Van Belle et al., 2011). Accordingly, it has been shown that patients with prosopagnosia fixate more on the mouth rather than the eyes (Orban de Xivry et al., 2008). This paper discusses an interesting theoretical framework that they credited to Caldara et al. (2005) to explain the pattern of observations. They suggest that the eye region is densely packed with fine details, which calls upon higher-level visual processing that can integrate these features into a whole. When the higher-level face processing areas are compromised, as in the case of acquired prosopagnosia, the observers fail to integrate these fine parts into a whole. However, this impairment reflects the strongest at the eye region, simply because there are more features to be integrated. Therefore, prosopagnosics naturally tend to increase reliance on the mouth region for face perception. This theoretical framework lower lookers process faces in a less holistic manner because there are fewer features that need to be integrated near their PFL compared to the upper lookers. This would be reflected as the reduced strength of the CFE. Our results are in-line with the predictions of this theory.

In summary, we suggest that long-term learning of the statistical variations in the incoming visual information from faces at the PFL can partially account for variations in the CFE.

D. Experiment 2: The Face Adaptation Effect

1. Participants

a) Prescreening Task.

A total of 234 observers participated in a short free eye movement face identification screening task. These participants were undergraduate or graduate students at the University of California, Santa Barbara, and participated in the study either for course credit or a small monetary reward. The experimental paradigm was identical to the prescreening task used in Experiment 1.

b) Main study.

We used a more stringent thresholding procedure in this study to only select two groups whose mean PFLs were 3.6° apart on the face on an average. This amounted to inviting only the top and bottom 4% of the prescreening participants. We found a total of 7 upper lookers and 6 lower lookers during the prescreening. All the upper lookers were female, with ages ranging from 19-23. The lower looker group consisted of 3 males and 3 females with ages ranging from 20-25. The participants completed the study in exchange for hourly monetary compensation. The whole study took 9-14 hours to complete and was completed by participants in multiple 1.5 hour-long sessions across a span of 1-2 weeks.

2. Stimuli.

a) Prescreening task.

The prescreening task used the same stimuli as in the composite face experiment (see Figure 4a).

b) Main study.

The main study consisted of a free-eye movements face ID task and an enforced fixation face adaptation task. We selected 4 Caucasian male faces from an in-house dataset of frontally photographed faces for the face ID task (see figure 10a). These faces were first standardized by rotating, cropping, resizing them such that the eyes and chin were centered and aligned. These standardized faces were then converted to an 8-bit grayscale format and embedded in a mask that only revealed frontal facial features. Both the luminance and contrast energy were matched so that skin color or texture variations cannot be used to judge the identity. A mask was applied to the faces that covered all external features, i.e., the hairline and the ears. The faces were 12.6° tall x 9.9° wide (see figure 10b).

We selected two out of the four faces (say A and B, see figure 10a) used in the face ID for the main face adaptation task. We created eight morphs of intermediate faces such that the morphs contained 15%,25%,35%,45%,55%,65%,75%and 85% of face B blended with face A (see figure 10c). To do this, we first used a state-of-the-art deep learning-based face landmark registration algorithm to fit 60 landmarks to the two faces, outlining the various features (Bulat & Tzimiropoulos, 2017). The images were then divided into triangles using Delaunay Triangulation. A parametrized affine transform was applied to each of these triangles to generate morphs with the required levels of A and B. One of the original faces was used as the adapter, while the various morphs were used as test faces.

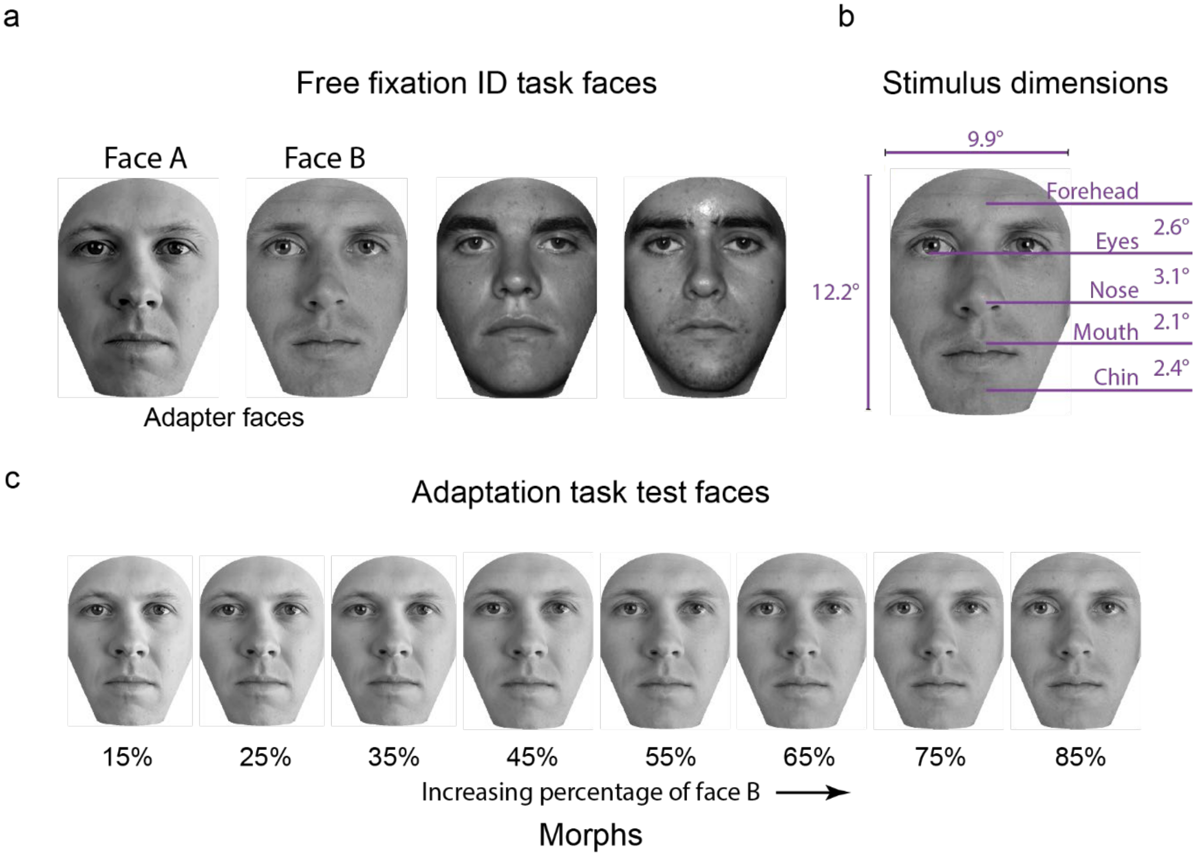


Figure 10. Various face stimuli used for the adaptation experiment. (a) shows the faces used in the free fixation face ID task. The first two faces in this panel were also used as adapter faces in the subsequent enforced fixation matching task. (b) shows the dimensions of the faces used in this study. (c) shows the morph faces used as test stimuli in the enforced fixation matching task.

3. Apparatus

The stimuli were presented on a Barco monitor with a resolution of 1280 x 1024 and a refresh rate of 60 Hz. The monitor was calibrated linearly with a maximum luminance of 114.7 cd/m^2. The screen was placed at 75 cm from the participant's eyes, such that each pixel on the monitor subtended a visual angle of 0.021° on their eyes. The stimulus display was controlled by software written using PsychToolbox-3 (Brainard, 1997) running on MATLAB 2018.

An Eye Link 1000 Plus tower mount eye tracker was used for tracking the left eye of each participant. The sampling rate of the tracker was 250 Hz. A 9-point calibration procedure was used at the beginning and repeated periodically to ensure accurate gaze data recording. Standard algorithms from the EyeLink toolbox were used to identify saccade and fixation events from the gaze data.

4. Experiment Design.

The full experiment consisted of two tasks: the free fixation face identification task and the enforced fixation face adaptation task. The design of the face ID task was identical to that in the composite face effect experiment, except that we used a different set of 4 faces (see the section *Stimuli* above). We now describe the enforced fixation face adaptation experiment.

a) Enforced-fixation face adaptation task.

We used this task to measure the Face adaptation after effect while observers maintained their gaze at a specific location on the face (which was manipulated). The two fixation locations were chosen on the vertical midline of the face, corresponding to the average PFLs of the two groups (upper and lower lookers). Each observer completed 3 (adapter conditions) x 2 (test face conditions) x 8 (morph levels) x 48 (repeats)= 2304 trials. The trials were distributed across 24 evenly sized blocks. The three adapter conditions were: no adapter, adapter at the own group's mean PFL, and adapter at the other group's average PFL. Likewise, the test face could be shown at the average PFL of the observer's own group or that of the other group. The test face was sampled from one of the eight possible morphs.

The fixation positions on the adapter and the test face were fixed within a block, while the test face was chosen uniformly from one of the eight possible morphs.

5. Trial design.

The trial design for the free fixation task was identical to that in the composite face effect experiment (see figure 11, upper panel). We now describe the events that occurred within each trial in the face adaptation experiment.

In this task, the initial fixation always occurred at the center of the screen. The fixation cross was black and overlaid on a gray background (luminance ~ 57 cd/m^2) which remained unchanged throughout the block. At the beginning of the trial, the observer maintained their gaze at the fixation cross, indicated readiness by a keypress. After the keypress, the program verified if the observer's gaze stayed within 1° of the fixation cross for a variable delay period of 500 msec to 1500 msec. If the gaze drifted beyond the 1° threshold, the trial was aborted. If the observer successfully maintained their gaze through the delay period, the trial progressed. There were 3 (no adapter/ adapter at PFL/adapter at other groups PFL) x 2 (test at PFL/test at other group's PFL) = 6 unique conditions in this experiment. Within a block, the observer would only see one condition. If the trial had an adapter, an adapter face was displayed on the screen. The position of the adapter was adjusted such that the observer's fixation (which they maintained at the center of the screen) fell on the average PFL of their own group or that of the other group, based on the condition. The adapter was presented at full contrast for 4 seconds. The fixation cross was lightened and overlaid on the face as a reference for the observer. During the 2 seconds, the program checked the gaze of the observer to prevent eye movements. If the gaze position drifted beyond 1° from the fixation cross, the trial was terminated. After adaptation, a test face that was uniformly sampled from the eight morphed faces was flashed for 200 msec. The position of the test face was also adjusted based on the condition (test face at own group's average PFL or the other group's average PFL). After that, a white Gaussian mask with the mean luminance matched to the background and a standard deviation of 11.2 cd/m^2 was flashed for 500 msec. The purpose of the mask was to wash out any lingering percept of the test face. Then a response screen with the two original faces used for the adaptation experiment appeared. The screen stayed on till the observer indicated which test face was shown with a mouse press. After the response, no feedback was given, and the next trial was initiated. See the lower panel of Figure 11 for the trial schematic.

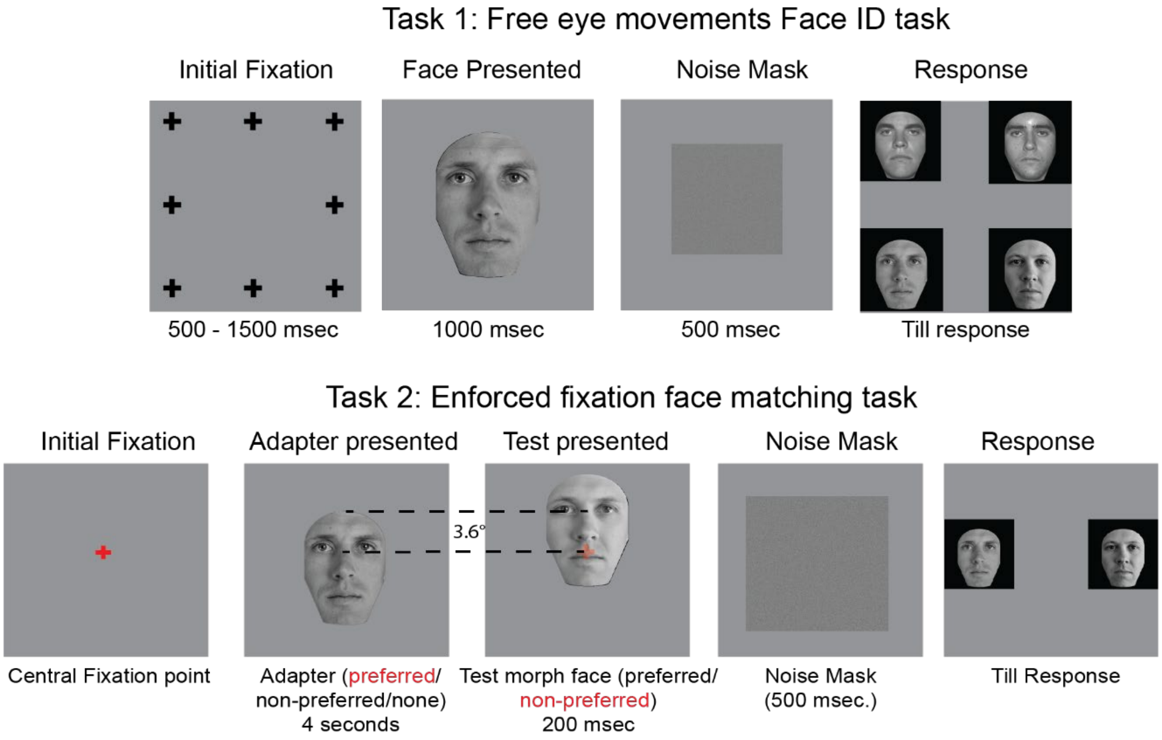


Figure 11. This figure shows the schematics of the two tasks. (a) depicts the free fixation face identification task. Observers initiated the trial fixating one of the 8 possible peripheral locations. A face is then presented in the center of the screen, and observers can freely move their eyes while studying the face. In the next screen, they are required to indicate which face was shown using a mouse click. (b) depicts the enforced fixation matching task. Observers initiated the trial fixating at the center of the screen. On adaptation blocks, an adapter face appeared such that the fixation was either at the observer's own group or other group's PFL. After that, a test face appeared either at the observers' own group's or other group's PFL. After a brief noise mask, observers were tasked to indicate which test face they saw. On one-third of the blocks, there was no adapter, and the test face was shown directly.

The experiments were administered by trained graduate or undergraduate researchers in accordance with protocols approved by the Institutional Review Board (IRB) of The University of California, Santa Barbara. Participants were first briefed about the nature of the study and compensation agreements. After obtaining consent to be a part of the study, they were given instructions about the task. The main experiment consisted of 2 tasks: the free fixation face identification task and the enforced-fixation face adaptation aftereffects task. The two tasks were always performed in the same order, i.e., the free fixation task followed by the enforced fixation task. This was to verify that the participant had a consistent PFL on faces used in the adaptation experiment that matched their PFL as measured in the prescreening task. Note that these tasks were conducted in parallel across observers – as soon as a participant finished the free fixation task, they could start with the forced fixation task. This timeline was separate for each participant. Each task was further divided into blocks that took 15-20 minutes to complete. Participants were encouraged to take breaks between sessions and were not allowed to spend more than 1.5 hours per session to avoid the effects of fatigue. We recalibrated the eye tracker between blocks and whenever a participant took a break to maintain eye-tracking quality throughout each session.

6. Analysis

a) Preferred fixation location.

The preferred first fixation location is defined as the first location inside the face that an observer's foveal region lands on when they make an eye movement to a face from a peripheral fixation location. For each observer, following the completion of all blocks of the free fixation task, the preferred first fixation location was estimated as the mean fixation location across all the first fixations on the face across trials. The vertical coordinates of the first fixation location on the face were used to analyze the dependence of FAE on the gaze position.

b) Strength of adaptation

We first calculated the fraction of times the observer responded to each morph level as face B. We had 48 responses for each morph level in each condition. Thus the smallest difference in response rate we could measure was ~ 0.2%. For each condition, we fit a psychometric function of the form

$$\psi(x;\gamma,\lambda,\mu,\sigma) = \gamma + (1-\gamma-\lambda)F(x;\mu,\sigma) \ \ldots \ (1)$$

where x is the % morph level, γ is the guess rate, λ is the lapse rate, μ and σ are parameters of the fitting distribution(Wichmann & Hill, 2001). The accuracy of the fitting was ensured by using the least absolute residual (LAR) method and further by manually examining the fits in each condition to remove any poorly fitted data. The *point of subjective equality* (PSE), defined as the strength of the stimulus that elicits both possible responses with equal probability, was computed from the fit equation for each condition. The strength of adaptation was then calculated as the difference between PSE with adaptation and the PSE with no adaptation, all else being equal.

c) Power analysis.

The choice of the number of subjects per group was made based on apriori power analysis done using data from a pilot experiment, where we measured the effect of manipulating the enforced fixation position by 3.6° on the face. In the pilot experiment, we did not use a between-subjects design. The pilot experiment revealed a strong interaction between adapter position and test position (*Cohen's f* = 1.03) and a weaker within-subjects main effect of adapter position (*Cohen's f* = 0.95). We estimated that we would need around six subjects per group to capture within-subjects effects with a (1-β) rate of 0.95, assuming the distance between the PFLs of the two groups was 3.6°. Based on the power analysis for the composite face effect (see figure 6), we estimated requiring to screen around 300 participants to find two groups with the requisite mean distance between PFLs. We also conducted a post hoc power analysis to compute the achieved power, the results of which are described in the results section.

d) Statistical Testing.

We used a 3-way mixed factor ANOVA for analyzing the trends in the adaptation strengths from different conditions. In addition, a non-parametric bootstrap analysis was carried out to validate the outcomes. The bootstrap analysis aimed to test if the change in adaptation strength between the own group's PFL and the other group's PFL was significantly different for upper and lower lookers. For this, we first calculated the relevant difference ($\Delta FAE = FAE_{PFL} - FAE_{non\text{-}PFL}$) for each observer. Then we divided the observers into upper and lower lookers and created 5000 bootstrap samples with replacement for each group. We then estimated the 95% confidence intervals of the difference between the two groups. The results are visualized using Gardner Altman plots with the help of software adapted from Ho et al. (2019).

7. Results

a) Prescreening task.

The distribution of the vertical coordinates of the PFL's of 286 observers in the prescreening task is shown in Figure 12. The distribution was skewed towards lower locations on the face (sample skewness = 1.04). The upper and lower lookers were selected from the 4 percentile tails of the distribution shown in purple and green, respectively. We selected 8 participants from each group. This was done to maximize the chances of finding a significant effect of manipulating the PFL on the face adaptation aftereffect. One lower looker was excluded because we could not replicate their lower looking behavior in the main experiment. One lower looker and one eye looker could not continue with the study, resulting in a final tally of 5 lower lookers and 7 upper lookers.

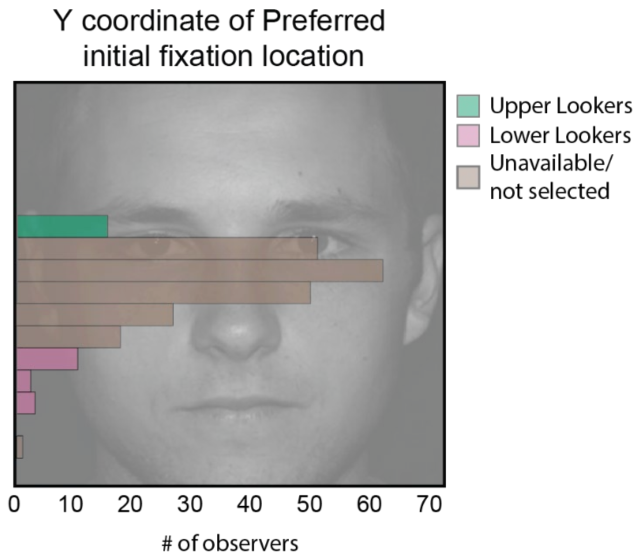


Figure 12. The panel shows the actual distribution of the vertical coordinates of the PFLs obtained from 234 screening participants. We selected the top and bottom 2.5% of the participants and invited them for further experiments. The upper lookers are depicted in green, and the lower lookers are depicted in pink. Those that were unavailable or not selected (due to poor quality data) are depicted in brown

b) Free eye movements face ID task.

First fixations across trials of one upper looker and one lower looker are shown in the left panel of Figure 13. The PFLs of the two groups are shown in the right panel of Figure 13. The PFLs of upper and lower lookers were significantly different ($t(10) =$ p <0.05). The average PFL of the upper lookers was 0.8° below the eye level, while that of the lower lookers was 4.4° degrees below the eye level. There was no significant difference in the face identification performance between the two groups (stats). The two fixation locations in the enforced fixation task were chosen to be the average PFLs of the two groups (see right panel of figure 13).

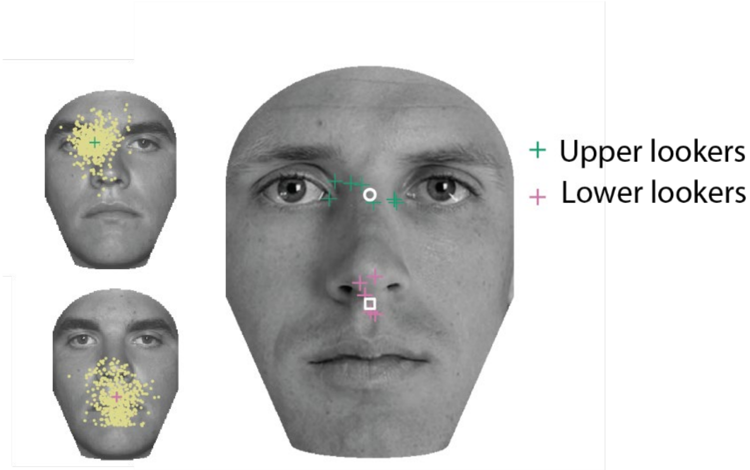


Figure 13. The left panels show the first fixation of an example upper and lower looker. The right pane shows the PFLs upper and lower lookers as green and pink crosses, respectively. The white circle and square represent the mean PFLs of the two groups, which were chosen as the points of fixation on the face for the enforced fixation face-matching task. The distance between these two fixation positions was 3.6°.

c) Enforced fixation face adaptation task.

To characterize the strength of the adaptation, we extracted the PSEs for each condition by fitting a psychometric function to the responses of the observer to different morph levels. The strength of the FAE for a condition was calculated as the shift in the PSE between conditions with and without the adapter, all else being the same. The fits for one eye looker and one lower looker are visualized in figure 14.

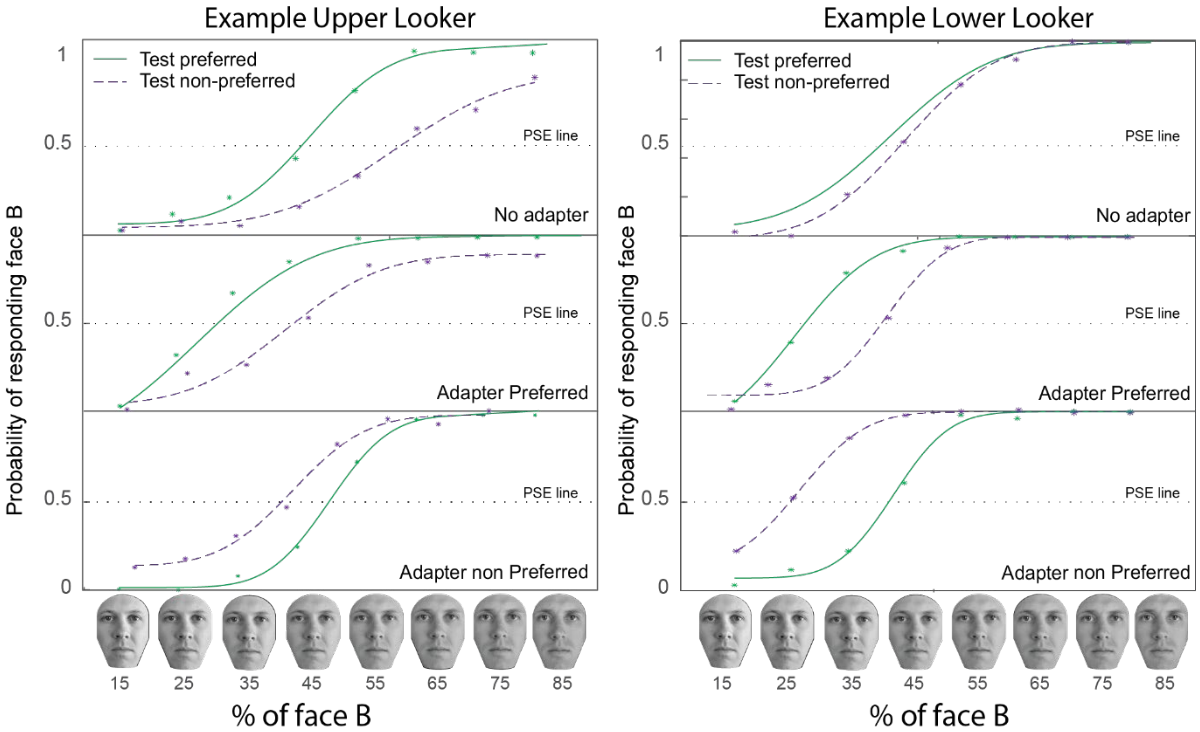


Figure 14. This figure shows the psychometric fits for various conditions for one upper looker and one lower looker. The first, second, and third subplots within each plot represent the *no adapter,* the *adapter at the own group's PFL*, and the *adapter at the other group's PFL* conditions, respectively. The green and purple traces represent the *test at the own group's PFL* and the *test at the other group's PFL* conditions, respectively. The X and Y-axis denote the percentage of face B in the test face and the probability of responding with face B for the test stimulus. The PSE was calculated as the abscissa of the psychometric curve at which there was an equal probability of responding as A or B. The strength of the FAE for a given adapter position was calculated as the difference between the PSE of that condition and an equivalent condition with no adapter

The strength of the face adaptation effect (FAE) was defined as the difference in the PSEs between conditions with and without the adapter, all else being the same. Figure 17 illustrates the procedure of calculating the PSE from the psychometric function fits for one upper looker and one lower looker. To characterize the variation of the FAE in our experiment, we conducted a 3-way mixed factors ANOVA with *looker type* (upper vs. lower), *adapter position* (adapter at PFL/adapter at non-PFL), and test position (PFL vs. non-PFL) as factors on the strength of adaptation in each condition. Here, non-PFL refers to the other group's PFL (see Experiment Design). Data points that had a median absolute deviation higher than 1.5

within each condition were treated as outliers and removed. We found two outliers and removed them. Keeping or removing the outliers did not qualitatively change the results of the analysis. There was homogeneity of variances in the within-subjects factors, as assessed by Levene's test of equality of variances ($p > 0.05$). Mauchly's test of sphericity indicated that the assumption of sphericity was valid ($\chi^2(5) = 9.22$, $p = 0.10$). The ANOVA showed no significant main effects, although the main effect of *adapter position* tended towards significance ($F(1,9) = 5.03$, $p = 0.052$). There was a significant *adapter position* x *test position* interaction effect ($F(1,9) = 58.47$, p <0.005), and a significant *adapter position* x *looker type* ($F(1,9)= 7.55$, p =0.023). No other effects were significant. Figure 15 shows the adapter strengths for various conditions.

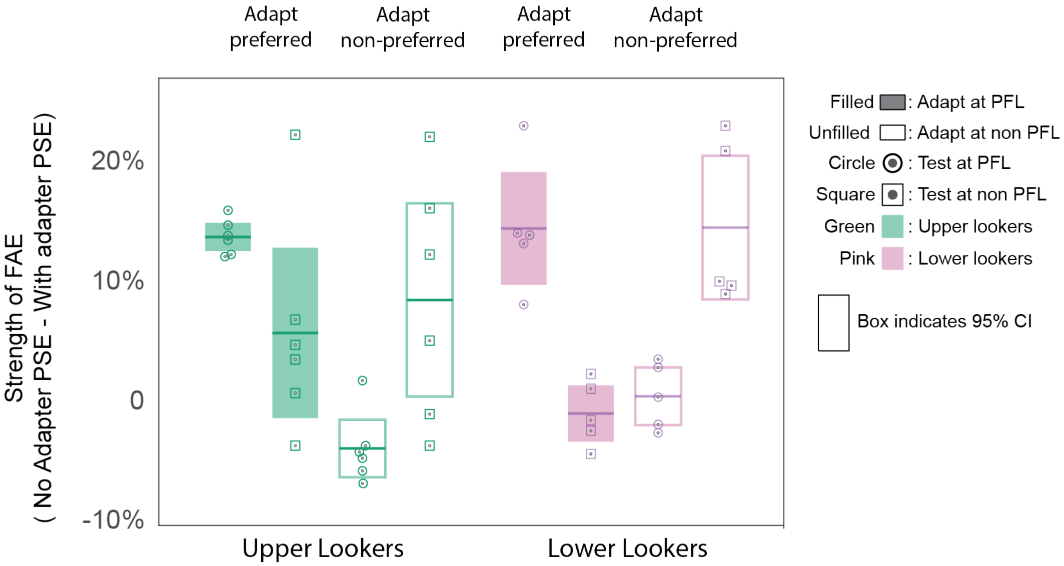


Figure 15. This figure shows the strength of the FAE for different conditions. The green and pink data points represent upper and lower lookers, respectively. The columns with filled boxes denote conditions where the adapter was at the observer's own group PFL whereas the columns with unfilled boxes represent the conditions where the adapter was at the other group's PFL. Likewise, Circle and square markers denote the conditions where the test was at the own group's PFL or the other group's PFL, respectively. The boxes indicate 95% confidence intervals.

The strong *adapter position* x *test position* interaction suggests that the adaptation was stronger when the adaptation and testing occurred at the same retinotopic location than when the location was different. Planned post hoc contrasts revealed this trend. When adaptation happened at the PFL, there was a significant difference in adaptation strength between the two testing locations (*t* (9) = 5.92*, p* <0.0005). This was also true when adaptation happened at a non-PFL (*t* (9) = - 4.08, *p* <0.005).

The *adapter position* x *looker type* interaction tells us how variation in the PFL affects the FAE. Planned post hoc contrasts revealed that there was a significant difference between the FAE when adaptation happened at the observer's own group PFL compared to the other group's PFL (averaged across both test positions) for the upper lookers (*t*(8) = 3.49, *p* = 0.0082). This comparison was not significant for lower lookers (*t*(8) = 0.026, *p* = 0.73). This shows that the upper lookers had a higher cost to the FAE of adapting to faces farther away from their PFL than lower lookers.

Even though we verified various underlying assumptions of the ANOVA, to allay the concerns of having a small sample, we repeated the analysis using a non-parametric bootstrap test (see analysis section for details). The bootstrap analysis revealed that the change in the adaptation strength across positions was significantly greater for upper lookers than for lower lookers (ΔFAE_{Upper} = 6.69%, ΔFAE_{Lower} = -0.75%, 95% CI of difference = [2.26% 12.2%], Cohen's *d* = 1.51). The results of the bootstrap sampling are visualized in Figure 16.

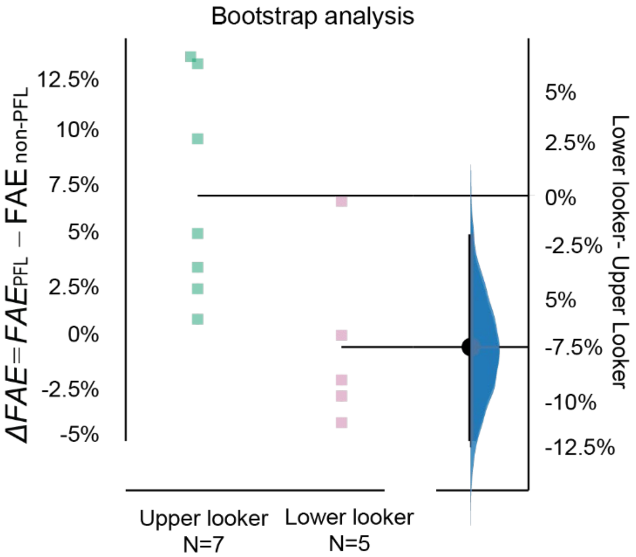


Figure 16. Non-parametric bootstrap analysis to establish the significance of the *adapter position* x *looker type* interaction effect. In this Gardner -Altmann chart, the green dots denote upper lookers, while the pink dots denote lower lookers. The Y-axis is the difference between the FAE at own group PFL and that of the other group. This suggests that the upper lookers had a higher cost of fixating away from their PFL compared to lower lookers.

A post hoc power analysis revealed an achieved (1-*β*) power of 0.60 for between-subjects factors and 0.37 for within-subjects factors. Thus, the experiment may be underpowered to detect some weaker effects. However, the experiment was able to reject the null hypothesis for the effects of interest at this sample size.

8. Discussion

Our goal in this study was to test whether visual experiences of faces (mediated by eye movement strategies) modulate the face adaptation aftereffects (FAE) for identity. For this, we measured the FAEs for upper and lower lookers (see methods for precise definition) while manipulating the position of the adapter and test stimuli to fall at either the mean preferred fixation location (PFL) of their group or that of the other group. Our results showed that the FAE was stronger when adaptation and testing occurred at the same retinal location for both groups. We also found that upper lookers showed a greater reduction in the FAE than lower lookers when the adaptation happened at their own group's PFL vs. that of the other group's PFL. In other words, upper-lookers were more sensitive to changes in the adapter position than lower-lookers.

Adaptation aftereffects have been referred to as the *psychophysicist's electrode* because they allow us to test if the same neuronal population mediates two effects (Webster, 2015). These aftereffects arise from the gradual but temporary desensitization of neurons when activated by features they are tuned to. Hence, by comparing adaptations effects, one can infer whether the same neuronal populations are active. Firstly, our finding a strong *adapter position* x *test position* interaction suggests that different subpopulations of neurons mediate face recognition at the PFL and a non-PFL. This is in agreement with earlier results showing that neuronal representations of faces are tuned to typical the retinotopic position of faces (Issa & DiCarlo, 2012; Stacchi et al., 2019). Secondly, our finding of a significant *adapter position* x *looker type* interaction showed that the FAE varied more with a change in adaptation position from PFL to non-PFL for upper lookers compared to lower lookers. We interpret this to mean that neural representations of faces in upper lookers are more tuned to the PFL than in lower lookers. In other words, the neural representations of lower lookers are more invariant to fixation position relative to the PFL.

While fMR-Adaptation based paradigms (Grill-Spector & Malach, 2001) have been used extensively in research on face processing, fewer studies have studied the locus of the FAE itself (Cziraki et al., 2010; Furl et al., 2007; Kovács et al., 2008). Most research has focused on the role of two areas: the Fusiform Face Area (FFA) and the Occipital Face Area (OFA). On the one hand, Kaiser et al. (2013), Rotshtein et al. (2004) showed that the FFA is involved only in identity after effects, while on the other hand, Xu et al. (2009) maintain that the FFA encodes the physical features (in the absence of an identity change) of the person also. Finally, Tsantani et al. (2021) show that both the FFA and OFA encode identity information: the FFA is involved more in high-level aspects of identity, while the OFA is involved in lower-level features. The FFA and OFA are thus the most likely loci for the face identity after effect. Kovács et al. (2008) manipulated the position of the face horizontally with respect to the fixation position and found that the adaptation effects in FFA are position invariant, while those in OFA are position-specific. However, in our experiment, we manipulated the adaptation and test locations on the face vertically. Given earlier literature suggesting that horizontal and vertical dimensions of faces are processed differently (Dakin & Watt, 2009; Goffaux & Rossion, 2007), we cannot relate our findings to the results of Kovács et al., (2008).

Our results also add to the ongoing discussion on whether the FAE is retinotopic. Some studies have reported that the FAE is position invariant, while the others maintain that the FAE is position-specific (Zimmer & Kovács, 2011). For example, Leopold et al. (2001) manipulated the position of the test stimulus on the retina relative to that of the adapter. Subjects adapted at different locations on the midline of the face and made a saccade to the mid-point of the eyes when they heard a beep to indicate the onset of the test stimulus. They found that the adaptation transferred in saccades up to 6° long. Afraz & Cavanagh (2008) also obtained a similar result. Whether there is transfer across retinal locations or not is also based on the duration of adaptation: Kovács et al. (2007) snd Xu et al. (2008) found no transfer across retinal locations in a face expression adaptation task. Our results also suggest that while the FAE is largely retinotopic, and more so for lower lookers than upper lookers. Thus, the PFL of the observer should also be considered in the debate of position specificity of the FAE. Evidence is also emerging that the position specificity of the FAE is also related to the adaptation time (Kovács et al., 2007, 2008). Adaptation for shorter time scales (~500 msec) may make the FAE more position invariant compared to longer time scales (>2 seconds) which may make the FAE more position-specific(Zimmer & Kovács, 2011). Our finding of a low level of retinotopic transfer across retinal locations (as indicated by the strong *adapter position* x *test position* interaction) may thus be due to the use of a long adaptation time of 4 seconds.

E. General Discussion

The goal of this chapter was to test if long-term differences in visual experience mediated by differences in oculomotor strategies of viewing faces modulate face effects like the composite face effect (CFE) and the face adaptation after effect (FAE). The role of the first fixation position on the face (preferred fixation position, PFL) in common face tasks like face recognition, gender identification, and expression recognition has already been established (Peterson & Eckstein, 2012; Tsank, 2019). Here we aimed to learn more about the mechanisms driving face perception by testing if the PFL modulates these well-studied face effects.

We found that the location of the PFL modulates both the CFE and the FAE, as evidenced by interaction effects with the *looker type* variable. The top half-CFE was found to reduce with the distance of the PFL to the eyes. While the magnitude of the FAE itself was not different for upper and lower lookers, we found that the FAE was more position-specific to the actual adaptation location on the face for the upper lookers. On the other hand, for the lower lookers, the strength of the FAE was more invariant to the adaptation location. These results highlight the role of visual experience in day-to-day face tasks and in influencing the outcomes of tasks designed to understand the mechanisms underlying face processing.

What do these findings tell us about the role of visual experience in shaping face processing mechanisms? First, we discuss the CFE. As noted in the introduction, the CFE represents an inflexibility of the visual system to process only a part of the face when the full face is available. Our finding that the CFE is stronger at the PFL than at the non-PFL adds to an increasing body of literature that this inflexibility can be related to learned internal representations when the faces are viewed in a familiar context. For example, the CFE is abolished or diminished when tested on inverted faces (Fernando et al., 2013), other-race faces (Michel, Caldara, et al., 2006), faces viewed at a distance of 24 meters (Ross & Gauthier, 2015, familiarity with size), and stereo distortion (Taubert & Alais, 2009). Further, the finding that upper lookers have a stronger CFE than lower lookers suggests that the content of the learned representations also matters for the CFE. One possible explanation for this is that upper lookers need to pool a larger area at their PFL to integrate all identity-relevant features. The CFE is designed so that the line separating the two halves cuts through this region of the processing. A larger portion of the stimulus features processed by upper lookers falls in the irrelevant half, compared to lower lookers. Therefore, the interference of the irrelevant half is larger for upper lookers. More research is needed to establish the exact mechanism by which this variation occurs. However, our findings establish that the the contents of the learned representations of face features at the PFL also modulate the CFE.

Now, we turn our attention to the results of the FAE task. Our finding that the FAE is more position-specific for upper lookers than for lower lookers suggests that the learned internal representations of upper lookers are more tuned to the PFL than those of the lower lookers. Position specificity of the FAE has been investigated often with varying outcomes (Zimmer & Kovács, 2011). The emerging consensus seems to be that both position-specific and position-invariant neural mechanisms mediate the FAE. An influential model of face processing suggests that these mechanisms may be involved in processing changeable and invariant aspects of faces (Haxby et al., 2000). Changeable aspects include gaze direction, mouth shape, eyebrow to eye distance, etc. Invariant aspects can be face shape, the configuration of facial features, skin tone, etc. This sets up our interpretation of the role of prolonged visual experience on the FAE: upper and lower lookers may differ in the degree of position-invariance of the FAE based on the distribution of changeable and invariant features around their PFL. However, it is unclear at the moment whether learning more changeable features leads to a greater or lesser degree of position invariance of the FAE. Further research is required to clarify the exact mechanisms by which the FAE is modulated.

In summary, our results further extend the evidence that differences in the visual experience of faces resulting from long-term use of different oculomotor strategies to view faces mediate face perception not only in day-to-day face tasks but also in face tasks designed to investigate various possible mechanisms underlying face perception.

III. The role of prolonged visual experience in the face inversion effect: evidence from a spatial summation task

A. Abstract

It is well-known that humans are worse at recognizing inverted faces than upright faces. Literature suggests that the poorer performance for inverted faces is due to the disruption of *holistic processing*, a mode of information processing that pools information from the whole face. Further, this disruption has been linked to an individual's visual experience: typically, humans see upright faces far more often than inverted faces. However, not known is how experience can bring about holistic processing in upright faces. We hypothesized that holistic processing arises from the visual system's ability to integrate information efficiently from a larger area. To study this, we compared the efficiency of information processing for upright and inverted faces viewed through an aperture of varying sizes. We show that face processing is more efficient for upright faces compared to inverted faces only for larger apertures but not for small apertures. To further understand how this may occur in the visual cortex, we measured the efficiency of a Convolutional Neural Network trained on full upright faces on the upright and inverted faces behind apertures. We found a similar divergence between efficiencies of the network for upright and inverted faces, such that upright faces were processed more efficiently as the size of the aperture increased. However, if we restricted the effective receptive field size of the CNN by reducing the number of layers, the divergence in efficiency with increasing aperture size was abolished. Therefore, the network's ability to integrate information from larger spatial areas is linked to the origin of the face inversion effect. To establish the role of visual experience in shaping this relationship, we repeated this

analysis with a CNN that was trained on full inverted faces and tested on upright or inverted faces behind apertures. We still found a divergence of efficiencies with increasing aperture size, but this time the pattern was flipped: inverted faces were processed more efficiently than upright faces for larger aperture sizes. Taken together, our results show that long-term visual experience shapes our preference for upright vs. inverted faces by selectively improving the visual system's efficiency in pooling information across space for upright faces.

B. Introduction

Inverted faces are hard to recognize. This fact was perhaps known for a long time but did not garner much attention from the scientific community. This changed when Yin (1969) demonstrated that the inversion effect is larger for faces than for other objects and suggested that there much be something special about faces. He reported that his subjects mentioned being unable to see the face as a *whole* when it was inverted. He speculated that the ability to see upright faces as wholes distinguishes them from objects. Further, this occurs because we usually see faces as mono-oriented (mostly upright). This paper triggered three new lines of research. In no particular order, these are the debate of domain specificity of face processing (Kanwisher, 2000), the role of expertise in face and non-face processing (Gauthier, 2020), and holistic processing of upright faces (Richler et al., 2012). While each of these research areas has been richly explored in the past 50 years, an integrative account of why upright faces but not inverted faces are processed holistically and how visual experience brings gives rise to this domain-specific mechanism is lacking.

Poltoratski et al. (2021) recently showed that holistic processing arises due to the spatial summation of information in face-selective regions. They mapped population-receptive fields (pRFs) of various brain regions to upright and inverted faces and showed that the pRFs for regions higher up in the visual hierarchy (like FFA) but not for lower regions shrank for inverted faces when compared to upright faces. This result provides a strong unifying link between domain specificity and holistic processing. Holistic processing occurs because the brain is able to pool information from a larger spatial extent for stimuli that activate the FFA, a region that has been implicated repeatedly in face processing (Kanwisher & Yovel, 2006). While Poltoratski et al. (2021) discuss the implications of this finding for how visual experience might shape holistic processing, they do not present evidence supporting any related claims. Our aim in this chapter is to provide evidence for this link.

To do this, we implemented a spatial summation experiment similar in design to Tyler & Chen (2006). They demonstrated that observers display a larger region of spatial summation relative to inverted faces in a face detection task. We aimed to replicate the results with a face identification task. For this, we used an adaptive staircase procedure to measure the contrast thresholds of humans on an upright or inverted face identification task where apertures of different sizes covered the faces. The efficiency of spatial processing was estimated by comparison with a Bayesian Ideal Observer. We hypothesized that as the aperture size increases, humans would continue to efficiently process upright but not inverted faces. We repeated the above comparisons using Convolutional Neural Networks (CNNs) to show the role of experience in shaping such an effect. Here, we hypothesized that the training set modulates the relationship between the efficiency and aperture size for upright and inverted faces. If the network is trained on upright faces, the efficiency must improve with increasing aperture size for upright but not inverted faces. Likewise, if the network is trained on inverted faces, the efficiency must improve with aperture size for inverted but not upright faces.

Finally, we test whether the degree of pooling in the CNN modulates the divergence in efficiency between upright and inverted faces as the aperture size increases.

C. Methods

1. Participants

A total of 8 observers (2 male and 6 females, mean age = 21.6) participated in an enforced-gaze position adaptive staircase face identification task. These participants were all undergraduate or graduate students at the University of California, Santa Barbara, and participated in the study either for course credit or monetary reward. All participants had normal or corrected-to-normal vision.

2. Stimuli

a) Face stimuli

We used an in-house dataset of 5 Caucasian male frontally photographed faces. These faces were standardized by rotating, cropping, and resizing to align the eyes and chin across stimuli. We also matched the luminance and contrast energy across all faces to ensure that low-level features don't drive identification performance. The full faces were 10.5° x 8.4° in size, but during the experiment, they were presented behind smooth apertures of varying sizes. In the experiment, the faces would appear either in the upright orientation or the inverted orientation.

b) Apertures

We created circular apertures to cover different degrees of visual information from the face. All apertures were centered at the midpoint of the face, which was a point on the nose for each face. To create the apertures, we first created binary circular masks of diameters 0.88°, 2.42°, 3.96°, and 5.50°. Next, we applied a radial linear-decay filter that attenuated the visual signal linearly as a function of the distance from the center of the mask at a specified rate of 80% per degree of eccentricity beyond the perimeter of the mask. We did this to avoid any cropping artifacts. We used a linear-decay filter instead of a Gaussian filter because the monitor was calibrated linearly (more details in the apparatus section below). These apertures revealed 2.0%, 15.2%, 40.7%, and 78.5% of the face when applied. Figure 17 shows all the stimuli as viewed through these apertures.

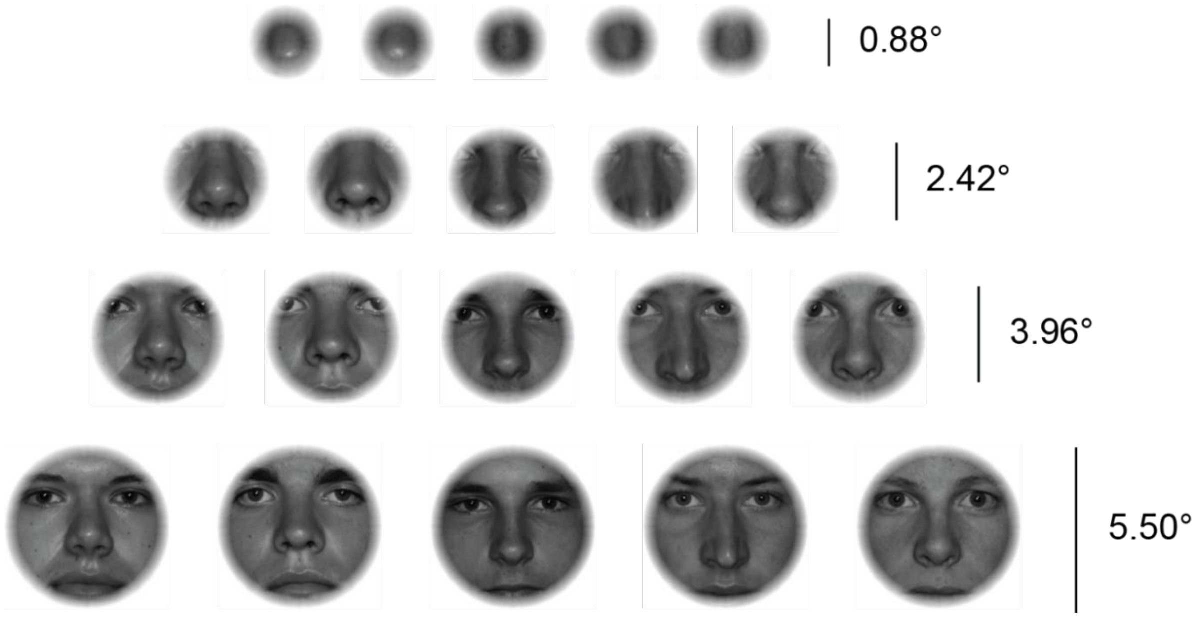


Figure 17. Various stimuli used in this study. Each column depicts increasing aperture sizes. The same set of stimuli were inverted for use in the inverted face condition. Note that these are noise-free full contrast versions of the stimuli. These stimuli had added white Gaussian noise during the actual task, and their contrast was varied adaptively within each block based on the observer's responses

3. Apparatus

We used a Barco monitor with a resolution of 1280 x 1024 and a refresh rate of 60 Hz. The monitor was calibrated linearly with a maximum luminance of 114.7 cd/m^2. Linear calibration ensures that a unit change in the pixel value amounts to approximately the same change in absolute luminance on the monitor across the dynamic range of the monitor. This is an important assumption for calculating the Bayesian Ideal Observer (BIO) performance, which assumes independence of white Gaussian noise in terms of luminance (rather than just in pixel values). The experiment took place in a dark room with minimal ambient light. The screen was placed 75 cm from the participant's eyes, such that each pixel on the monitor subtended a visual angle of 0.021° on their eyes. The stimulus display was controlled by software written using PsychToolbox-3 (Brainard, 1997), running on MATLAB 2020.

An EyeLink 1000 Tower mounted eye tracker was used to monitor the left eye of each participant during each trial to prevent eye movements. The sampling rate of the tracker was 1000 Hz. A 9-point calibration procedure was used at the beginning and repeated periodically to ensure the validity of gaze measurements.

Convolutional neural networks were trained on an NVIDIA Tesla V100-SXM2 GPU with 16 GB ram.

4. Experiment Design

We used a gaze-contingent adaptive staircase procedure to measure the 70% contrast threshold for participants on a 1-in-5 match to sample task. For each observer, we estimated 4 (aperture sizes) x 2 (upright/inverted face) x 4 (repeats) = 32 contrast thresholds. To measure each contrast threshold, we used QUEST+ (Watson, 2017). The details of the fitting method are described in the Procedure section. We estimated one contrast threshold per block, which consisted of 64-128 trials based on the participant's responses. The actual experiment also used an additional watches condition (not reported in this chapter). Thus, participants completed a total of 48 blocks. The block order was randomized across participants. The experiment was completed across several sessions, usually over 7-14 days.

5. Trial Design

To begin a trial, participants maintained fixation on a central cross and hit a key to indicate readiness. After the keypress, the program checked whether the participant maintained their fixation on the cross for a variable period of 500-1500 msec. We used a variable period to prevent the participant from anticipating stimulus onset time. If the observer's gaze position drifted more than 1° from the fixation cross during this interval, the trial was aborted and restarted. If the participant successfully maintained fixation for the variable delay period, the cross disappeared, and a noisy contrast reduced stimulus consisting of either an upright or inverted face as viewed through an aperture of one of four possible sizes depending on the condition appeared at the center of the screen and stayed on for 200 msec. The face was chosen randomly from the set of five faces used for the experiment. During this time, observers were instructed to avoid making eye movements. After the stimulus disappeared, a Gaussian white noise mask of matched mean luminance and standard deviation of ~6.4 cd/m^2 was displayed for 500 msec. This was followed by a response screen containing noise-free full contrast versions of all the five faces as viewed from the same aperture as the stimulus. The response screen stayed on until the participant indicated which stimulus they saw through a mouse click. After the response, the experiment progressed to the next trial. No feedback was given. A schematic depicting the events within each trial is shown in Figure 18.

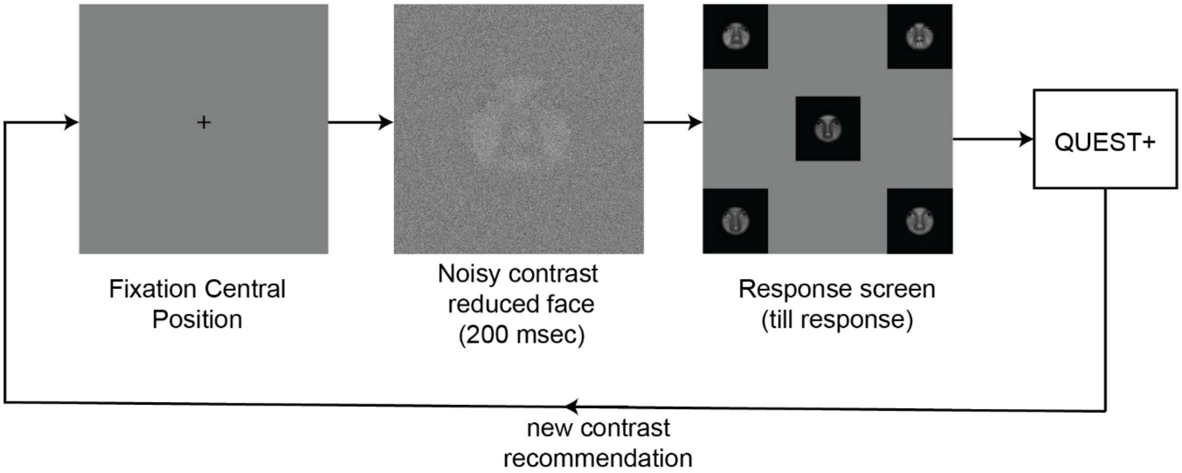

+
Fixation Central Position
Noisy contrast reduced face (200 msec)
Response screen (till response)
QUEST+
new contrast recommendation

Figure 18. Trial schematic for the spatial integration task. Observers initiated the trial by fixating at a cross and pressing a key. After a variable delay of 1000-1500 msec, a noisy contrast reduced face as viewed through an aperture was flashed for 200 msec. Then a response screen was shown containing full contrast faces as viewed through the same aperture. The contrast was adjusted using a staircase procedure (QUEST+) over 64-128 trials to yield an estimate of the contrast threshold for the condition.

6. Procedure

The experiment was administered by trained graduate or undergraduate researchers in accordance with protocols approved by the Institutional Review Board (IRB) of the University of California, Santa Barbara. The main task consisted of 48 blocks to be completed across multiple sessions. Each block took between 8-20 minutes to complete, based on the speed and accuracy of the participant. At the beginning of each session, after the participant was seated comfortably, an initial eye tracker calibration was performed, and then the participant started the task. The eye tracker was recalibrated whenever there were too many aborted trials due to broken fixations. The participant was encouraged to take a break between blocks to prevent eye discomfort and were not allowed to spend more than 1.5 hours per session to avoid the effects of fatigue.

a) Adaptive Staircase.

To measure the contrast threshold, an adaptive staircase procedure based on QUEST+(see Watson, 2017 for the full derivation) was used. The implementation of the algorithm was adopted from (Jones 2018). For each participant, we assumed the psychometric function ψ based on a Weibull CDF (Watson & Pelli, 1983):

$$\psi(x,a,\beta)=\gamma+(1-\gamma-\lambda)1-e^{\left(\frac{x}{a}\right)^{\beta}} \quad \dots \ (2)$$

where x is the multiplicative stimulus contrast, a and β are free parameters fit to each observer, sampled from [0.002,1], and [1,4], in discrete steps of 0.002 and 0.1, respectively. γ and λ were the chance performance (20%) and the expected mechanical error rate (5%), respectively. Initially, we assumed a uniform prior probability across all possible (a, β). On each block, we started with an initial guess of $x = 0.2$. Based on the response of the participant, the probability of each possible outcome r (correct or incorrect response) on the next trial is computed. The joint posterior probability density of the parameters a and β is also updated using the Bayes' Theorem applied with the prior distribution and the likelihood of seeing the responses till that trial. The expected entropy of the outcome r at a chosen stimulus x is then estimated for each possible value of x. The value of x that minimizes this entropy is chosen for displaying the next trial. As the experiment progresses, the estimates of a and β improve. The minimum number of trials in each staircase procedure was set to 64. The adaptive staircase process was terminated either at a maximum of 128 trials or when the entropy of the outcome r was less than 8, whichever happened first. This choice was made based on pilot experiments to determine a criterion to have a sufficiently accurate estimate of the contrast threshold while keeping the experiment short.

7. Analysis and Modelling

a) Determining the contrast threshold

The 70% contrast threshold was computed for each observer by solving equation (1) by setting ψ $(x, a, \beta) = 0.70$ and substituting the estimates of a and β obtained from the data from each adaptive staircase.

b) Bayesian Ideal Observer (BIO)

Ideal observers are hypothetical devices that are designed to perform optimally in a perceptual task given task-relevant information. Ideal observers have been used in many ways in vision research (see Geisler, 2011 for a review). Here, we used the BIO as a benchmark against which we will compare humans and CNNs on the occluded face identification task. We present a conceptual derivation of the BIO here.

In this task, there is no variation in the signal location within a block. Each stimulus was 500x500 pixels in size. The ideal observer is provided with exact templates of these stimuli. On each trial, an independently drawn sample of white Gaussian noise is added to the stimulus. This variation causes statistical variation in the stimulus across trials which limits the performance of the ideal observer. The decision that the ideal observer makes is determined using the maximum likelihood principle applied to the five possible responses for each trial (see the derivation below). The average performance is assessed by simulating 10,000 such trials and computing the fraction of correct responses.

c) Convolutional Neural Networks

Convolutional neural networks can be treated as model observers that are loosely based on the architecture of the visual system in animals. These models take in an image input and iteratively apply a set of convolution filters to the image, rectify the outputs, and then pool the resulting “feature maps” to generate responses that may be considered similar to the responses of cells in the visual pathway (Lindsay, 2021). After some number of iterations of the above steps, the output feature map is passed through fully connected layers that learn the weights to be assigned to various features for a given image. The final fully connected layer consists of as many units as there are categories, which allows the network to output a category label for an image. In CNNs, as successive iterations of convolution and pooling are applied, the size of the theoretical receptive field of the units in the deeper layers becomes larger(Luo et al., 2016). In other words, each unit’s response is affected by a larger patch of the image. We use this property of CNNs to manipulate the region of spatial summation allowed within the model.

(1) Architecture.

We aimed a create a set of CNNs with progressively increasing receptive field (RF) sizes that mimic RFs in humans. For this, we used 5 CNNs. We will refer to them as CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5. The CNN architectures are shown in figure 22. CNN-4 was the deepest architecture with four convolutional layers. CNNs 5, 4, 2, and 1 were progressively shallower neural networks, each created by removing a pair of convolutional and max-pooling layers from the previous CNN. CNN-3 was created by changing the stride of the first convolutional layer from 3 to 2. Figure 19 shows the schematics of these CNN architectures. In each network, the extracted features after the cascade of convolutional and max pooling operations we flattened and connected with a dense layer with 32 units. Up to this layer, we used the ‘ReLU’ activation function for each layer. The dense layer then further connected with a final 5-unit that used a softmax activation function to return a categorical output reflecting the decision of the CNN. The receptive fields of the four CNNs were 0.80° (CNN1), 1.95° (CNN 2), 2.8° (CNN 3), 4.22° (CNN 4), and 10.26° (CNN 5) in size respectively.

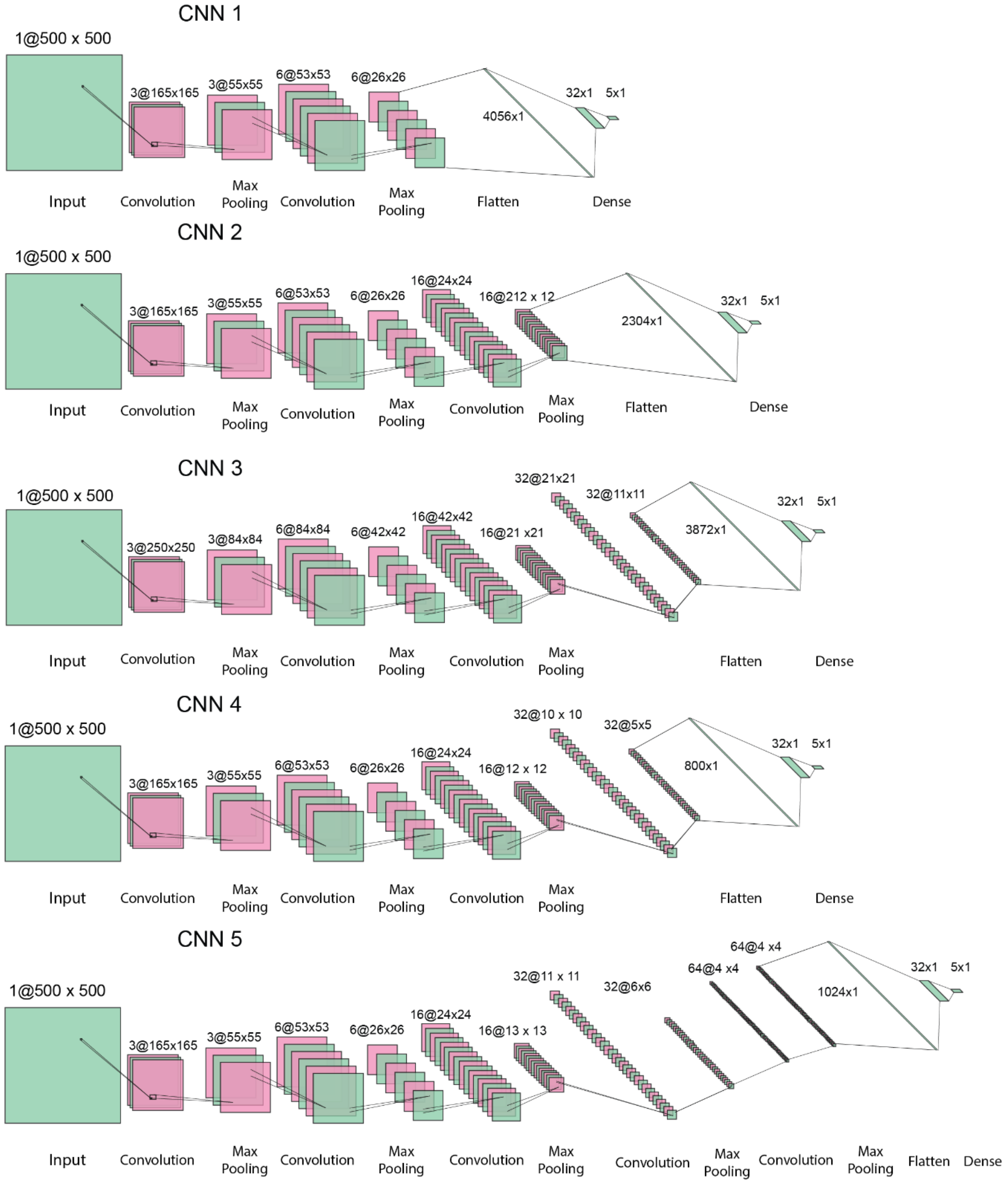


Figure 19. This figure shows the architectures for different neural networks used in this chapter. Architectures 1, 2, 4 and 5 were created by successively stacking convolutional and max-pooling layers. CNN 3 had a similar architecture, except that it had a lower stride in the first convolutional layer, resulting in an intermediate RF size between CNNs 2 and 4

(2) Receptive Field Size Estimation.

Receptive field (RF) estimation was done using the technique of gradient-based RF estimation with the help of code developed by Krzysztof Kolasinski, available free under GPL-3 License.

(3) Training and Testing Datasets.

We created training datasets by adding Gaussian noise equivalent to the noise level in the stimuli human and ideal observers were shown. The primary training of the CNNs happened on a 7,500-sample dataset of either full (without aperture) upright or full inverted faces. The dataset contained 1,500 samples of each of the five faces used in this study. xA separate dataset with 600 images/class was used for validation. The faces were generated by adding white Gaussian noise with zero mean and standard deviation equal to 14 gray levels. Since we needed to estimate the neural network's contrast threshold, we used different levels of multiplicative contrast uniformly sampled from the range 0.002 to 1 in discrete steps of 0.002. This step allowed the units to learn features in full-face images relevant for face identification at various contrasts, like how humans perform the task (Geirhos et al., 2018). For transfer learning, we created smaller datasets with 300 images/class for each face condition (upright/inverted) and each aperture (4 levels). These retraining datasets also used varying levels of contrast like the primary training dataset. The validation dataset consisted of 150 images/class. For testing, we created a dataset that consisted of noise-added faces at full contrast consisting of 300 images/class. The actual contrast of the images was manipulated by multiplication with a scalar while fitting the contrast threshold (see details of fitting below).

(4) Training.

All CNNs were trained on 7500 noisy samples of full (without aperture) upright and inverted faces separately. We used the Adam optimizer with an initial learning rate of 0.001, trained for a maximum of 100 epochs, stopping earlier if the validation loss did not reduce after 20 epochs. We used a dropout layer that randomly dropped 20% of the nodes while training to improve training robustness.

(5) Transfer Learning.

Even though the preponderance of human experience to faces consists of full faces, we can flexibly use this learning on faces covered with apertures. A face stimulus covered by an aperture activates neurons across different stages of visual processing based on their feature tuning. These activations form the basis for the decision made by humans. However, this property does not automatically translate the CNNs. If a CNN is trained on full faces and tested on a face covered with apertures, it is most likely to perform at chance level. To model the human ability to flexibly use stimulus features at different stages of visual processing to make identity judgments, we need to limit the CNN's ability to use feature maps learned during training on full faces and retrain it to perform categorization on stimuli masked by apertures. For this, we used the technique of transfer learning. For this, we created a fresh CNN with the architecture chosen from one of the 4 CNNs we are studying. We copied the weights of the corresponding layer from the largest network (CNN 5) into this architecture trained on upright or inverted full faces. For the case of CNN 3, we copied weights from the full architecture trained on full faces. Then the weights of all the copied layers were frozen. The dense layer was initiated randomly. This network was then retrained separately on each condition (upright/ inverted faces) x (4 apertures). All other hyperparameters were set identical to the primary training paradigm.

d) Fitting the contrast thresholds

We fit 70% contrast thresholds for the ideal observer and the CNN. The contrast threshold for the ideal observer and the CNNs were determined by a staircase procedure. The procedure started with an initial guess for the contrast threshold. The testing of the model observer was done with the contrast reduced multiplicatively by the guess contrast. If the ideal observer's performance was greater than 70%, then guess contrast was reduced by half, and if it was lower than 70%, it was increased by half of its distance to the upper threshold, which was set to be 1. Based on the outcome, the lower or upper threshold was updated to the previous guess threshold, and the algorithm progressed to the next step. This procedure was repeated until the performance of the model observer eventually converged to 70% with a permissible difference threshold (0.5% for the ideal observer and 2% for the CNN). The final guess contrast was taken as the contrast threshold for the condition.

In the inverted face condition, CNN 1 reached a peak performance of only 62%. For this condition, to compare the performance profile of CNN 2 with other CNNs, we used a 50% contrast threshold instead.

e) The efficiency of an observer or model

To benchmark the performance of a human or model observer, we use a measure known as efficiency (Barlow, 1980), which is the ratio of the squared contrast thresholds of the ideal and human observers for a chosen performance level.

$$Efficiency = \eta = \frac{c_{ideal}^2}{c_{observer}^2} \quad (2)$$

Since we estimated the contrast thresholds for the human and model observers at 70% accuracy, we performed a binary search to compute the contrast threshold at which the BIO performed at 70% and used that contrast threshold in equation (2) to compute the efficiency of the human and CNN observers. The efficiency always lies between zero and 1. A higher efficiency indicates that the observer is able to use a larger fraction of the stimulus information to perform the task.

8. Results.

a) Results for human observers.

To establish the variation of human efficiency with aperture size and face orientation, we conducted a 2-way ANOVA with aperture (4 levels) and *orientation (*upright/inverted) as factors. The results indicated a main effect of aperture size (*F*(3,248) = 83.06, *p* <<0.05), and a main effect of face *orientation* (*F* (1,248) = 92.69, *p* << 0.05). The interaction effect was also significant (*F* (3,248) = 28.19, *p* << 0.05). Post-hoc Tukey's HSD contrasts revealed that the efficiencies of human observers were not significantly different for upright and inverted faces for the first and second aperture sizes. Crucially, human efficiency for upright faces was significantly higher than inverted faces for larger apertures. We also found that human efficiency for inverted faces increases slightly but significantly for larger apertures (*p* = 0.005, see figure 20).

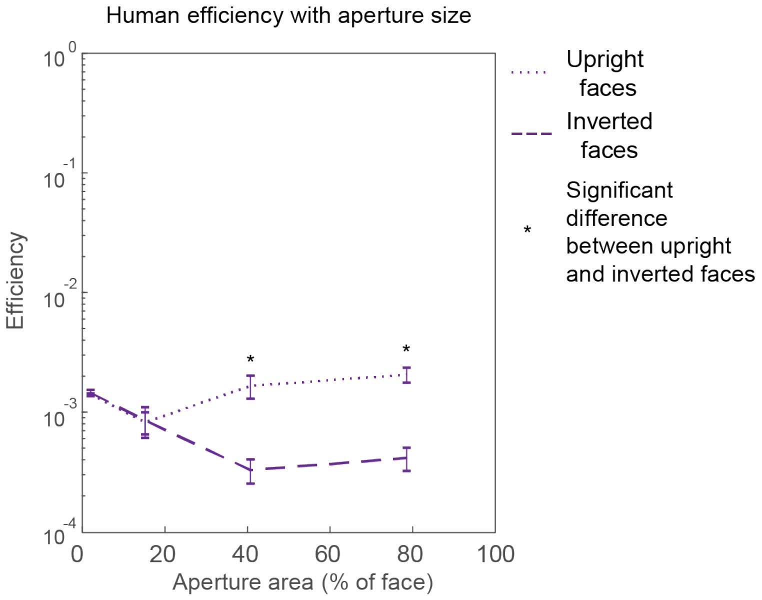


Figure 20. The variation of spatial integration efficiency with aperture size for upright and inverted faces. The efficiency of humans initially reduces with aperture size for very small apertures. However, as the apertures become larger, human efficiency for upright faces increases while that of inverted faces continues to reduce.

b) *The role of effective receptive field size*

To understand how prolonged exposure to faces spatial summation in CNNs, we plotted the efficiency of the CNNs that were trained on full upright faces as a function of the mask area fraction (Figure 21). Our results show the efficiencies of CNNs 1, 2, and 3 for upright and inverted faces are approximately equal with increasing aperture size. CNN 4 shows poor spatial summation for the smaller apertures, but as the aperture size increases, its spatial summation capability for upright faces beats that for inverted faces. CNN 5 shows similar efficiency for upright and inverted faces for smaller apertures but shows superior integration efficiency for upright rather than inverted faces as the aperture size increases.

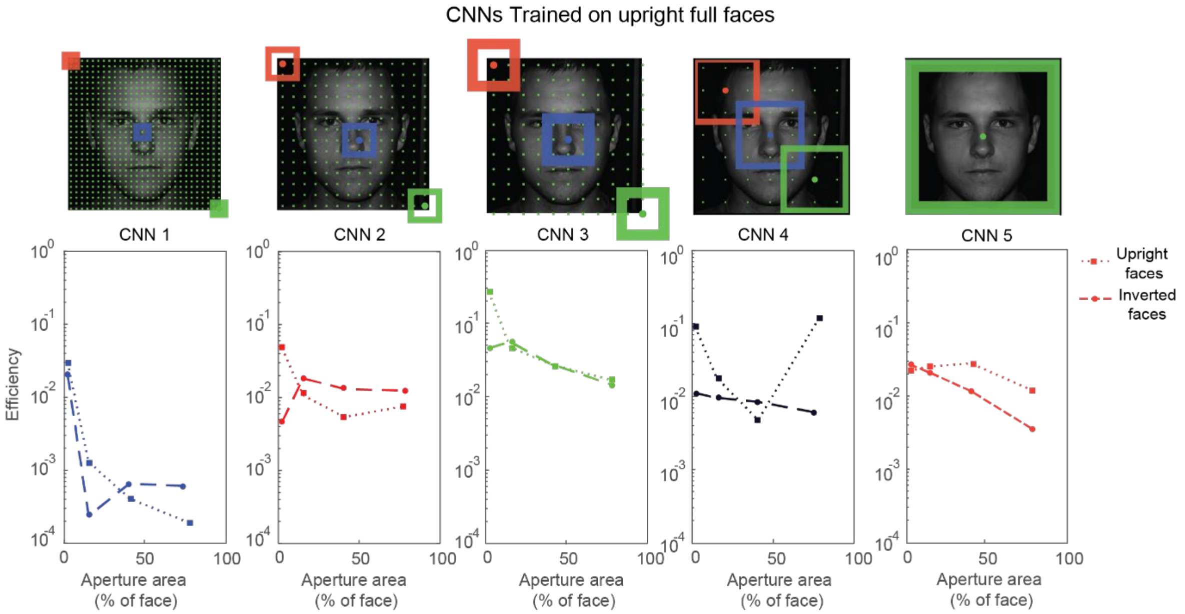


Figure 21. These charts show the efficiency of the 4 CNNs in various conditions. The red, blue, and green squares simply represent three placements of the receptive field overlaid on the image as an aid to visualize the spatial extent of the RF at different locations on the image. The top row shows the sizes of the RFs for these CNNs. The plots show the efficiency of CNNs trained on full upright faces on a task of categorizing upright or inverted faces viewed through apertures of increasing sizes. Blue, red, green, black and orange traces represent CNNs 1, 2, 3, 4, and 5 respectively. The dotted and dashed lines represent testing on upright and inverted faces, respectively.

9. Discussion

We implemented a spatial summation task on upright and inverted faces to understand how possible computational mechanisms behind the Face Inversion Effect (FIE). Specifically, we aimed to understand why *holistic processing,* i.e., a mode of information processing that considers all the parts together as a unit, occurs only for upright but not inverted faces. Possible mechanisms behind the face inversion effect have been proposed from a variety of approaches like developmental psychology (Cashon & Holt, 2015); psychophysics, (Diamond & Carey, 1989; Yin, 1969), computational modeling (Loftus et al., 2004); neuroscience (Yovel & Kanwisher, 2005), cognitive science (Rakover, 2013; Rossion, 2009), and computational neuroscience (Wallis, 2013). When we consider all the proposed theories, three factors are dominant. Firstly, the inversion effect has been linked to overtraining for upright orientation of faces, but not for inverted (Wallis, 2013). Secondly, it has been linked to using a larger *perceptual field* in the upright orientation compared to inverted (Rossion, 2009). Finally, the inversion effect has been connected to higher-level face-sensitive regions (primarily the Fusiform face area, or FFA) in the ventral cortex (Yovel & Kanwisher, 2005). Over the last 20 years, various novel techniques to measure these factors have been developed and used to test these theories. Here, we provide evidence linking these factors to the face inversion effect by applying two new techniques that have never been applied together in this context: spatial efficiency measurements and deep learning.

Our results indicate that humans can integrate information across a larger spatial extent for upright compared to inverted faces. This result is broadly in agreement with the *perceptual field* hypothesis. Our results are also consistent with the findings of Tyler & Chen (2006), who reported a larger region of spatial summation for upright vs. inverted faces on a detection task. Their study was based only on four individuals, and the effects did not reach statistical significance. We demonstrated a much larger effect size (f = 0.61) of inversion on the efficiency for face identification at a sample size of 8 individuals. This indicates that face identification (but not face detection) elicits stronger spatial summation in upright faces than inverted faces.

Recent research has shown that the population receptive fields (pRFs) of higher-level face-sensitive areas like the FFA in the ventral stream shrink in response to the face inversion. This results in reduced spatial integration of features in these areas, ultimately manifesting behaviorally as the face inversion effect (Poltoratski et al., 2021). While this result elucidates the neural mechanism underlying the face inversion effect, it does not reveal the connection to the third factor listed above: the role of long-term visual experience in shaping this mechanism.

We demonstrated this connection by using five Convolutional Neural Networks (CNN) trained on upright and inverted faces and testing their efficiency at the spatial summation task. CNN 1, with a small receptive field, showed reducing the efficiency with increasing aperture size irrespective of its training. This result is consistent with the idea that lower visual areas do not mediate the face inversion effect. The size of the theoretical receptive field was approximately 0.8° x 0.8°. In contrast, CNN 5 with larger receptive fields trained on full upright faces showed a similar pattern of spatial summation as in humans: the efficiency of spatial summation was similar for upright and inverted faces for smaller aperture sizes, but superior for upright rather than inverted faces for larger aperture sizes.

What aspects of upright faces allow humans (and CNNs with sufficiently large RFs) to acquire superior spatial integration capabilities? To answer this, we draw inspiration from an influential model of face processing (Haxby et al., 2000), which postulates that faces have changeable and invariant features. Examples of invariant features are feature configuration, the position of the jawline, skin tone, etc. Based on this model, prolonged exposure to faces allows the brain to optimize for invariant features making it more efficient at integrating information from these features. Since these features are spread out on the face, the neurons in higher regions with larger RFs are recruited for processing faces. These neurons show superior efficiency when the features of the incoming stimulus are aligned with their preferred features. On the other hand, they are also more inefficient at representing stimulus features that they are not tuned to. Our research provides preliminary evidence supporting the idea that visual experience mediates the face inversion effect by increasing the spatial summation capability in regions of the brain with larger receptive fields. However, more research is needed to identify the exact mechanism of this influence. One possible direction is to explore the roles of changeable and invariant features during training in bringing about superior spatial summation abilities neural network.

Bearing in mind the risk of inappropriate inference when using CNNs as models of the visual system (Borowski et al., 2019), we do not suggest that our models completely capture all elements of how experience shapes face perception. Rather, we are using this model to establish a specific claim that the preferential use of efficient spatial summation mechanisms in humans can be linked to biases in long-term visual experience. The value of such efforts for psychological science has been recognized and emphasized (Guest & Martin, 2021). Therefore, our results suggest that the face inversion effect in humans occurs as a result of superior spatial integration efficiency for upright rather than inverted faces. Our results also suggest a possible mechanism by which such superior spatial integration efficiency may arise in regions higher up in the visual processing hierarchy as a result of long-term visual experience of full upright rather than inverted faces.

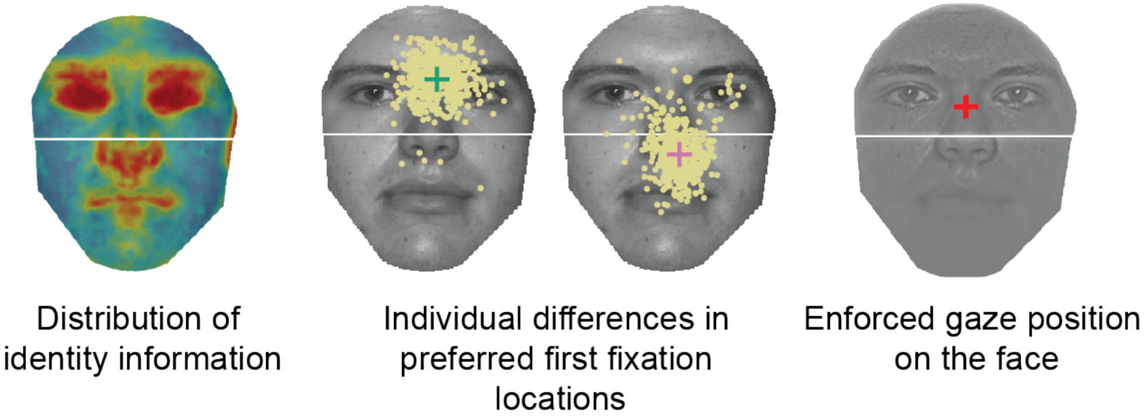
IV. General Discussion

This thesis aimed to explore the role of the visual experience of faces in shaping face perception mechanisms in humans. The emergent picture is that prolonged exposure to faces such that the features occur in a fixed configuration around an observer's preferred point of fixation on the face may allow the learning of larger feature templates. These templates are efficient in most real-life face recognition scenarios where the incoming face information is formatted to match the learned templates. However, efficient processing may break down in some contrived and unfamiliar situations, resulting in peculiar perceptual effects that appear to be unique to faces. We presented empirical evidence supporting the hypothesis that the visual experience of faces shapes these perceptual phenomena in humans and model observers.

Recent results showed that the visual experience of faces as mediated by the PFL plays a functional role in task performance across face tasks like person, gender, and emotion recognition.(Peterson & Eckstein, 2012; Tsank, 2019). Chapter II extended these results by demonstrating that the PFL also modulates two well-studied face effects: the Composite Face Effect (CFE) and the Face adaptation effect (FAE). For this, we compared groups of upper and lower-lookers on a Composite Face Effect (CFE) task and a Face adaptation effect (FAE) task. We found that the location of an individual's PFL influenced the outcomes of both tasks. Upper lookers generally showed stronger CFEs than lower lookers. The second task revealed that upper lookers were more position-specific in their adaptation effects, such that adapting away from their PFL reduced the strength of the FAE irrespective of the testing location. On the other hand, the FAE for the lower lookers was more position invariant. These results form the basis for the central argument in this thesis: long-term visual experience of faces shapes the underlying neural mechanisms of face processing, which gives rise to efficient face processing and other perceptual effects commonly reported with faces.

Here we discuss possible interpretations of these results. Over a prolonged period of viewing faces at a specific fixation location, the brain learns internal representations of faces that are tuned to that fixation location (Tsank, 2019). The first panel in Figure 22 shows the performance of a Region of Interest Ideal Observer (ROI) ideal observer (see Peterson & Eckstein, 2012 for a derivation). The heatmap represents the distribution of task-relevant information on the face, with the hotter regions representing more informative regions. The image shows that there is more information density near the eye region, which is also spread across a larger area than that near the mouth region. Thus, there is a 'U' shaped distribution of visual information on the face. A foveated observer who usually fixates higher up on the face would need to integrate information across a larger area to identify faces effectively. An upper looker would likely develop efficient visual processing over a larger span than a lower looker, which increases the interference from the irrelevant half in the CFE task. This possibly explains the reducing CFE as the PFL moves away from the eyes.

The interpretation of the effect of the PFL on the FAE is more counter-intuitive. One possible theoretical interpretation of these results is based on a well-established theory of face processing (Haxby et al., 2000). The theory postulates that some aspects of the face are more invariant than others. For example, the configuration of features on the face remains unchanged across human faces. On the other hand, several features like mouth shape, gaze direction, etc., are changeable. The brain may learn more rigid representations of invariant features around the PFL. This will result in stronger tuning to the fixation position, and consequently, higher position specificity of the FAE. On the other hand, if the features around an observer's PFL are more changeable, the learned representations may be more flexible, resulting in a weaker fixation tuning and ultimately higher position invariance. This may be a possible explanation that accounts for our findings. While we have provided empirical evidence supporting the role of the PFL on the FAE, an exact mechanism for this effect would need to be established by a more rigorous future study.



Distribution of identity information

Individual differences in preferred first fixation locations

Enforced gaze position on the face

Figure 22. Face recognition is an interplay between the distribution of stimulus distribution, long-term oculomotor strategies, and the foveated visual system. The first image depicts a heat map of face identity relevant information computed using a Region of interest Ideal observer (see Peterson & Eckstein, 2012 for implementation). The next two faces depict individual differences in preferred first fixation locations on the face. Yellow dots represent the landing position of the first eye movement to faces across 320 trials. The green and pink crosses represent the average first fixation location of an example upper and lower looker, respectively. The right-most face depicts the fixation position in a gaze-contingent sequential face part matching task. The red cross shows the location where the observer maintains their gaze throughout the stimulus presentation time.

Chapter III explored the computational basis of the face inversion effect (FIE). The FIE refers to the fact that humans are better at identifying upright rather than inverted faces. This effect is thought to arise because humans are mainly exposed to upright rather than inverted faces. Here we showed that this effect occurs due to increased efficiency of spatial integration of information for upright faces compared to inverted faces. Recent research also showed the importance of higher-level visual processing in the visual cortex in mediating the FIE (Poltoratski et al., 2021). Our results reconciled these findings by showing that a neural network model with a larger receptive field also demonstrates superior spatial integration for the learned stimulus orientation, provided the stimulus had sufficient spatial extent for the efficient spatial processing to kick in.

Chapter III establishes the role of visual experience to face differently: by using deep learning models. Convolutional neural networks are loosely based on the human visual system. However, many researchers have noted limitations in comparing the two systems (Borowski et al., 2019). However, we argue here that to characterize whether and how visual experience brings about the FIE, an exact correspondence between the visual system and the model observer is not necessary. Quoting David Marr, “In order to understand bird flight, we have to understand aerodynamics; only then the structure of the feathers and different shapes of bird’s wings makes sense” (Marr, 1982). Likewise, the CNN helps us test the hypothesis that long-term statistical learning of visual features in upright faces in a system capable of spatial summation over a sufficiently large spatial extent can efficiently process upright but not inverted faces. Our confirmation of this hypothesis thus suggests the role of visual experience in shaping the FIE at an algorithmic level. However, we cannot make any inferences about *how* this process happens in humans. Combined with recent research confirming the role of spatial summation in higher-level regions in the ventral stream in the brain (Poltoratski et al., 2021), we can make a stronger case for a spatial summation account of face processing.

Our finding that visual experience influences face perception is not novel: many authors have noted this in passing. However, our findings show that this influence is exerted even at smaller spatial scales, such as between two different fixation locations on the face. This underscores the importance of developing techniques that measure and characterize the *face diet* accurately across the human lifespan to better understand face processing mechanisms.

References

Afraz, S. R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research*, *48*(1), 42–54. https://doi.org/10.1016/J.VISRES.2007.10.028

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Van Esesn, B. C., Awwal, A. A. S., & Asari, V. K. (2018). *The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches*. http://arxiv.org/abs/1803.01164

Barlow, H. B. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 71–82. https://doi.org/10.1098/rstb.1980.0083

Borowski, J., Funke, C. M., Stosio, K., Brendel, W., Wallis, T. S. A., & Bethge, M. (2019). *The Notorious Difficulty of Comparing Human and Machine Perception*. 642–646. https://doi.org/10.32470/ccn.2019.1295-0

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*. http://bbs.bioguider.com/images/upfile/2006-4/200641014348.pdf

Braje, W. L., Kersten, D., Tarr, M. J., & Troje, N. F. (1998). Illumination effects in face recognition. *Psychobiology*, *26*(4), 371–380. https://doi.org/10.3758/BF03330623

Bruce, V., & Young, A. (2013). *Face perception*. Psychology Press.

Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). *Proceedings of the IEEE International Conference on Computer Vision*, *2017-October*, 1021–1030. https://doi.org/10.1109/ICCV.2017.116

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

1485. https://doi.org/10.1080/17470218.2013.800125

C-W Shyi, G., & Wang, C.-C. (2016). Testing Differential Holistic Processing Within a Face: No Evidence of Asymmetry from the Complete Composite Task. *Frontiers in Psychology*, *7*, 1506. https://doi.org/10.3389/fpsyg.2016.01506

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *Journal of Cognitive Neuroscience*, *17*(10), 1652–1666. https://doi.org/10.1162/089892905774597254

Cashon, C. H., & Holt, N. A. (2015). Developmental Origins of the Face Inversion Effect. In *Advances in Child Development and Behavior* (Vol. 48, pp. 117–150). https://doi.org/10.1016/bs.acdb.2014.11.008

Cassia, V. M., Picozzi, M., Kuefner, D., Bricolo, E., & Turati, C. (2009). Holistic processing for faces and cars in preschool-aged children and adults: evidence from the composite effect. *Developmental Science*, *12*(2), 236–248. https://doi.org/10.1111/j.1467-7687.2008.00765.x

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, Race, Age and Beauty as Factors in Recognition of Faces. *Perception*, *10*(6).

Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E., & Kalina, R. E. (1987). Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science*, *236*(4801), 579–582.

Cziraki, C., Greenlee, M. W., & Kovács, G. (2010). Neural correlates of high-level adaptation-related aftereffects. *Journal of Neurophysiology*, *103*(3), 1410–1417.

https://doi.org/10.1152/JN.00582.2009/ASSET/IMAGES/LARGE/Z9K0031099810004.JPEG

Dakin, S. C., & Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, *9*(4), 2–2. https://doi.org/10.1167/9.4.2

Diamond, R., & Carey, S. (1986). Why faces Are and Are Not Special: An Effect of Expertise. *Journal of Experimental Psychology: General*, *115*(2), 107–117. https://doi.org/10.14890/minkennewseries.64.2_258

Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *Journal of Neuroscience*, *35*(6), 2791–2802. https://doi.org/10.1523/JNEUROSCI.4037-14.2015

Fernando, D., Favelle, S., Wickramariyaratne, T., Crookes, K., Davies, A. A., McKone, E., Broughton, M., Zappia, S., Darke, H., & Fiorentini, C. (2013). Importance of the Inverted Control in Measuring Holistic Face Processing with the Composite Effect and Part-Whole Effect. *Frontiers in Psychology*, *4*(February), 1–21. https://doi.org/10.3389/fpsyg.2013.00033

Furl, N., van Rijsbergen, N. J., Treves, A., & Dolan, R. J. (2007). Face adaptation aftereffects reveal anterior medial temporal cortex role in high level category representation. *NeuroImage*, *37*(1), 300–310. https://doi.org/10.1016/J.NEUROIMAGE.2007.04.057

Gauthier, I. (2020). What We Could Learn About Holistic Face Processing Only From Nonface Objects. *Current Directions in Psychological Science*, *29*(4), 419–425. https://doi.org/10.1177/0963721420920620

Gauthier, I., & Tarr, M. J. (1997). *Becoming a "Greeble" Expert: Exploring Mechanisms* *for Face Recognition* (Vol. 37, Issue 12). https://ac.els-cdn.com/S0042698996002866/1-s2.0-S0042698996002866-main.pdf?_tid=8344c232-c475-4215-adb4-160b5bb7334d&acdnat=1552173859_f0deb1bc84a7105e1a3a4a12b1d3afbd

Gauthier, I., Wong, A. C. N., & Palmeri, T. J. (2010). Manipulating Visual Experience: Comment on Op de Beeck and Baker. *Trends in Cognitive Sciences*, *14*(6), 235. https://doi.org/10.1016/J.TICS.2010.03.009

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *7th International Conference on Learning Representations, ICLR 2019*. https://arxiv.org/abs/1811.12231v2

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781. https://doi.org/10.1016/j.visres.2010.09.027

Ghazi, M. M., & Ekenel, H. K. (2016). A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 102–109. https://doi.org/10.1109/CVPRW.2016.20

Goffaux, V., & Rossion, B. (2007). Face Inversion Disproportionately Impairs the Perception of Vertical but not Horizontal Relations Between Features. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 995–1002. https://doi.org/10.1037/0096-1523.33.4.995

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The Perception of a Face Is No More Than the Sum of Its Parts. *Psychological Science*, *23*(4), 427–434.

https://doi.org/10.1177/0956797611427407

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*(1–3), 293–321. https://doi.org/10.1016/S0001-6918(01)00019-1

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585

Harrison, S. A., Gauthier, I., Hayward, W. G., & Richler, J. J. (2014). Other-race effects manifest in overall performance, not qualitative processing style. *Visual Cognition*, *22*(6), 843–864. https://doi.org/10.1080/13506285.2014.918912

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://doi.org/10.1016/S1364-6613(00)01482-0

Heering, A., Rossion, B., Turati, C., & Simion, F. (2008). Holistic face processing can be independent of gaze behaviour: Evidence from the composite face illusion. *Journal of Neuropsychology*, *2*(1), 183–195. https://doi.org/10.1348/174866407X251694

Heinke, D., & Mavritsaki, E. (2009). *Computational modelling in behavioural neuroscience: closing the gap between neurophysiology and behaviour* (Vol. 2). Psychology Press.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: data analysis with estimation graphics. *Nature Methods 2019 16:7*, *16*(7), 565–566. https://doi.org/10.1038/s41592-019-0470-3

Horry, R., Cheong, W., & Brewer, N. (2015). The other-race effect in perception and recognition: Insights from the complete composite task. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 508–524. https://doi.org/10.1037/xhp0000042

Hsiao, J. H., An, J., Zheng, Y., & Chan, A. B. (2021). Do portrait artists have enhanced face processing abilities? Evidence from hidden Markov modeling of eye movements. *Cognition*, *211*, 104616. https://doi.org/10.1016/J.COGNITION.2021.104616

Issa, E. B., & DiCarlo, J. J. (2012). Precedence of the Eye Region in Neural Processing of Faces. *Journal of Neuroscience*, *32*(47), 16666–16682. https://doi.org/10.1523/jneurosci.2391-12.2012

Itier, R. J., Alain, C., Sedore, K., & McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes! *Journal of Cognitive Neuroscience*, *19*(11), 1815–1826. https://doi.org/10.1162/jocn.2007.19.11.1815

Jiang, F., Blanz, V., & Rossion, B. (2011). Holistic processing of shape cues in face identification: Evidence from face inversion, composite faces, and acquired prosopagnosia. *Visual Cognition*, *19*(8), 1003–1034. https://doi.org/10.1080/13506285.2011.604360

Jones, P. R. (2018). QuestPlus: A Matlab Implementation of the QUEST+ adaptive Psychometric Method. *Journal of Open Research Software*, *6*, 1–5. https://doi.org/10.5334/jors.195

Kaiser, D., Walther, C., Schweinberger, S. R., & Kovács, G. (2013). Dissociating the neural bases of repetition-priming and adaptation in the human brain for faces. *Journal of Neurophysiology*, *110*(12), 2727–2738. https://doi.org/10.1152/JN.00277.2013/ASSET/IMAGES/LARGE/Z9K024132228000

4.JPEG

Kanwisher, N. (2017). The Quest for the FFA and Where It Led. *The Journal of Neuroscience*, *37*(5), 1056–1061. https://doi.org/10.1523/JNEUROSCI.1706-16.2016

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

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Kovács, G., Cziraki, C., Vidnyánszky, Z., Schweinberger, S. R., & Greenlee, M. W. (2008). Position-specific and position-invariant face aftereffects reflect the adaptation of different cortical areas. *NeuroImage*, *43*(1), 156–164. https://doi.org/10.1016/J.NEUROIMAGE.2008.06.042

Kovács, G., Zimmer, M., Harza, I., & Vidnyánszky, Z. (2007). Adaptation duration affects the spatial selectivity of facial aftereffects. *Vision Research*, *47*(25), 3141–3149. https://doi.org/10.1016/j.visres.2007.08.019

Kowalski, M., Naruniec, J., & Trzcinski, T. (2017). Deep Alignment Network: A Convolutional Neural Network for Robust Face Alignment. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2017-July*, 2034–2043. https://doi.org/10.1109/CVPRW.2017.254

Kurbel, D., Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2021). The composite face effect is robust against perceptual misfit. *Attention, Perception, and Psychophysics*, *83*(6), 2599–2612. https://doi.org/10.3758/s13414-021-02279-0

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94. https://doi.org/10.1038/82947

Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of Cognitive Neuroscience*, *33*(10), 2017–2031. https://doi.org/10.1162/jocn_a_01544

Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review*, *111*(4), 835–863. https://doi.org/10.1037/0033-295X.111.4.835

Luo, W., Li, Y., Urtasun, R., & Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *Nips*, 4905–4913.

Marr, D. (1982). Vision: A Computational Investigation of Visual Representation in Man. *Phenomenology and the Cognitive Sciences*, *8*(4), 397. https://mitpress.mit.edu/books/vision

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision*, *14*(7), 6. https://doi.org/10.1167/14.7.6

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*(1), 55–73. https://doi.org/10.1080/13506280500158761

Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic Processing Is

Finely Tuned for Faces of One ' s Own Race. *Psychological Science*, *17*(7), 608–615.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144.

Morton, J., & Johnson, M. H. (1991). Gesichtswahrnehmung bei Kindern_ Morton und Johnson.pdf. *Psychological Review*, *98*(2), 164–181.

Murphy, J., Gray, K. L. H., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin and Review*, *24*(2), 245–261. https://doi.org/10.3758/s13423-016-1131-5

Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in Cognitive Sciences*, *14*(1), 22–30. https://doi.org/10.1016/J.TICS.2009.11.002

Orban de Xivry, J. J., Ramon, M., Lefèvre, P., & Rossion, B. (2008). Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *Journal of Neuropsychology*, *2*(Pt 1), 245–268. https://doi.org/10.1348/174866407X260199

Oruc, I., Shafai, F., Murthy, S., Lages, P., & Ton, T. (2019). The adult face-diet: A naturalistic observation study. *Vision Research*, *157*, 222–229. https://doi.org/10.1016/j.visres.2018.01.001

Pernet, H. (2006). *Ritual masks: deceptions and revelations*. Wipf and Stock Publishers.

Peterson, M.F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), E3314–E3323. https://doi.org/10.1073/pnas.1214269109

Peterson, M.F., & Eckstein, M. P. (2013). Individual difference in eye movements during face identification reflect observer specific optimal points of fixation. *Psychological Science*, *24*(7), 1216–1225.

Peterson, Matthew F, Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual Differences in

Face Looking Behavior Generalize from the Lab to the World. *Journal of Vision*, *16*(2016), 1–18. https://doi.org/10.1167/16.7.12.doi

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology*, *3*(DEC), 1–11. https://doi.org/10.3389/fpsyg.2012.00559

Poltoratski, S., Kay, K., Finzi, D., & Grill-Spector, K. (2021). Holistic face recognition is an emergent phenomenon of spatial processing in face-selective regions. *Nature Communications 2021 12:1*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-24806-1

Quadflieg, S., Todorov, A., Laguesse, R., & Rossion, B. (2012). Normal face-based judgements of social characteristics despite severely impaired holistic face processing. *Visual Cognition*, *20*(8), 865–882. https://doi.org/10.1080/13506285.2012.707155

Rakover, S. S. (2013). Explaining the face-inversion effect: the face-scheme incompatibility (FSI) model. *Psychonomic Bulletin & Review*, *20*, 665–692. https://doi.org/10.3758/s13423-013-0388-1

Ramon, M., Busigny, T., & Rossion, B. (2010). Impaired holistic processing of unfamiliar individual faces in acquired prosopagnosia. *Neuropsychologia*, *48*(4), 933–944. https://doi.org/10.1016/j.neuropsychologia.2009.11.014

Rezlescu, C., Susilo, T., & Caramazza, A. (2017). The Inversion, Part-Whole, and Composite Effects Reflect Distinct Perceptual Mechanisms. *Journal of Experimental Psychology Human Perception & Performance*, *43*(12), 1961. https://doi.org/10.1037/xhp0000400

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*(5), 1281–1302. https://doi.org/10.1037/a0037004

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic Processing of Faces: Perceptual and Decisional Components. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(2), 328–342. https://doi.org/10.1037/0278-7393.34.2.328

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology*, *3*(DEC), 1–6. https://doi.org/10.3389/fpsyg.2012.00553

Rivers, C. (1994). *Face value: Physiognomical thought and the legible body in Marivaux, Lavater, Balzac, Gautier, and Zola*. Univ of Wisconsin Press.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79. https://doi.org/10.1016/j.cognition.2006.02.008

Ross, D. A., & Gauthier, I. (2015). Holistic processing in the composite task depends on face size. *Visual Cognition*, *23*(5), 533–545. https://doi.org/10.1080/13506285.2015.1049678

Rossion, B. (2008). *Picture-plane inversion leads to qualitative changes of face perception*. *128*, 274–289. https://doi.org/10.1016/j.actpsy.2008.02.003

Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica*, *132*(3), 300–312. https://doi.org/10.1016/j.actpsy.2009.08.002

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*(2), 139–253. https://doi.org/10.1080/13506285.2013.772929

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2004). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience 2004 8:1*, *8*(1), 107–113. https://doi.org/10.1038/nn1370

Stacchi, L., Ramon, M., Lao, J., & Caldara, R. (2019). *Neural Representations of Faces Are Tuned to Eye Movements*. https://doi.org/10.1523/JNEUROSCI.2968-18.2019

Sun, X., Wu, P., & Hoi, S. C. H. (2018). Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing*, *299*, 42–50. https://doi.org/10.1016/j.neucom.2018.03.030

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46A*(2), 225–245. https://doi.org/10.1080/14640749308401045

Taubert, J., & Alais, D. (2009). The composite illusion requires composite face stimuli to be biologically plausible. *Vision Research*, *49*(14), 1877–1885. https://doi.org/10.1016/j.visres.2009.04.025

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*.

Tsank, Y. (2019). *Face Perception: The Interaction of Eye Movements with Internal Face Representations*. University of California, Santa Barbara.

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L. (2021). Ffa and ofa encode distinct types of face identity information. *Journal of Neuroscience*, *41*(9), 1952–1969. https://doi.org/10.1523/JNEUROSCI.1449-20.2020

Tyler, C. W., & Chen, C. C. (2006). Spatial summation of face information. *Journal of Vision*, *6*(10), 1117–11. https://doi.org/10.1167/6.10.11

Van Belle, G., Busigny, T., Lefèvre, P., Joubert, S., Felician, O., Gentile, F., & Rossion, B.

(2011). Impairment of holistic face perception following right occipito-temporal damage in prosopagnosia: Converging evidence from gaze-contingency. *Neuropsychologia*, *49*(11), 3145–3150. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2011.07.010

Wallis, G. (2013). Toward a unified model of face and object recognition in the human visual system. *Frontiers in Psychology*, *4*(AUG), 1–25. https://doi.org/10.3389/fpsyg.2013.00497

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3), 10–10. https://doi.org/10.1167/17.3.10

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.

Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science*, *1*, 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(April), 357–360. https://doi.org/10.1038/nature02361.1.

Wichmann, F. A., & Hill, N. J. (2001). *The psychometric function : I . Fitting , sampling , and goodness of fit*. *63*(8), 1293–1313.

Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the Cortical Hierarchy: Low-Level Curve Adaptation Affects High-Level Facial-Expression Judgments. *Journal of Neuroscience*, *28*(13), 3374–3383. https://doi.org/10.1523/JNEUROSCI.0182-08.2008

Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): Image or person? *Vision Research*, *49*(23), 2800–2807.

https://doi.org/10.1016/J.VISRES.2009.08.021

Yang, N., Shafai, F., & Oruc, I. (2014). Size determines whether specialized expert processes are engaged for recognition of faces. *Journal of Vision*, *14*(8), 1–12. https://doi.org/10.1167/14.8.17

Yin. (1969). Looking At Upside-Down Faces. *Journal of Experimental Psychology*, *81*(1), 141–145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747–759. https://doi.org/10.1068/p160747n

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, *15*(24), 2256–2262. https://doi.org/10.1016/j.cub.2005.10.072

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, *8*(August), 1–9. https://doi.org/10.3389/fnhum.2014.00562

Zebrowitz, L. A. (2018). *Reading faces: Window to the soul?* Routledge.

Zimmer, M., & Kovács, G. (2011). Position specificity of adaptation-related face aftereffects. *Philosophical Transactions of the Royal Society B, Biological Sciences*, *366*, 586–595. https://doi.org/10.1098/rstb.2010.0265