

**DEVELOPING A NON-HUMAN PRIMATE MODEL TO DISSECT THE
NEURAL MECHANISMS OF FACIAL EMOTION PROCESSING**

SHIRIN TAGHIAN ALAMOOTI

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTERS OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

AUGUST 2024

© Shirin Taghian, 2024

Abstract

Facial emotion recognition is a cornerstone of social cognition, vital for interpreting social cues and fostering communication. Despite extensive research in human subjects, the neural mechanisms underlying this process remain incompletely understood. This thesis investigates these mechanisms using a non-human primate model to provide deeper insights into the neural circuitry involved in facial emotion processing. We embarked on a comparative analysis of facial emotion recognition between humans and rhesus macaques. Using a carefully curated set of facial expression images from the Montreal Set of Facial Displays of Emotion (MSFDE), we designed a series of binary emotion discrimination tasks. Our innovative approach involved detailed behavioral metrics that revealed significant parallels in emotion recognition patterns between the two species. These findings highlight the macaques' potential as a robust model for studying human-like facial emotion recognition. Building on these behavioral insights, the second phase of our research delved into the neural underpinnings of this cognitive process. We conducted large-scale, chronic multi-electrode recordings in the inferior temporal (IT) cortex of rhesus macaques. By mapping the neural activity associated with the classification of different facial emotions, we uncovered specific neural markers that correlate strongly with behavioral performance. These neural signatures provide compelling evidence for the role of the IT cortex in processing complex emotional cues. Our findings bridge the gap between behavioral and neural perspectives on facial emotion recognition, offering a comprehensive understanding of the underlying mechanisms. This research not only underscores the evolutionary continuity of social cognition across primate species but also sets the stage for future explorations into the neural basis of emotion processing. The integration of behavioral analysis with advanced neural recording techniques presents a powerful framework for advancing our knowledge of social cognition and its disorders.

Acknowledgements

I would like to express my sincere gratitude to Dr. Kohitij Kar for giving me the opportunity to complete my master's thesis in his lab and the welcoming environment he created, allowing me to explore different projects. His exceptional guidance and advice during our weekly meetings were vital for accomplishing the work presented here. I am also deeply thankful to Dr. Steven Connor, who guided me as my committee supervisor.

My gratitude goes to the Vista Foundation and the Center for vision research (CVR) at York University for supporting my research throughout this year.

I would like to extend my warmest gratitude to all my colleagues at the VITA lab, who made my working environment that much more enjoyable, especially Dr. Hamid Ramezanour who has helped me a lot and led me to a better understanding, and Anaa Zafer Salim, for the many insightful and joyful discussions, her incredible kindness and friendship.

Finally, my deepest thanks are to my Family, especially my sister and my partner who have always been there, I owe much of my journey to their unconditional support.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	vi
1 Introduction	1
1.1 Behavioral aspects of Facial Emotion Recognition	2
1.2 The Neural Mechanism of Facial Emotion Recognition	3
1.2.1 The Ventral Visual Stream	4
1.2.2 Amygdala	6
1.3 Using a Non-Human Primate Model to Explore Facial Emotion Processing .	7
1.4 Thesis Outline	10
2 Aim1: Compare human and monkey behavior on a battery of a facial emotion judgment task.	11
2.1 Background	11
2.1.1 Objectives of this chapter	13
2.2 Hypothesis	13
2.3 Methods	14
2.3.1 Visual Stimuli and Task	14
2.3.2 Human Behavior	16
2.3.3 Monkey Behavior	17
2.3.4 Behavioral metrics	20
2.4 Results	23
2.4.1 Monkey Behavioral consistency	24
2.4.2 Monkey and Human share behavioral variance	25
2.4.3 Statistical Tests	29
2.5 Inferences	29

3	Aim2: Large-scale, chronic multi-electrode recordings across the macaque ventral visual cortex to identify candidate neural biomarkers of facial emotion processing	32
3.1	Background	32
3.1.1	Objectives of this chapter	35
3.2	Hypotheses	36
3.3	Methods	38
3.3.1	Visual Stimuli and Task	38
3.3.2	Macaque Neural Data Collection and Processing	38
3.3.3	From Neural Activity to Predicting the Behavior	42
3.4	Results	45
3.4.1	IT accurately categorizes different emotions	45
3.4.2	Comparison of macaque IT-based predictions with monkey behavior .	47
3.4.3	Comparison of neural models with Human Behavior	51
3.5	Inferences	54
4	Discussion	58
4.1	Limitations of the current study	58
4.2	Clinical Implications	61
4.2.1	Improving Diagnosis	61
4.2.2	Symptom Mitigation	64
4.3	Future Direction	67
	Bibliography	71

List of Figures

1.1	Facial emotion recognition in non-human primates	3
1.2	The primate ventral visual stream	5
1.3	Human vs. Macaque Ventral Stream	8
1.4	Face-Selective Regions	9
2.1	Conceptual schematic of the binary emotion discrimination task	16
2.2	Replication of the psychometric function for human behavior	17
2.3	Monkey Training	19
2.4	Metrics for Behavioral data	23
2.5	Monkey Behavioral consistency	25
2.6	Human vs Monkey behavioral consistency at emotion level	26
2.7	Human vs Monkey behavioral consistency at image level	27
2.8	Human vs Monkey behavioral consistency	28
3.1	Conceptual schematic of Monkey recording	42
3.2	Conceptual schematic of Decoding	44
3.3	Neural signatures	46
3.4	Consistency of Monkey IT population with Monkey Behavior for different time bins	49
3.5	Consistency between Monkey IT population and Monkey Behavior for the most consistent time bins	50
3.6	Consistency of Monkey IT population with Human Behavior for different time bins	52
3.7	Consistency between Monkey IT population and Human Behavior for the most consistent time bins	53
4.1	Fine-tuned representations from IT-analogous layers of several ANNs	63
4.2	ANN Models of Vision Enable ASD Phenotype Control	66

Chapter 1

Introduction

Facial Emotion Recognition plays a crucial role in how humans understand and navigate social interactions. This skill is essential for effective communication and helps us make decisions in various social contexts [1]. The human face serves as a dynamic canvas, capable of expressing a wide range of emotions without the need for words [2, 3]. This nonverbal language of facial expressions greatly influences our daily lives, affecting our relationships, professional interactions, and even outcomes in important situations like courtroom trials and elections.

The importance of Facial Emotion Recognition is not limited to academic or scientific discussions. It is also reflected in popular culture, such as movies and TV shows, and the widespread use of emojis in digital communication. These cultural examples highlight how universally recognized the face is as a key means of expressing and understanding emotions. For many years, researchers in psychology and cognitive neuroscience have been studying the brain processes that allow us to perceive both fleeting emotional states and lasting social traits from facial expressions [4, 5]. The ability to accurately read and interpret emotional and

social information from faces is not only important for communication but can also be seen as a basic survival skill. Many of these assessments happen very quickly and automatically, which shows how important this cognitive process has been in human evolution [5, 6].

1.1 Behavioral aspects of Facial Emotion Recognition

The study of facial emotion recognition represents a relatively nascent field compared to the extensively explored domain of object recognition [7, 8]. Object recognition has been comprehensively examined across various species, including humans [9], rats [10], mice [11, 12], and non-human primates [13]. Comparative analyses of object recognition between species, particularly between non-human primates and humans, have been conducted [14, 8]. Furthermore, the neural underpinnings of object recognition have been thoroughly investigated [15]. Specifically, within the domain of object recognition, "face" processing has received special attention [16, 17]. Decades of research have discovered a specific sub-region of the primate brain dedicated to processing face-like stimuli [18, 19].

Facial emotion recognition has similarly garnered significant attention in the scientific literature [20, 3], with a predominant focus on human subjects, including clinical populations [21, 22], and notably individuals with autism spectrum disorder [23, 24, 25]. The majority of these investigations have emphasized the neural mechanisms underlying facial emotion recognition [26, 27]. However, there is a relative paucity of research examining the behavioral aspects of facial emotion recognition in non-human primates [28]. Such studies typically employ images of conspecifics exhibiting emotional expressions rather than human faces and

show monkeys can perfectly do the task (Figure 1.1).

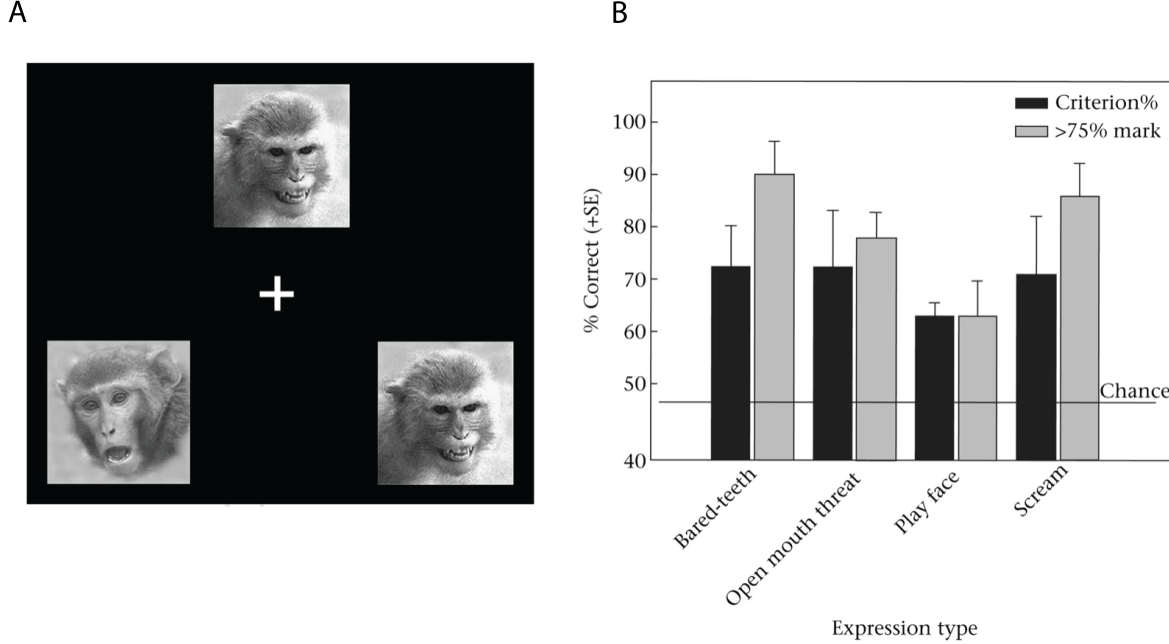


Figure 1.1: **Facial emotion recognition in non-human primates** **A.** Example of a matching-to-sample trial used in the experiment by Lisa A. Parr and Matthew Heintz [28]. **B.** The mean performance of subjects on each facial expression category during the experiment, assessed at two criteria, demonstrates the ability of monkeys to perform facial emotion recognition tasks (adapted from Lisa A. Parr et al. [28]).

A significant lacuna exists in our understanding of the neural correlates of non-human primate behavior in response to human facial emotions. Bridging this gap is critical for advancing our comprehension of the neural and behavioral mechanisms underlying facial emotion recognition across species.

1.2 The Neural Mechanism of Facial Emotion Recognition

Facial emotion recognition involves a complex interplay of various neural mechanisms. Among these, the ventral visual stream plays a key role [29], integrating sensory information to

generate appropriate emotional responses. The following section investigates the key neural structures involved, starting with the ventral visual stream, a crucial component in the processing of facial emotions, and then covers the role of the amygdala in processing facial information [30].

1.2.1 The Ventral Visual Stream

The primate visual system begins its intricate processing pathway at the retina, where light is converted into electrical signals. These signals travel through the optic nerve to the lateral geniculate nucleus (LGN) in the thalamus for initial processing. They are then relayed to the primary visual cortex (V1) in the occipital lobe, where neurons specialize in detecting the orientation and direction of light stimuli [30, 31, 32].

Beyond V1, the visual processing stream divides into two distinct pathways: the dorsal and ventral streams. The dorsal stream, known as the "Where" pathway, is primarily responsible for processing motion and spatial information. In contrast, the ventral stream, referred to as the "What" pathway, is characterized by a hierarchical and feed-forward organization that is essential for object recognition including facial expressions [33, 22]. The ventral stream is anatomically divided into several regions: V1, V2, V4, and the inferotemporal cortex (IT), which is further subdivided into posterior, central, and anterior regions (pIT, cIT, aIT) [33], as depicted in Figure 1.2. These regions sequentially process visual information, evolving from basic feature detection in V1 to more complex facial emotion representations in IT. This hierarchical progression is also evident in the increasing size of receptive fields (RFs) of

neurons from V1 to IT [34].

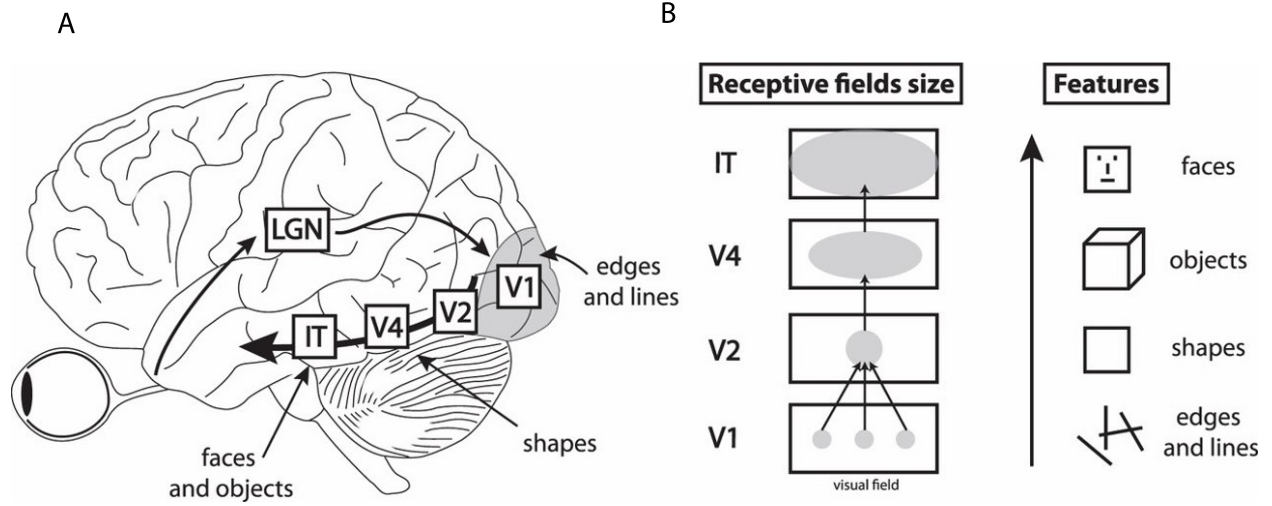


Figure 1.2: **The primate ventral visual stream** **A.** The primate ventral visual stream processes information from the retina through the LGN, V1, V2, V4, to IT. **B** Receptive fields expand from lower to higher visual areas, integrating more of the visual field. V1 codes basic features like edges, while higher-level neurons integrate these, representing more intricate features. (Adapted from Herzog et al. [35])

While V1 is primarily engaged in basic feature extraction, such as the detection of orientations and edges [31], V2 undertakes more complex processing tasks, including the analysis of contours and textures [36]. V4, although still involved in orientation processing [37], plays a significant role in color processing and maintaining color constancy, ensuring consistent color perception under varying lighting conditions [38]. The inferotemporal cortex (IT) is crucial for the recognition of facial emotion attributes. IT neurons often display selectivity for specific categories, including fruits, faces, bodies, and locations [39]. One remarkable property of IT neurons is their invariance to visual transformations. A neuron that responds to a particular object will typically maintain its response despite changes in the object's size, position, or rotation within the visual field [40]. This invariance allows

primates to recognize faces under a variety of viewing conditions [41, 42].

1.2.2 Amygdala

Primates have developed a specialized visual system dedicated to processing facial information, an essential function for social communication and survival [43, 29, 27, 44]. Within this sophisticated network, the amygdala stands out as a critical component [45]. The human amygdala is extensively linked to the recognition and interpretation of facial emotions, as empirical evidence has shown that individuals with amygdala damage often have significant difficulties recognizing fearful facial expressions [46]. Functional magnetic resonance imaging (fMRI) studies have consistently demonstrated that the amygdala exhibits heightened activity in response to fearful faces [47, 48].

Nevertheless, the amygdala’s response is not confined to fear-related stimuli. It has also been observed that the amygdala responds to neutral or happy facial expressions, as indicated by both fMRI studies [49] and single-neuron recordings [50, 51]. Some research, however, suggests that the amygdala shows a preferential response to facial expressions associated with threat, such as fear and anger, compared to neutral or positive expressions [52]. Single-neuron recordings from the human amygdala have revealed that these neurons encode subjective judgments of facial emotions and the detailed content of these emotions [53]. Moreover, amygdala neurons are found to encode social trait judgments and a broad social trait space that informs first impressions based on facial cues [54]. Additionally, these neurons encode key facial features, such as the mouth and eyes, which are crucial for interpreting facial

emotions and social traits [54].

Despite significant advances in our understanding of the ventral visual pathway and the amygdala, limitations in human brain recording techniques have posed challenges in quantitatively assessing whether and how neural representations in this pathway are repurposed for processing facial emotions. Addressing these limitations is crucial for a more comprehensive understanding of the neural mechanisms underlying facial emotion recognition.

1.3 Using a Non-Human Primate Model to Explore Facial Emotion Processing

Intriguingly, the ventral visual processing stream—a hierarchically organized network of interconnected neocortical areas [34]—appears to be remarkably well conserved across numerous primate species. This conservation extends to Old-World monkeys, such as rhesus macaques (*Macaca mulatta*) and baboons (*Papio papio*), which diverged from the human lineage approximately 25 million years ago [55]. Indeed, decades of meticulous research have revealed strong anatomical and functional homologies within the ventral visual hierarchy between humans and macaque monkeys [56, 57, 58]. This hierarchical structure encompasses areas from the primary visual cortex (V1) to the inferotemporal cortex (IT), enabling complex visual processing, including object and face recognition.

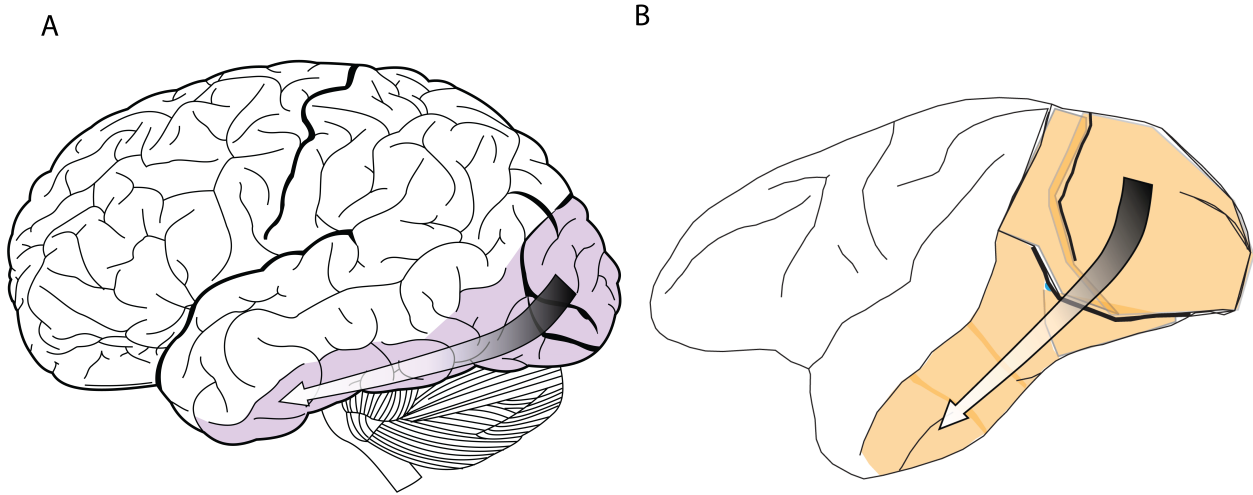


Figure 1.3: **Human vs. Macaque Ventral Stream** **A.** Human ventral visual stream from V1 to IT. **B.** Macaque ventral visual stream homologous to human.

A burgeoning body of research has focused on elucidating how facial identity is encoded within the primate brain, with particular emphasis on the fusiform face areas in humans [27, 29] and the topographically specific face patch systems of the inferior temporal (IT) cortex in rhesus macaques [59, 60]. Recent fMRI investigations have identified several brain regions involved in analyzing the visual characteristics of faces in both humans and macaque monkeys. These findings indicate the presence of multiple functionally specialized "face patches" within the face-processing system [26]. These face patches are regions within the temporal lobe that are selectively responsive to faces, suggesting a highly specialized network for face processing.

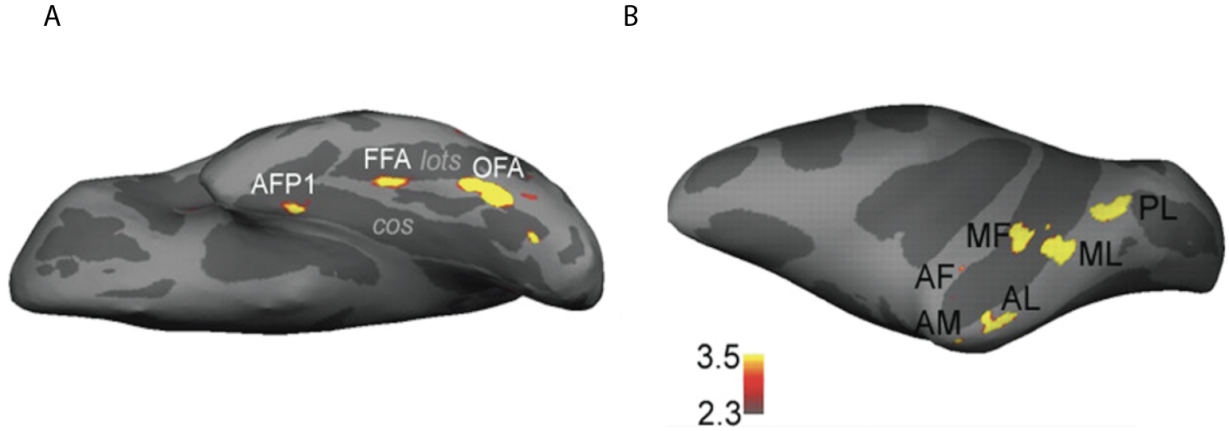


Figure 1.4: **Face-Selective Regions** **A.** Face-selective regions in the temporal lobe of humans. **B.** Face-selective regions in macaques, a lateral view of the inflated hemispheres. Adapted from Tsao et al., 2008 [26]

Furthermore, it has been established that macaque monkeys possess the ability to discriminate between different emotional expressions [61]. Specifically, macaque monkeys have demonstrated proficiency in facial discrimination tasks utilizing monkey faces displaying various emotions [28]. These abilities underscore the functional capabilities of the macaque’s face-processing system, which appears to mirror the complexity and specificity of human facial emotion recognition systems. This ability to discern and react to facial emotions makes macaques an invaluable model for understanding the neural mechanisms underlying facial emotion processing in primates.

The comparative studies between human and macaque face-processing systems highlight the evolutionary continuity and offer significant insights into the neural substrates of social cognition [26]. By leveraging the anatomical and functional similarities between humans and

macaques, we can better understand the intricate neural circuits involved in facial recognition and emotional processing, which has profound implications for understanding human social behavior and related disorders.

1.4 Thesis Outline

This rest of my thesis is structured as follows:

Chapter 2 addresses Aim 1 of our research, focusing on the behavioral alignment between humans and monkeys in facial emotion discrimination. We present a series of behavioral metrics, ranging from coarse to fine-grained measures, developed to probe this relationship comprehensively.

Chapter 3 explores Aim 2, investigating the predictive power of population neural activity in the monkey inferior temporal cortex for the behavioral patterns observed in Aim 1. This chapter builds upon previous research implicating this cortical region in face and object processing.

Chapter 4 provides a critical discussion of our findings, addressing the limitations of the current work and its potential implications for autism research. We evaluate the strengths and weaknesses of our approach and consider future directions for this line of inquiry.

Collectively, these chapters offer a multifaceted examination of facial emotion discrimination across species, integrating behavioral and neurophysiological perspectives to advance our understanding of this crucial cognitive process.

Chapter 2

Aim1: Compare human and monkey behavior on a battery of a facial emotion judgment task.

2.1 Background

Recognizing facial emotions is crucial for understanding social cues and interacting with others. It helps people empathize, cooperate, and form social connections. Most adults naturally understand facial emotions, but children and adults with autism often struggle with this skill to varying degrees. These challenges are believed to contribute to the social and communication difficulties typical of autism [23, 62, 63, 64].

Functional neuroimaging studies have shown abnormal brain activity in people with autism in the areas involved in face processing [65, 66]. When comparing brain activity in response to dynamic and static facial expressions using event-related functional magnetic resonance imaging (fMRI), researchers found reduced activity in the amygdala (AMY) and fusiform gyrus (FFG) in individuals with high-functioning autism when they viewed dynamic

emotional expressions [67].

The pattern of activity in the inferior temporal (IT) cortex in primates is considered a potential neural marker for facial emotion processing and can help develop hypotheses for future studies on autism in both humans and nonhuman primates [68]. However, due to limitations in human brain recording methods, it has been difficult to quantitatively test how neural representations in the ventral visual pathway might be used for emotion recognition processing.

To overcome this limitation of recording from the human brain, here we aimed to use scalable behavioral testing methods to precisely characterize primate behavior at the resolution of individual facial emotion images and to record from the primate brain and directly compare primate behavior to those neural models of primates [15] over the domain of facial emotion recognition behavior at this high resolution.

In a previous study by Wang and Adolph [63], facial emotion recognition was measured in neurotypical adults and individuals with autism. They discovered specific and subtle deficits in emotion perception among people with autism, attributed to reduced specificity caused by increased noise in processing facial stimuli. The primary objective of my study is to extend the seminal work of Wang and Adolph [63] to non-human primates while expanding the emotional spectrum under investigation. My approach aims to elucidate the evolutionary continuity of facial emotion processing and potentially uncover shared neural mechanisms underlying emotion recognition across species. By adapting the paradigm for monkeys, I seek to establish a comparative framework that may provide deeper insights into the fundamental

processes of facial emotion perception, including those that may be altered in conditions such as autism.

2.1.1 Objectives of this chapter

The primary objectives of this study are 1. To develop and implement a facial emotion recognition task suitable for both human and macaque subjects, ensuring comparability. 2. To collect and analyze behavioral data from both species using this task, focusing on their ability to discriminate between different facial emotions across various intensities. 3. To compare the performance of humans and macaques at both the emotion level and image level, using multiple behavioral metrics to provide a comprehensive analysis. 4. To assess the validity of using macaques as a model for studying the neural mechanisms underlying facial emotion recognition in humans. These objectives aim to bridge the gap between human and non-human primate research in emotion perception, potentially providing insights into the evolutionary basis of facial emotion recognition and its neural underpinnings.

2.2 Hypothesis

Hypotheses

This aim is predicated on two central hypotheses:

- **H0:** Monkeys and humans exhibit a similar ability to recognize facial emotions in a behavioral task, suggesting a shared cognitive mechanism for emotional perception.

- **H1:** Monkeys and humans show different behavioral performances in facial emotion recognition tasks, highlighting the unique aspects of primate emotional perception and the need for further investigation.

The ability of monkeys and humans to recognize facial emotions will be tested through a simple task in which they are presented with a sample image and then displayed two choices and asked to choose the correct one (match it to the sample). The probability of choosing the correct emotion for each picture will be calculated for both species, and a statistical analysis will be performed to determine if the probabilities are correlated. If a significant correlation is found, it will be concluded that both monkeys and humans have a consistent ability to recognize facial emotions, and monkeys can potentially be used as a model for the second aim.

2.3 Methods

2.3.1 Visual Stimuli and Task

I have replicated the facial emotion task utilized in Wang and Adolph’s study [63] as illustrated in Figure 2.1.A. To adapt the task for both humans and monkeys, I modified it by replacing the written choices with image choices, making it accessible for monkeys as well. Additionally, I altered the dataset, employing the Montreal Set of Facial Displays of Emotion (MSFDE) dataset [69] as the foundational sample, as referenced in Figure 2.1.B. This dataset is more extensive than the one used in the original study, featuring 12 identities instead of 4 and 6

emotions instead of 2.

To ensure that low-level visual features did not confound the emotion recognition task, all facial images from the MSFDE dataset were standardized using the SHINE (Spectrum, Histogram, and Intensity Normalization and Equalization) toolbox [70]. This preprocessing step equalizes parameters such as luminance, contrast, and spatial frequency across all stimuli, allowing for a more controlled assessment of emotion recognition based solely on facial expressions.

The binary emotion discrimination task for our study is meticulously designed to assess the ability of subjects to discern and categorize emotions presented in a series of emotion discrimination tasks.

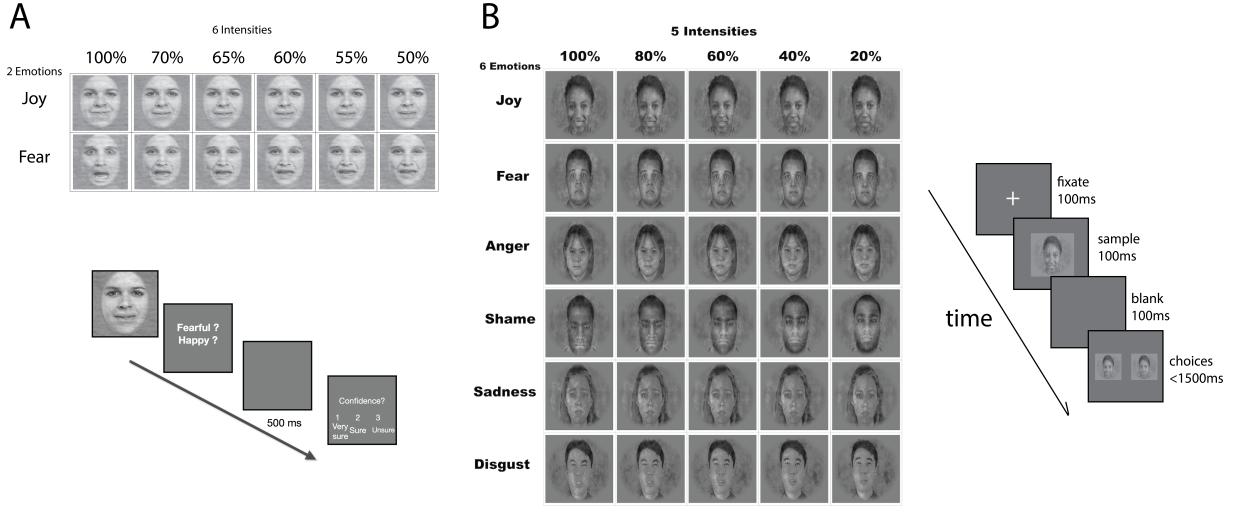


Figure 2.1: **Conceptual schematic of the binary emotion discrimination task A.** The dataset and task used for Wang and Adolph’s study [63] involved presenting a face from the Shinecut image set for 1 second, followed by a question asking participants to identify the facial emotion as either fearful or happy. After a blank screen for 500 ms, participants were asked to indicate their confidence in their decision by choosing ‘1’ for ‘very sure’, ‘2’ for ‘sure’, or ‘3’ for ‘unsure’. An example of each emotion at all intensities was shown for two random identities from the image set. **B.** An example of 6 different identities randomly chosen from MSFDE dataset [69] to show emotions and intensities of the image-set used for both the neural and behavioral experiment. The binary emotion discrimination task for this study shows the timeline of events for each trial. Subjects fixated on a cross, and then the test image at 8 degrees containing one of six possible emotions, is shown for 100 ms. After a 100-ms delay, a canonical view of the target emotion (the same as that presented in the test image) and a distractor emotion (one of the other five emotions) appear, and the human or monkey indicates which emotion was present in the test image by touching one of the two choices on the screen.

2.3.2 Human Behavior

To implement the task for human subjects (illustrated in Figure 2.1.B), we employed the JavaScript library JsPsych. This framework enabled us to craft a task wherein a sample image is presented for 100ms at an 8-degree central visual field following a central fixation symbol ("+"). Subsequently, two choices appear, consisting of human faces displaying emotions at an

extreme level. These faces represent the target (the correct choice) and a distractor represents a random emotion. Participants are instructed to select the target. This task was deployed on Amazon Mechanical Turk (MTurk), allowing subjects to partake in the experiment online via a computer or mobile device for a small payment (Figure 2.2.C). Additionally, to validate the online results, we replicated the task in the lab (Figure 2.2.B), with 3 individuals from the lab completing the task (Pearson correlation = 0.91 ± 0.08).

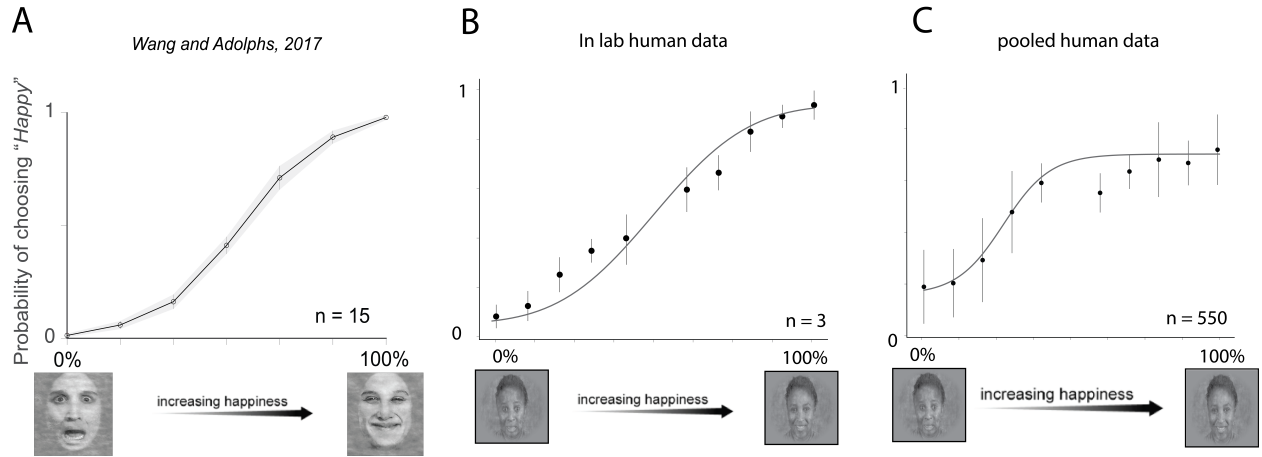


Figure 2.2: **Replication of the psychometric function for human behavior** **A.** Group average of emotion judgments adapted from Wang and Adolphs's [63]. The graph shows how often trials were judged as joyful based on morph levels (from 0% fearful 100% happy on the left to 100% fearful 0% happy on the right). The shaded area represents the standard error of the mean (SEM) across participants. **B.** Replicating the results shown in A but using different images from the MSFDE dataset from 3 lab members, showing how often trials were judged as joyful based on morph levels. The error bars represent the mean (SEM) across participants. **C.** The same approach for combined data from online participants on Amazon Mechanical Turk, showed how often trials were judged as joyful based on morph levels. The error bars represent the 95% confidence intervals.

2.3.3 Monkey Behavior

Monkeys performed a similar task. We collected data from two male rhesus macaques. Images were displayed on a 24-inch LCD monitor (1920×1080 at 60 Hz) placed 42.5 cm in front of

the monkey. The monkeys had to touch a small white dot (0.2°) to start a trial. Each trial began with a sample image (chosen from a set of 359 images) shown for 100 ms, followed by a blank gray screen for 100 ms. Then, a choice screen appeared, displaying a standard image of the target emotion (the correct choice) and a standard image of the distractor emotion. The monkeys could view the choice images for up to 1500 ms. Before laboratory testing, we trained the monkeys in their home cages using kiosks [71] to perform delayed match-to-sample tasks using the same images (but with high-intensity emotions). We obtained a minimum of 52 trials per image from the combined responses of the two monkeys.

Monkey Training

To ensure valid comparisons, it was essential for both monkeys and humans to perform similarly structured tasks. In our study, we trained two monkeys [72] on the facial emotion discrimination task. These monkeys were previously trained in object recognition tasks [73].

The training setup included enclosures with touchscreen devices—using iPads programmed with the MWorks platform (<https://mworks.github.io>). Initial training used a subset of stimuli from the human task (two identities: numbers 1 and 12) and three emotional expressions (joy, fear, and shame) at extreme intensities (100% and 80%). Baseline performance was 58% accuracy.

Training sessions continued until the monkeys consistently achieved over 80% accuracy, reaching the human accuracy range for these images. Then, additional identities and emotions for extreme-intensity images were added, and training continued until they again surpassed

the human accuracy range. Evolution of average response in a behavioral training program for monkeys over the number of trials, highlighted by key training phases (Figure 2.3.B).

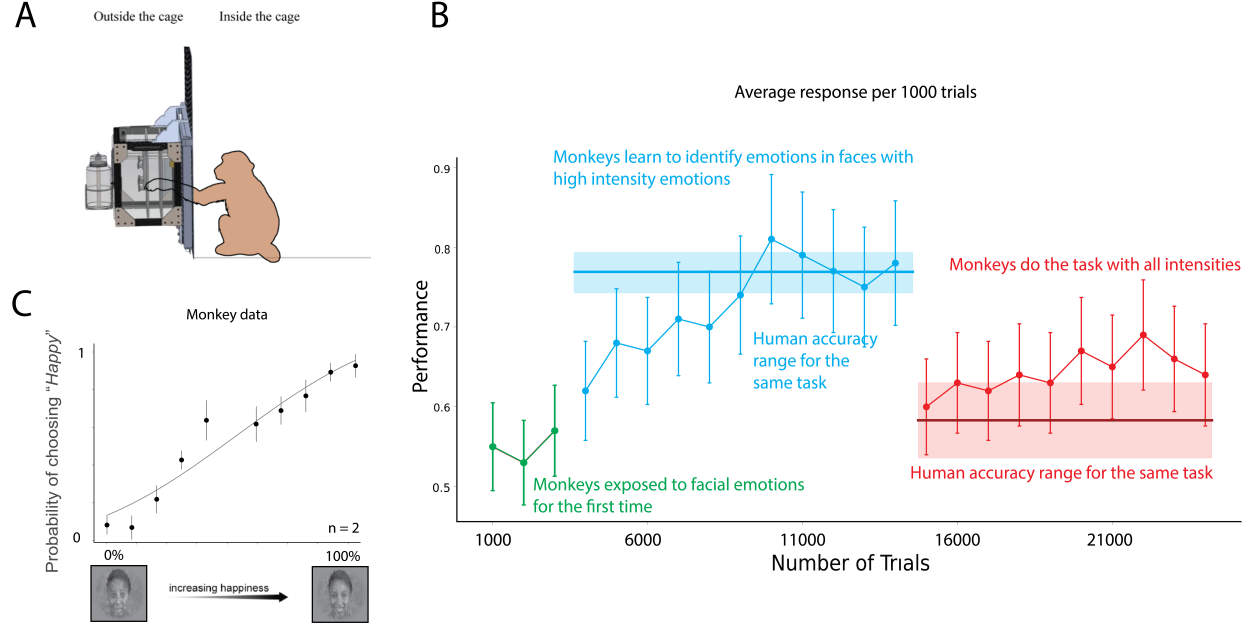


Figure 2.3: Monkey Training **A.** The monkey accesses the touch screen by positioning themselves in front of the kiosk on a horizontal divider panel within the cage [71]. **B.** Training process for the macaque monkey. The monkey was initially trained on the object recognition task. We started training the monkey with extreme-intensity images (green curve). However, monkeys quickly learn to recognize images with extreme intensity (blue curve) and the horizontal line shows the human accuracy for the same task and the lighter band indicates the 95% confidence intervals. Furthermore, their ability was tested for all images (red curve) and the horizontal bar indicates the human accuracy range in which the monkey could hit that. **C.** The graph shows how often trials were judged as joyful based on morph levels (from 0% fearful 100% happy on the left to 100% fearful 0% happy on the right). The error bars represent the 95% confidence intervals.

Once the monkey performances reached the human accuracy range in performing the task with extremely intense emotions, the entire image set was used for testing, matching the task parameters for both monkeys. This sequential training approach effectively prepared monkeys for the comprehensive emotion discrimination task, allowing for a direct comparison of monkey behavior with human behavior.

2.3.4 Behavioral metrics

To thoroughly analyze the behavior of monkeys and humans, we employed four behavioral metrics at two distinct levels: emotion and image. These metrics are the hit rate resolution at the emotion level (E1), emotion and task level (E2), and the more detailed image level (I1), image and task level (I2), as elaborated in previous researches [74]. Each metric calculates a pattern of unbiased behavioral performance by averaging the accuracies per image across all trials. By applying these metrics to the behavioral accuracies, we derived a distinct biological or artificial "signature" for each system, providing a comprehensive assessment of their performance.

The one-versus-all emotion-level performance metric (E1) is designed to assess the discriminability of each emotion category. This is achieved by pooling accuracies across all images of a specific emotion type and all corresponding emotion/distractor pairs. Given that six emotion categories were tested, the E1 signature comprises six independent values. These values represent the ability of the subjects to correctly identify each emotion when compared against all other emotions combined. Figure 2.4.A (left) displays the E1 behavioral signatures for both monkeys and human participants, highlighting how each group discerns among the six emotions.

The one-versus-other emotion-level performance metric (E2) evaluates the discriminability of each emotion category against each different distractor emotion. This metric pools accuracies across all images of a specific emotion type and one emotion/distractor pair at a time. Since six emotion categories were tested, the E2 signature contains 30 independent

values (6×5). These values offer a more detailed look into how well the subjects can distinguish each emotion from each of the other five emotions individually. Figure 2.4.A (right) presents the E2 behavioral signatures for both primates and human participants, providing a nuanced understanding of their discriminative abilities at the emotion level.

The one-versus-all image-level performance metric (I1) determines the discriminability of each image containing an emotion from all other emotions, pooling across all possible distractor choices. Focusing on the primary image test set of 359 images, the resulting I1 signature includes 359 independent values. Each value corresponds to the accuracy of recognizing a specific image when all other images serve as distractors. This metric evaluates the performance for an image i of object e against all five distractor emotions ($d \neq e$). Figure 2.4.B (left) depicts the I1 signatures, demonstrating how each system handles a large set of individual images across different emotions.

The one-versus-other image-level performance metric (I2) assesses the discriminability of each image containing an emotion from each other emotion by pooling across each possible distractor choice individually. With the same set of 359 images, the I2 signature comprises 359×5 independent values. This metric evaluates the performance for image i of object e against each of the five distractor emotions ($d \neq e$), providing a highly detailed analysis of image-level discriminability. Figure 2.4.B (right) illustrates the I2 signatures, showing the subjects' ability to differentiate each image from specific distractor emotions. We computed

the performance per image versus distractor (I2 metric) as:

$$I2_i^e = \sum_{d=1}^6 PC_i^{e,d \neq e}$$

where PC - percent correct, is the fraction of correct responses for the binary task between emotions e and d. And we computed the average performance per image (I1 metric) as:

$$I1_i^e = \frac{\sum_{d=1}^6 I2_i^d}{5}$$

Considering every image i of emotion type e, the E2 performance for each context type is the mean across all images' with the same emotion and distractor:

$$E2_{e,d} = \frac{\sum_{d=1}^6 I2_i^d}{60}$$

We computed the average performance per emotion (E1 metric) as:

$$E1_e = \frac{\sum_{d=1}^6 E2_e}{6}$$

All of these behavioral signatures are however tightly linked. For instance, images with emotions that is challenging to discriminate generally would display lower performance metrics as opposed to images with less challenging emotional features.

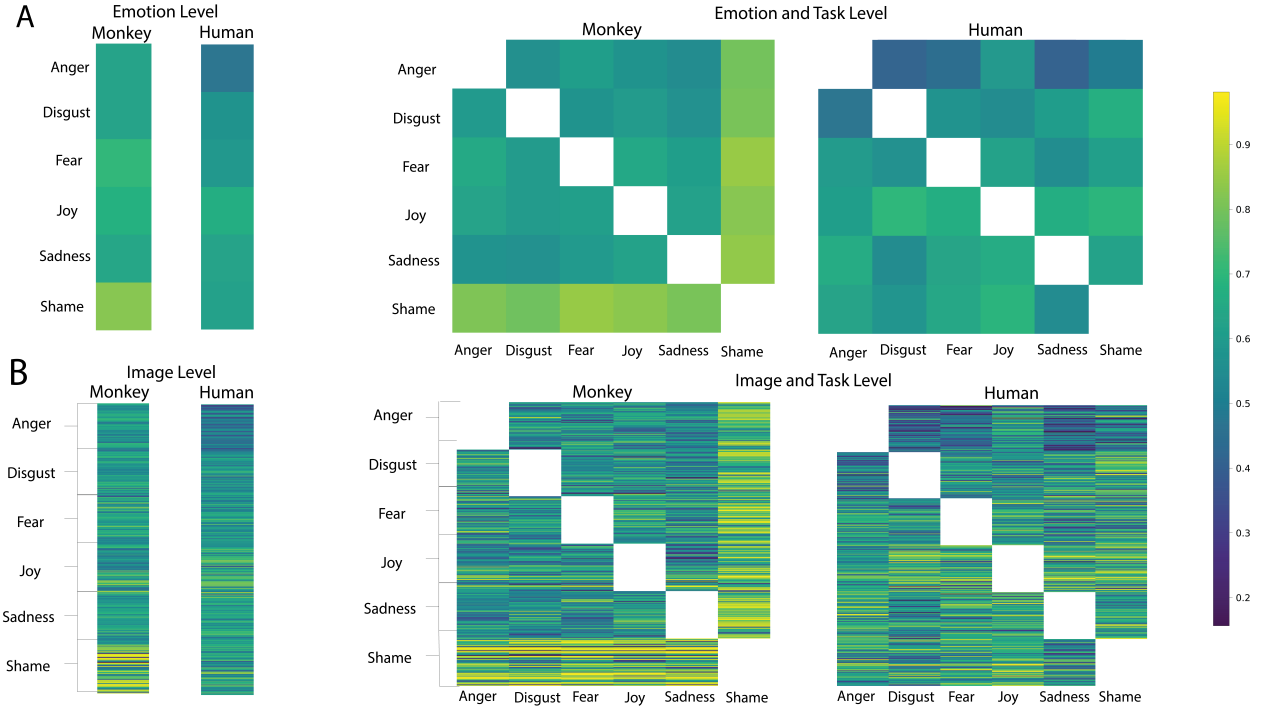


Figure 2.4: **Metrics for Behavioral data** A matrix was created to measure the ability of participants to correctly identify emotions in a set of 359 images. Each image was presented with two choices, and participants were asked to choose the correct emotion. The resulting matrix consisted of n rows (n being the number of images) and m columns (m being the number of different options available for measurement). In our study, the measurement was focused on emotions (e.g., happy vs. sad), but the matrix could also be used to measure other factors such as gender or ethnicity. The result for each image was the probability of the correct emotion being chosen, which was used to fill the emotion column in the matrix. A similar matrix is relevant for the emotion change detection task (rows will be arranged according to the starting emotion, and columns are ending emotions).

2.4 Results

We measured human and macaque behavior on the task. To understand the neural processes behind the emotional influences, we require a more detailed examination of the neural networks involved. We observed reliable emotional modulation in the monkeys' behavioral accuracy, which was significantly correlated, and aligned with those observed in humans at

the image-by-image level.

2.4.1 Monkey Behavioral consistency

We trained two monkeys to perform the task (see Figure 2.3). To compare their behavior with human behavior, we first need to assess the consistency between the two monkeys' data. We calculated behavioral metrics at both the image level and emotion level. For each metric, we computed the correlation between Monkey 1 and Monkey 2. We then adjusted these correlation values using the noise ceiling correction, calculated as the square root of the internal consistency (by calculating the split-half reliability across trial repetitions [15]) for each monkey's metrics, image level (noise-corrected Pearson $R = 0.71$) and emotion level (noise-corrected Pearson $R = 0.94$). This correction accounts for differences in individual capacities between the monkeys (see Figure 2.5).

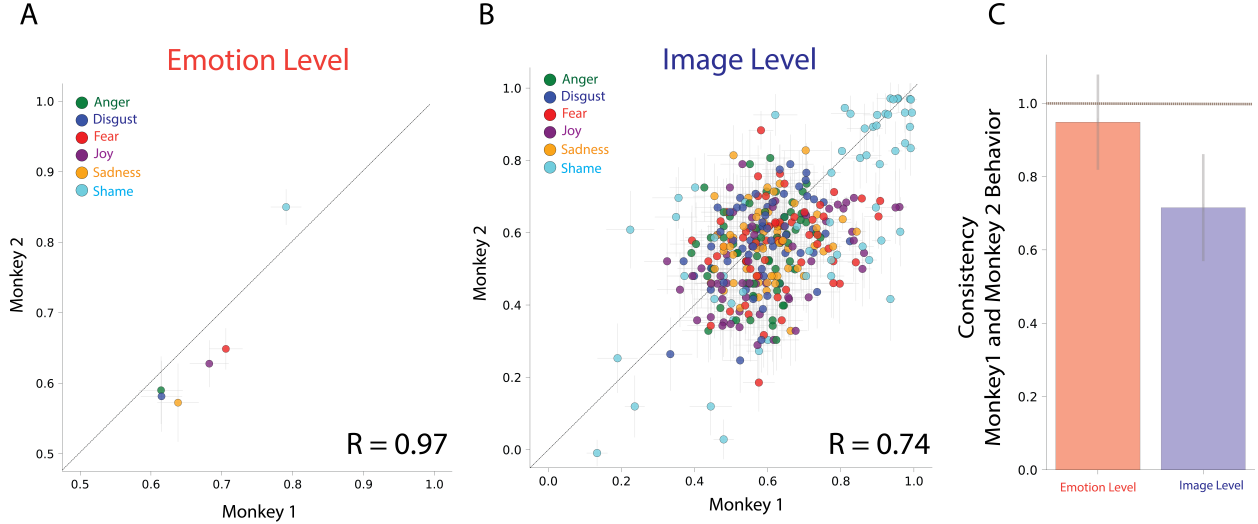


Figure 2.5: **Monkey Behaviors are consistent at both emotion and image level** The correlations between the two monkeys for each metric have been computed and corrected for the monkeys’ noise ceiling. The image level correlation has a p-value of $1.432e-27$, and the emotion level correlation has a p-value of 0.004. Error bars are calculated using bootstrapping over the means.

2.4.2 Monkey and Human share behavioral variance

We measured the correlation between monkey and human behavior at both image and emotion levels and then corrected the correlation by monkey and human noise ceiling. We calculated the internal consistency by calculating the split-half reliability across trial repetitions [15] of the human’s behavioral accuracy as the emotion varied. We obtained a high correlation of Pearson $R = 0.67$. The pooled humans and macaques have very similar facial emotion recognition patterns. Among the various levels of analysis, our results revealed the highest degree of consistency between humans and macaques at the emotion category level (Figure 2.6). This finding suggests that both species exhibit similar patterns in recognizing and categorizing distinct emotional expressions, despite potential differences in their responses to

individual images.

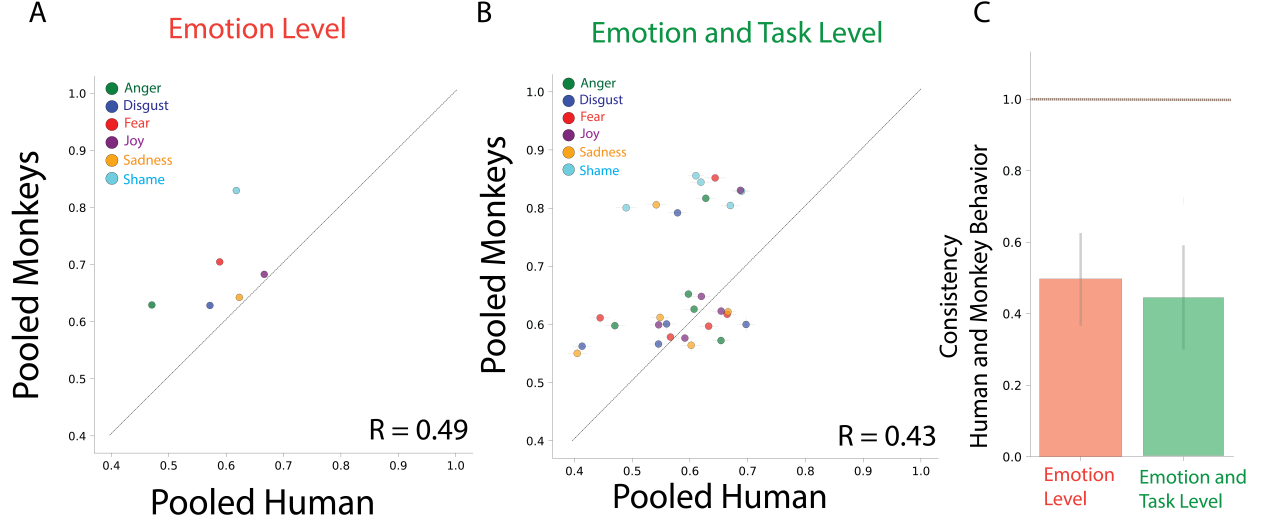


Figure 2.6: **Human and Monkey behavior are the most consistent at emotion level** **A.** The correlation between human and monkey emotion-level E1 noise-corrected Pearson $R = 0.49$ and $p\text{-value} = 0.0437$, **B.** Emotion and task level E2: noise-corrected Pearson $R = 0.43$ and $p\text{-value} = 0.0044$ **C.** The summary of correlations for both metrics and error bars indicates the 95% confidence interval.

We have also measured the consistency between monkeys and humans at finer-grained metrics in image level metrics and we found a significant correlation at this level as well. Image level I1: noise-corrected Pearson $R = 0.39$ and Image and task level I2: noise-corrected Pearson $R = 0.26$. (Figure 2.7).

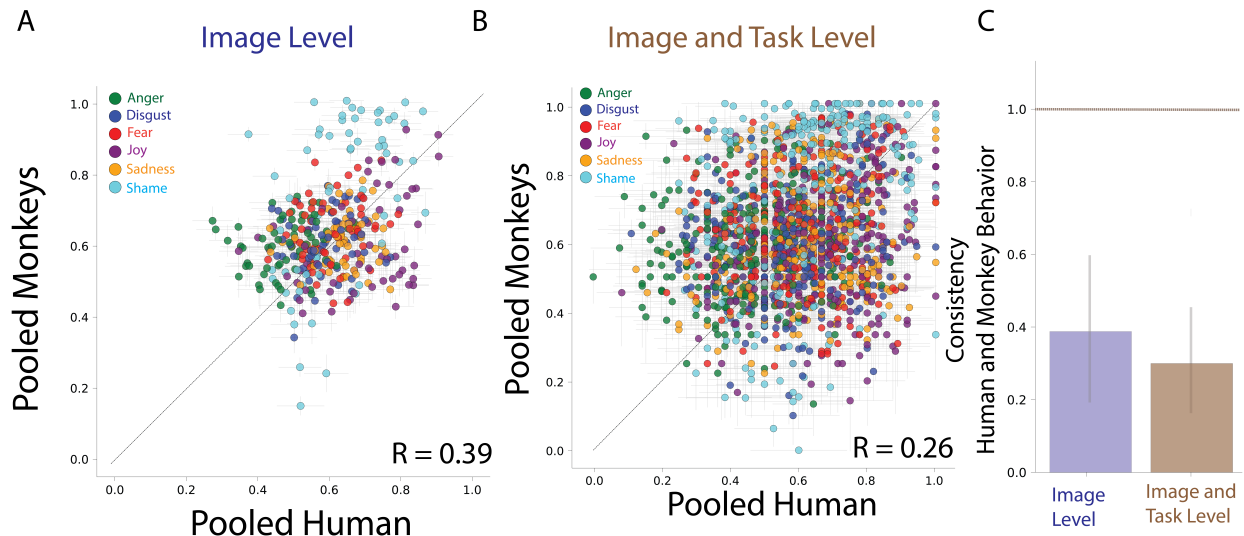


Figure 2.7: **Human and Monkey behavior are consistent at image level A.** The correlation between human and monkey at Image-level I1: noise-corrected Pearson $R = 0.39$ and $p\text{-value} = 2.84e-09$, **B.** Image and task level I2: noise-corrected Pearson $R = 0.26$ and $p\text{-value} = 4.99e-17$. we are seeing a significant correlation in both metrics. **C.** The summary of correlations, and error bars indicate the 95% confidence interval.

In Figure 2.8 we have a summary of the consistency between monkey and human behavior for all four metrics from coarser to the finer-grained metrics.

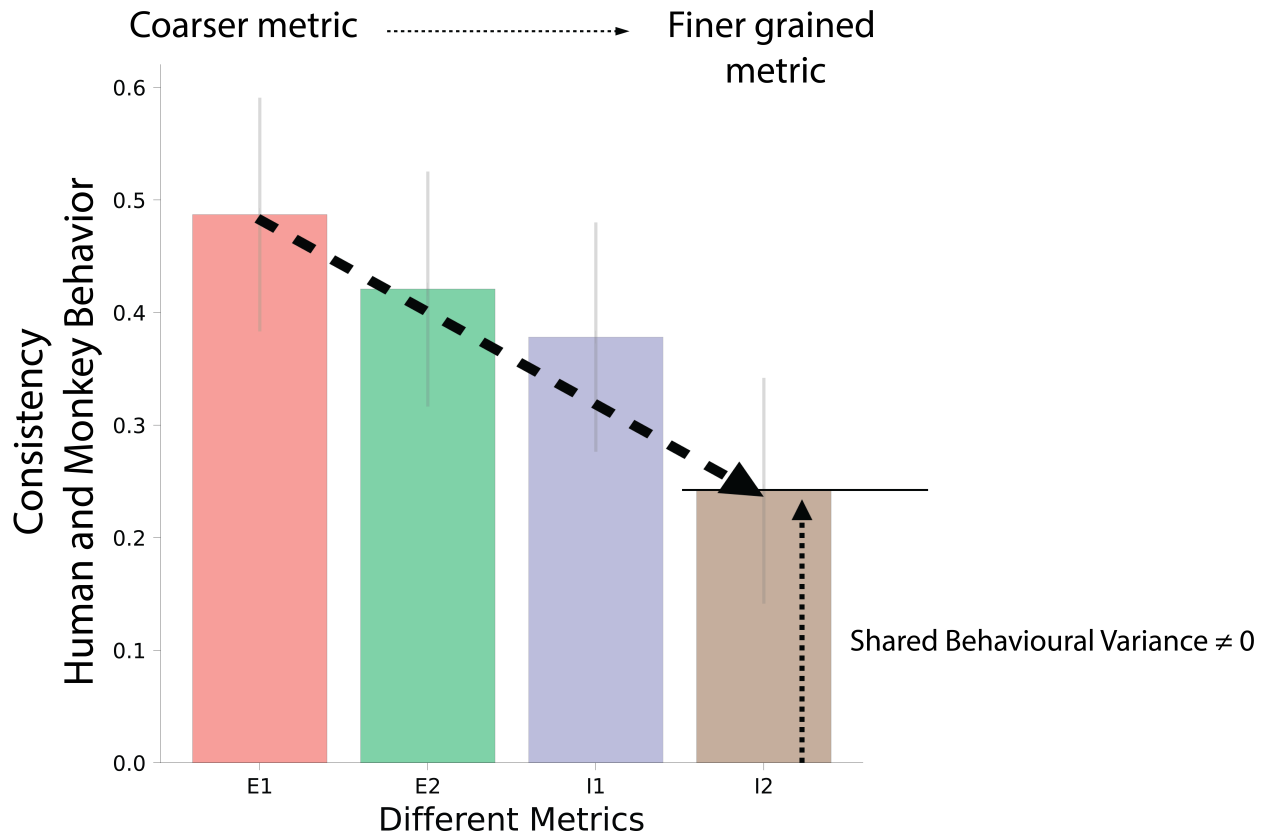


Figure 2.8: **Summary of Human vs Monkey behavioral consistency for all metrics**
 The summary of correlations for all metrics from coarser metric to finer-grained, error bars indicate the 95% confidence interval..

As expected, we see a decrease in the consistency between humans and monkeys as we use a finer-grained metric but we still see a shared variance between them even at the I2 level. These observations establish monkeys as a good model of humans to further study the neural mechanisms of facial emotion recognition.

2.4.3 Statistical Tests

In this study, all correlation coefficients were calculated using the Pearson formula. To evaluate the significance of these correlations between two variables of interest, a bootstrapped permutation test was employed. This process began with the construction of a null hypothesis by combining the two variables into a single dataset. From this mixed dataset, two vectors were generated by random sampling, with replacement, the same number of elements as in the original variables. This procedure was repeated multiple times, typically exceeding 1000 iterations, to generate a distribution of correlation coefficients under the null hypothesis. The actual observed correlation was then compared to this null distribution to compute the p-value, which indicates the probability of rejecting the null hypothesis.

2.5 Inferences

In this chapter, I have demonstrated the successful development and implementation of a facial emotion recognition task that could be performed by both human and macaque subjects. This allows for direct comparisons between monkeys and humans. The adaptation of the task for macaques, including the use of touch screens and image-based choices instead of identifying object names as words (as used in Wang and Adolphs, 2017) demonstrates the feasibility of conducting such experiments in monkeys.

The high correlation observed between the two macaque subjects (noise-corrected Pearson $R = 0.71$ at the image level and $R = 0.94$ at the emotion level) suggests a strong consistency

in facial emotion recognition abilities within the species. This consistency provides a solid foundation for using macaques as a model for studying the neural mechanisms underlying facial emotion recognition. The high internal consistency also indicates that the task is reliable and produces reproducible results across individual macaque subjects.

The study revealed significant correlations between human and macaque performance across multiple behavioral metrics, particularly at the emotion level (E1 and E2) and image level (I1 and I2). These correlations suggest that there are fundamental similarities in how humans and macaques process and recognize facial emotions. This finding supports the idea of a shared evolutionary basis for facial emotion recognition. It validates using macaques as a model for studying the neural mechanisms underlying this cognitive process in humans.

The study observed a decreasing trend in the consistency between human and macaque performance as the behavioral metrics became more fine-grained (from E1 to I2). This gradient suggests that while the overall patterns of emotion recognition are similar between species, there may be more variability in how specific images or subtle emotional cues are processed. This finding emphasizes the importance of using multiple levels of analysis when comparing cognitive processes across species.

The significant correlations between human and macaque performance, particularly at the emotion level, support using macaques as a model for studying the neural mechanisms of facial emotion recognition. This validation is crucial for future research, as it allows for more invasive and detailed neural recordings in macaques that can inform our understanding of human brain function in emotion processing.

The study demonstrates the effectiveness of using carefully designed behavioral tasks and multiple analysis metrics to compare cognitive processes. Using online platforms (Amazon Mechanical Turk) for human data collection, alongside laboratory-based testing, showcases the potential for large-scale, diverse data collection in comparative cognitive research.

While this chapter provides valuable insights, it also highlights some limitations. The relatively small sample size of macaques (two subjects) suggests that replication with a larger sample would be beneficial. Additionally, the study focuses on behavioral comparisons, leaving open the question of how neural activity patterns compare between monkeys and humans during facial emotion recognition tasks. Future studies combining behavioral measures with neuroimaging or electrophysiological recordings in both species could provide a more comprehensive understanding of the neural basis of facial emotion recognition. Demonstrating similarities and differences in facial emotion recognition between humans and macaques provides a framework for investigating other complex cognitive processes. This approach can help bridge the gap between animal models and human cognition, potentially leading to more translational research in social cognition, decision-making, and emotional regulation.

Chapter 3

Aim2: Large-scale, chronic multi-electrode recordings across the macaque ventral visual cortex to identify candidate neural biomarkers of facial emotion processing

3.1 Background

Understanding the neural mechanisms underlying facial emotion recognition (FER) is pivotal for advancing our knowledge of social cognition. FER involves a network of brain regions that process and interpret emotional cues, playing a crucial role in social interactions and communication [4, 6].

Research has extensively studied the anatomical and functional aspects of the ventral visual cortex, particularly the inferotemporal (IT) cortex, which is integral to recognizing facial expressions and processing emotional information [33, 34]. The IT cortex's role in extracting meaningful information from visual stimuli provides insights into the neural representations

of emotions and their underlying mechanisms [26, 27].

The visual pathway’s ventral stream, comprising anatomically distinct regions including V1, V2, V4, and the IT cortex, exhibits a hierarchical and feed-forward organization crucial for facial expression recognition (FER) [33, 34]. This hierarchical structure is exemplified by the progressive enlargement of neuronal receptive fields from V1 to IT, facilitating the assimilation of increasingly complex visual stimuli [34, 30]. IT neurons display selectivity for specific categories, including faces, and exhibit invariance to visual transformations such as size, position, and rotation, which is crucial for consistent face recognition across different contexts [40, 41].

While much progress has been made, the tools we have for studying the human brain still have limits. These limits make it hard to measure exactly how the brain’s visual pathway uses information to understand facial emotions. To fully grasp how the brain recognizes facial expressions, we need to find ways around these challenges [35, 37].

Using non-human primate models, particularly macaques, has provided valuable insights due to the strong anatomical and functional homologies within the ventral visual hierarchy between humans and macaques [56, 58]. Decades of research have indicated that the macaque IT cortex is central to recognizing facial expressions and processing emotional cues [59].

Aim 2 of this thesis focuses on utilizing large-scale, chronic multi-electrode recordings in the macaque ventral visual cortex to identify candidate neural biomarkers of facial emotion processing [33, 8]. By examining the neural activity patterns associated with the perception of different facial emotions, this study aims to provide a detailed characterization of the

ventral visual cortex’s role in emotion processing.

This research builds on computational models of FER, such as artificial neural networks (ANNs), which simulate IT neuron activity and predict behavior. These models align closely with observed neural patterns, supporting the hypothesis that deviations in IT cortex activity underlie FER processes [14, 8]. By validating these computational predictions with biological data, this study aims to enhance the specificity of inferences regarding the neural markers of FER [33, 8].

The identification of neural biomarkers within the IT cortex could lead to early diagnostic techniques and personalized interventions in conditions such as autism spectrum disorder. Understanding the neural circuitry involved in facial emotion processing is essential for developing targeted therapeutic strategies [4, 26].

Recent studies have highlighted the use of multi-electrode arrays implanted in the IT cortex for neural recordings. These arrays allow for the sampling of neural activity across various IT regions, providing a comprehensive view of the neural dynamics involved in FER [42]. The data collected from these recordings can be analyzed to identify specific neural markers associated with different emotional responses, offering insights into the neural coding strategies employed during facial emotion recognition.

In the context of this thesis, large-scale neural recordings were conducted in macaques while they viewed sequences of facial expressions. The neural data were processed and analyzed to map the activity patterns within the IT cortex, aiming to link these patterns with behavioral outcomes observed in FER tasks [35, 37]. This approach bridges the gap

between computational models and biological observations, enhancing our understanding of the neural mechanisms underlying FER.

Overall, this chapter delves into the methods and findings from these neural recordings, providing a comprehensive analysis of the neural substrates involved in FER. By comparing these neural activity patterns with behavioral outcomes, this research aims to contribute significantly to the field of social cognition and improve diagnostic and intervention strategies.

3.1.1 Objectives of this chapter

The primary objective of this study is to leverage large-scale, chronic multi-electrode recordings in the macaque ventral visual cortex to identify and characterize neural biomarkers of facial emotion processing. Specifically, we aim to: 1. determine whether population activity in the macaque IT cortex can accurately categorize facial emotions; 2. assess the consistency between IT cortex activity-based predictions and observed macaque behavior in facial emotion recognition tasks, and 3. evaluate the alignment between macaque IT cortex activity-based predictions and human facial emotion recognition behavior.

By addressing these objectives, we seek to bridge the gap between computational models and biological observations, potentially uncovering conserved neural mechanisms underlying facial emotion recognition across primate species.

3.2 Hypotheses

The investigation into the neural underpinnings of facial emotion recognition in macaques is structured around several interconnected hypotheses. This section outlines these hypotheses, aiming to understand the predictive power of the inferotemporal (IT) cortex about facial emotion recognition behavior and to compare these predictions across species.

We are testing our hypothesis, predicated on a preliminary hypothesis, which must be tested sequentially:

Null Hypothesis: The population activity in the macaque IT cortex cannot categorize facial emotions.

Hypothesis: The population activity in the macaque IT cortex can accurately categorize facial emotion.

This hypothesis tests whether the neural activity patterns in the IT cortex are informative enough to determine the outcomes of facial emotion recognition tasks. If the null hypothesis is rejected, it would indicate that the neural signals within the IT cortex can recognize and discriminate between different facial emotions.

Null Hypothesis: The accuracy of facial emotion recognition based on IT cortex activity is not consistent with the behavior observed in macaques.

Hypothesis: The accuracy of facial emotion recognition based on IT cortex activity is consistent with the behavioral responses observed in macaques.

This hypothesis aims to validate the neural predictions against behavioral outcomes in macaques. If the null hypothesis is rejected, it would mean that the predictions made from

the IT cortex activity match the actual behavior exhibited by the macaques during facial emotion recognition tasks, confirming that the IT cortex activity is a reliable indicator of how macaques recognize and respond to facial emotions.

Hypothesis: The accuracy of facial emotion recognition based on IT cortex activity in macaques is not consistent with the behavior observed in humans.

Main Hypothesis: The accuracy of facial emotion recognition based on macaque IT cortex activity is consistent with human facial emotion recognition behavior.

This last hypothesis suggests that the neural mechanisms underlying facial emotion recognition are conserved between humans and macaques. If the null hypothesis is rejected, it would indicate that the neural activity in the IT cortex of macaques not only predicts their behavior but also aligns with human behavior in facial emotion recognition tasks. This consistency would suggest similar neural mechanisms are involved in facial emotion recognition in both macaques and humans, reinforcing the validity of using macaques as a model for human facial emotion research.

By systematically testing these hypotheses, this study aims to elucidate the neural basis of facial emotion recognition.

3.3 Methods

3.3.1 Visual Stimuli and Task

We utilized a dataset of 359 gray-scale facial images, representing six emotions: anger, disgust, fear, joy, sadness, and shame. These images were evenly distributed among the emotions and were used in our behavioral task, sourced from the Montreal Set of Facial Displays of Emotion (MSFDE) [69]. Detailed information about the dataset can be found in section 2.1.3.

3.3.2 Macaque Neural Data Collection and Processing

Our experiments involved two adult male rhesus macaques (*Macaca mulatta*), each about 10 years old. All procedures, including data collection, surgery, and animal care, followed NIH guidelines, were approved by the Massachusetts Institute of Technology Committee on Animal Care, and adhered to the Canadian Council on Animal Care standards, as well as the York University Animal Care Committee’s regulations.

Neural recordings were made using two Utah micro-electrode arrays implanted in the IT cortex, each array containing 96 electrodes arranged in a 4mm x 4mm grid with 400 μm spacing. This setup allowed sampling across various IT cortex regions. For analysis, we treated each neural site as a random sample from the pooled IT population, without considering their specific spatial locations.

Macaque Passive Emotion Task

In the passive viewing task (Figure 3.1.A,B), monkeys fixated on a white dot (0.2°) for 300 ms to start a trial. We displayed sequences of 5 to 10 images, each for 100 ms, with a 100 ms gray screen between images. This was followed by a fluid reward and a 500 ms interval before the next sequence. The neural data analyzed were collected while the monkeys ($n=2$) performed these passive fixation tasks.

Eye Tracking

We tracked eye movements using the vPIXX video eye tracker. Subjects were trained to fixate on a central white square (0.2°) within a $\pm 2^\circ$ window using operant conditioning and water rewards. Calibration tasks, involving saccades to spatial targets and fixation for 500 ms, were performed at the start of each session and repeated if drift occurred.

Real-time eye-tracking ensured eye jitter stayed within $\pm 2^\circ$; otherwise, the trial was aborted and data discarded. Stimulus display and reward control were managed by MWorks Software.

Electrophysiological Recording and Data Preprocessing

Before neural recordings, each monkey was surgically implanted with a titanium headpost under aseptic conditions. We implanted 96-electrode Utah arrays (Blackrock Microsystems) in the IT cortex of the monkeys: two arrays in the right hemisphere of monkey 1 and two in the left hemisphere of monkey 2. Each array was a 4.4mm x 4.2mm grid with 400 μ m spacing,

and 1.5mm long electrodes with platinum tips. The arrays were pneumatically inserted below the superior temporal sulcus and in front of the posterior middle temporal sulcus. Prior research has conducted analyses of multi-unit neuronal activity, resulting in the identification of consistently responsive neural loci [75].

During sessions, multiunit neural activity was recorded at 20 kHz using an Intan RHD Recording Controller (Intan Technologies, LLC). Raw voltage signals were bandpass filtered offline with a second-order elliptical filter (300 Hz to 6 kHz, 0.1 dB passband ripple, 50 dB stopband attenuation) and thresholded to detect multiunit spikes. A spike event was defined as the voltage crossing the threshold (falling edge) exceeding three times the standard deviation of the raw voltage. We saved 1.6 ms waveforms for each spike event, including a 0.4 ms pre-trough period. These waveforms were used to sort spikes into individual neurons and classify cell types.

Sorting Multiunit Neurons to Single Neurons

We normalized the 1.6 ms multiunit spike waveforms to a range of -1 to 1. For each multiunit site, the normalized waveforms were reduced to a lower-dimensional space using principal components analysis (PCA), with the top 2 principal components used for k-means clustering. Each waveform cluster was identified as an individual neuron. To ensure we selected neurons rather than noise, we calculated image rank order reliability for each neuron. This involved splitting responses into halves across repeated image presentations and computing the Spearman-Brown corrected Pearson correlation between the splits (Figure 3.1.C).

We also implemented a specialized normalization process to address the day-to-day variability in neural recordings from monkeys. This process comprised two primary steps. Firstly, we subtracted the background neural response—the average response to blank images—from the recorded response at each neural site. This step was essential to remove baseline neural activity unrelated to the stimulus, ensuring that subsequent analyses would capture genuine stimulus-driven activity. Secondly, we normalized these adjusted responses by dividing them by the standard deviation of all responses recorded at that site on the same day. This normalization by standard deviation helped to standardize the response data across different recording sessions, reducing sensitivity to daily variations in recording conditions such as electrode placement or neural responsiveness. This methodological approach aimed to minimize the impact of day-to-day fluctuations on the analysis, enabling more reliable comparisons of neural activity over time.

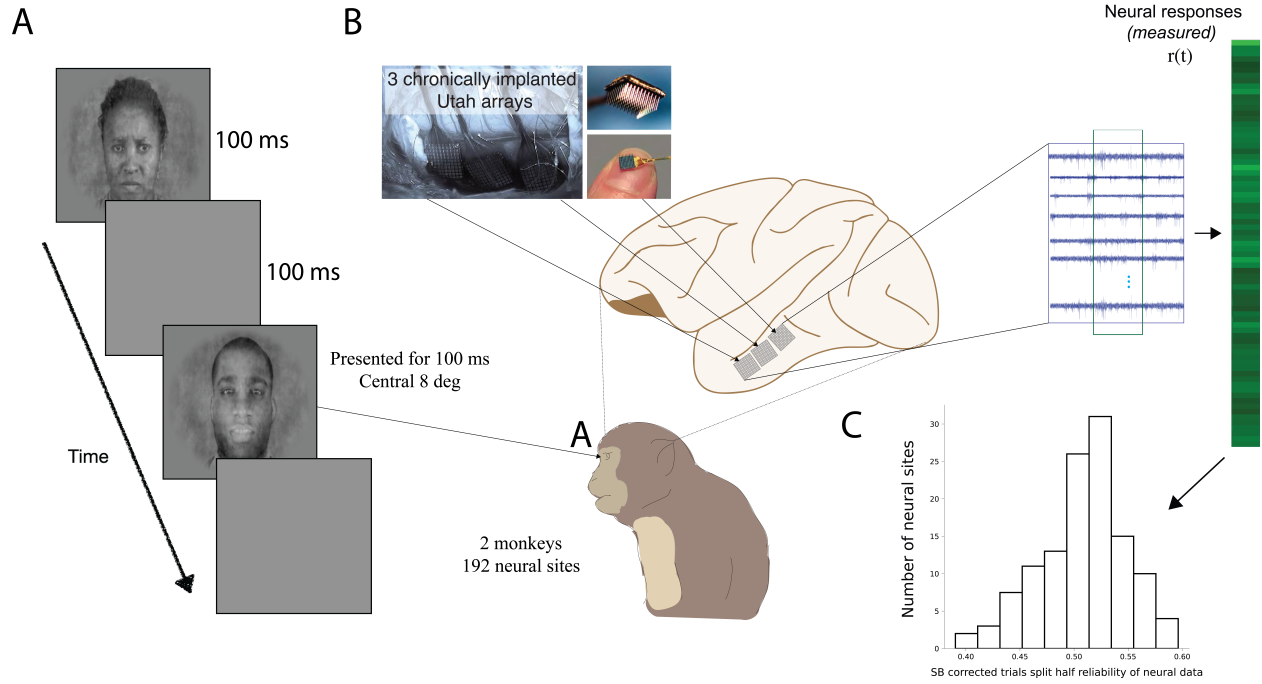


Figure 3.1: **Conceptual schematic of Monkey recording A,B.** We measured neural responses from the inferior temporal (IT) cortex using chronically implanted multi-electrode Utah arrays while the monkeys passively fixated on images ($n=359$). The inset shows a schematic of bandpass-filtered voltages across neural recording sites. **C.** The self-consistency, measured by the neural split-half correlation for each neuron, is shown on the x-axis. The y-axis depicts the number of neurons.

3.3.3 From Neural Activity to Predicting the Behavior

To estimate what information downstream neurons could easily “read” from a given IT neural population, we used a simple, biologically plausible linear decoder, specifically linear classifiers, which have been previously shown to effectively link IT population activity and primate behavior [75]. Linear decoders are designed to perform binary classifications by computing weighted sums of input features, where each weight is analogous to the strength of a synapse, and separating the outputs based on a decision boundary, akin to a neuron’s spiking threshold.

In this study, we employed a logistic regression-based classifier. This type of classifier is particularly suitable because it generates a decision boundary optimized to best separate images of the target emotion from images of the distractor emotions. Logistic regression works by estimating the probability that a given input feature set belongs to a particular class, in this case, a specific facial emotion. The model assigns weights to each feature, and these weights are adjusted during training to maximize the accuracy of the emotion classification.

The training and testing sets were pseudo-randomly selected multiple times to ensure that every image in our dataset was included in the held-out test set at least once. This method helps in assessing the generalization capability of the decoder across different subsets of the data. For each iteration, we trained the classifier on a portion of the data and tested its performance on the remaining unseen images.

We computed the binary task performances by calculating the percent correct (PC) score for each pair of possible binary tasks given an image. For example, if an image depicted the target emotion (e.g., Anger), we calculated the PC score for the binary classification task between the target emotion (Anger) and each distractor emotion (e.g., Joy, Fear). The PC score is a measure of how accurately the classifier can distinguish the target emotion from the distractor emotions (Figure 3.2).

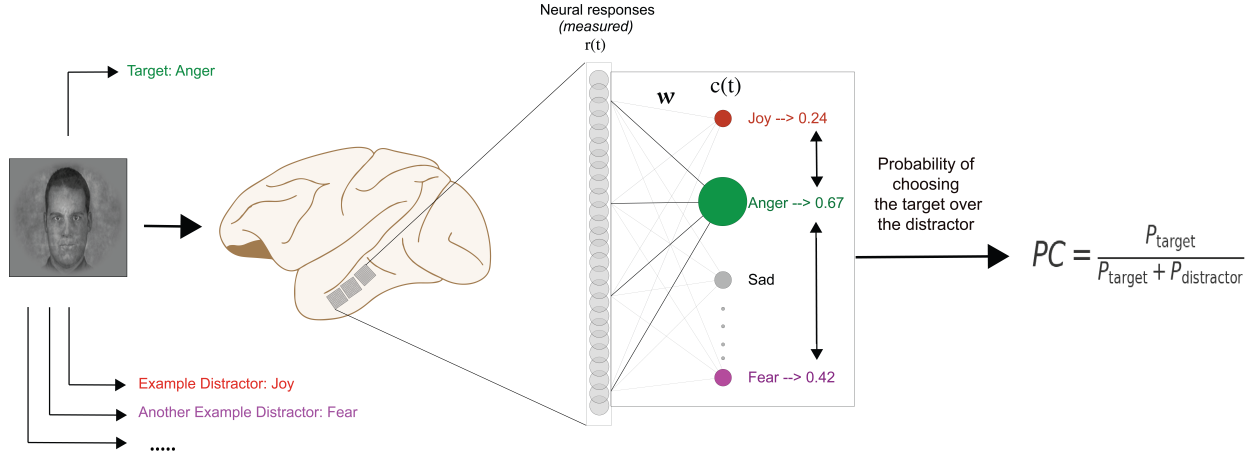


Figure 3.2: **Conceptual schematic of Decoding** Illustration of the logistic regression-based classifier used to decode emotion labels from IT neural population activity. The classifier is trained to predict the probability of selecting the target emotion (e.g., Anger) over various distractor emotions (e.g., Joy, Fear). The neural responses are split into training and test sets to predict the labels in the test set. The performance is evaluated by calculating the percent correct score for each binary classification task between the target and distractor emotions.

This process involves several steps:

1. **Data Splitting:** The dataset is divided into training and test sets multiple times, ensuring that every image serves as part of the test set in at least one iteration. This technique, known as cross-validation, helps to evaluate the decoder's performance robustly.
2. **Training the Classifier:** For each split, the logistic regression decoder is trained on the training set. During training, the decoder learns the weights that best separate the target emotion from the distractor emotions based on the neural activity patterns.
3. **Testing the Classifier:** After training, the decoder's performance is tested on the held-out test set. The decoder uses the learned weights to predict the class labels (emotions) of the test images.
4. **Performance Evaluation:** The decoder's accuracy is evaluated by calculating the percent

correct score for each binary classification task. This involves comparing the predicted emotion labels with the actual labels to determine how often the decoder correctly identifies the target emotion over the distractor emotions.

By using this logistic regression-based classifier, we can quantify how well the IT neural population activity predicts facial emotion recognition behavior. This approach provides insights into the neural coding strategies employed by the IT cortex and their relevance to behavior.

3.4 Results

3.4.1 IT accurately categorizes different emotions

After collecting and curating the data and building our classifier decoders, we tested our hypotheses, beginning with the first hypothesis to determine if the macaque IT population can predict facial emotion recognition behavior. We previously established that our neural data is reliable (Figure 3.1.C). However, to evaluate its accuracy, each stimulus was presented for 100 ms with 100 ms of neutral gray background interleaved between images. This allowed us to select different time bins to build various decoders. Based on previous studies on the visually-driven latency of IT neurons [75], we initially chose the 70-170 ms time bin from image onset to assess whether our neural data is visually responsive.

To decode the data collected from monkeys, we used linear classifiers averaged for this specific time bin. We calculated all four metrics [74]—ranging from emotion level (E1) to

image level (I1)—mirroring the approach used for behavioral data (see section 2.3). We then generated all four corresponding matrices (Figure 3.3).

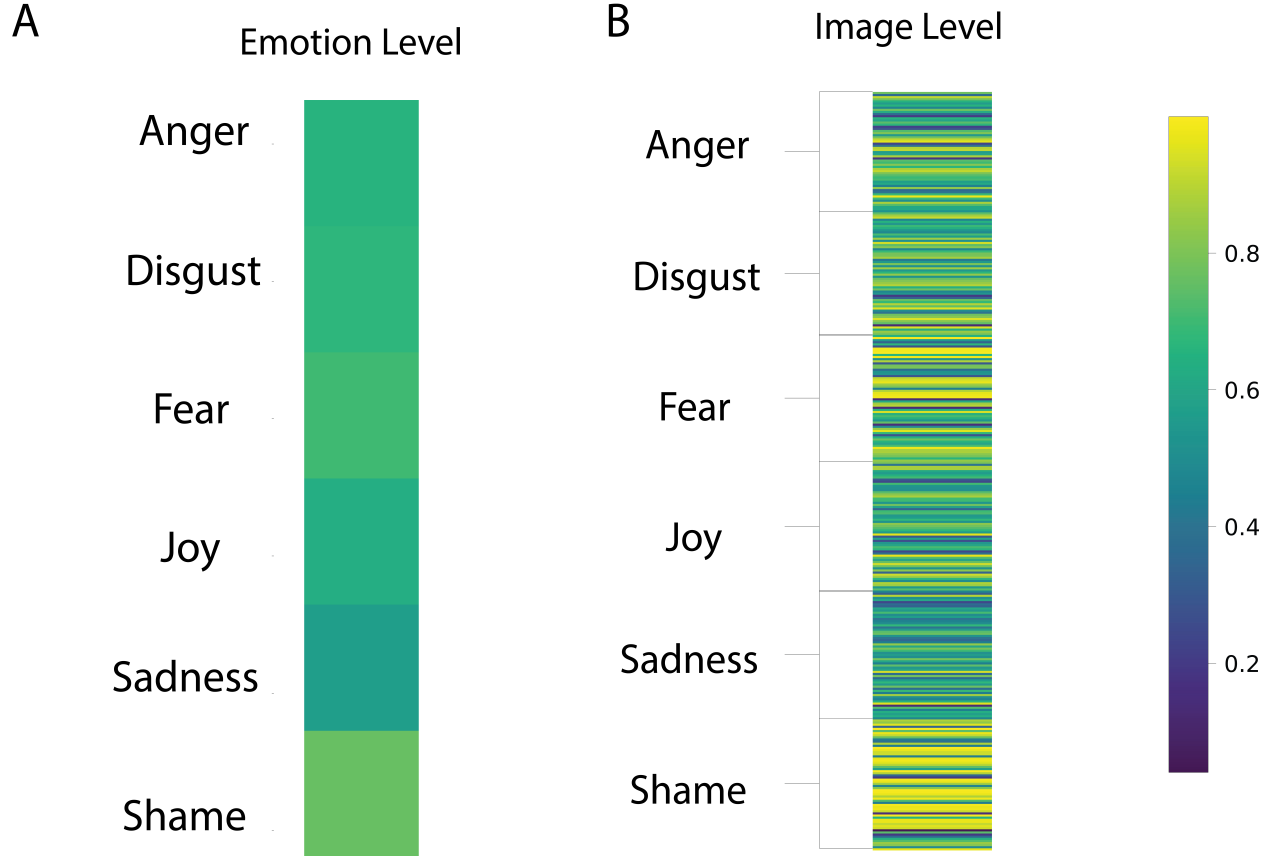


Figure 3.3: **decoding emotion categories from the neural population.** **A.** Showing the 6-dimensional Emotion level neural signature for the IT population, each value represents the accuracy grouped across all images of that emotion, for all distractor emotions. **B.** The Image-level neural signature Each line in the signature is the accuracy for an image, averaged across distractors.

The results showed that the neural data could decode emotions with an accuracy exceeding the chance level (62%). This finding indicates that our neural data can predict facial emotion recognition within this time bin, allowing us to reject the null hypothesis for the first part of our hypothesis. With this validation, we moved to the second step, determining if this accuracy is consistent with monkey behavior. However, before proceeding, we aimed to

identify if there is a time bin in which we can build a more consistent and accurate decoder.

3.4.2 Comparison of macaque IT-based predictions with monkey behavior

To find the most accurate and consistent decoder aligned with monkey behavior, we explored all possible time bins (205) to optimize the decoder’s performance (Figure 3.4.A). By adjusting the time bin parameters and re-evaluating the classifier’s accuracy, we aimed to identify the optimal time window that maximizes the predictive power of the IT neural population. This process involved systematically shifting the start and end points of the time bins and recalculating the decoder accuracy for each configuration.

After determining the accuracy of each decoder, we calculated the similarity between each decoder’s accuracy and the monkey’s behavioral data. To quantify this similarity, we used the "consistency" measurement [74] [76]. Consistency is computed as a noise-adjusted correlation of behavioral signatures between the two metrics (monkey behavior and macaque IT population). Specifically, we computed the correlation between monkey behavior and the macaque IT population for all possible time bins using Pearson correlation, and corrected it with the Spearman-Brown corrected internal consistency of the IT population for that time bin.

To obtain the IT population ceiling, we randomly split all possible trials for each image into two halves and applied each behavioral metric to each half. This resulted in two independent estimates of the system’s predicted behavioral signature concerning that metric. The self-consistency of the system is obtained by calculating the Pearson correlation between these two

estimates of the predicted behavioral signature, corrected using the Spearman-Brown formula. This self-consistency measure indicates the reliability of the predicted behavioral signature given the amount of data collected and is noted as the IT population self-consistency.

Similarly, we computed the primate behavioral ceiling by applying the same procedure to the monkey behavioral data. The purpose of using the primate ceiling is to account for the unpredictable variances in behavioral patterns due to differing factors—“noise” not reproducible by the experimental condition.

We computed the decoder accuracy and the consistency between monkey behavior and the macaque IT population for all time bins for both emotion-level and image-level metrics. To visualize this, we plotted the consistency of each decoder against the decoding accuracy for that specific decoder. This allowed us to identify the optimal time bin, which is the time bin that falls within the range of monkey accuracy and also shows high consistency with monkey behavior (Figure 3.4.B,C).

By following this detailed and systematic approach, we ensured that the selected time bin provided the most accurate and reliable prediction of facial emotion recognition behavior in macaques. This method not only optimized the decoder’s performance but also validated the consistency of the neural predictions with actual behavioral responses, thereby supporting the robustness of our findings.

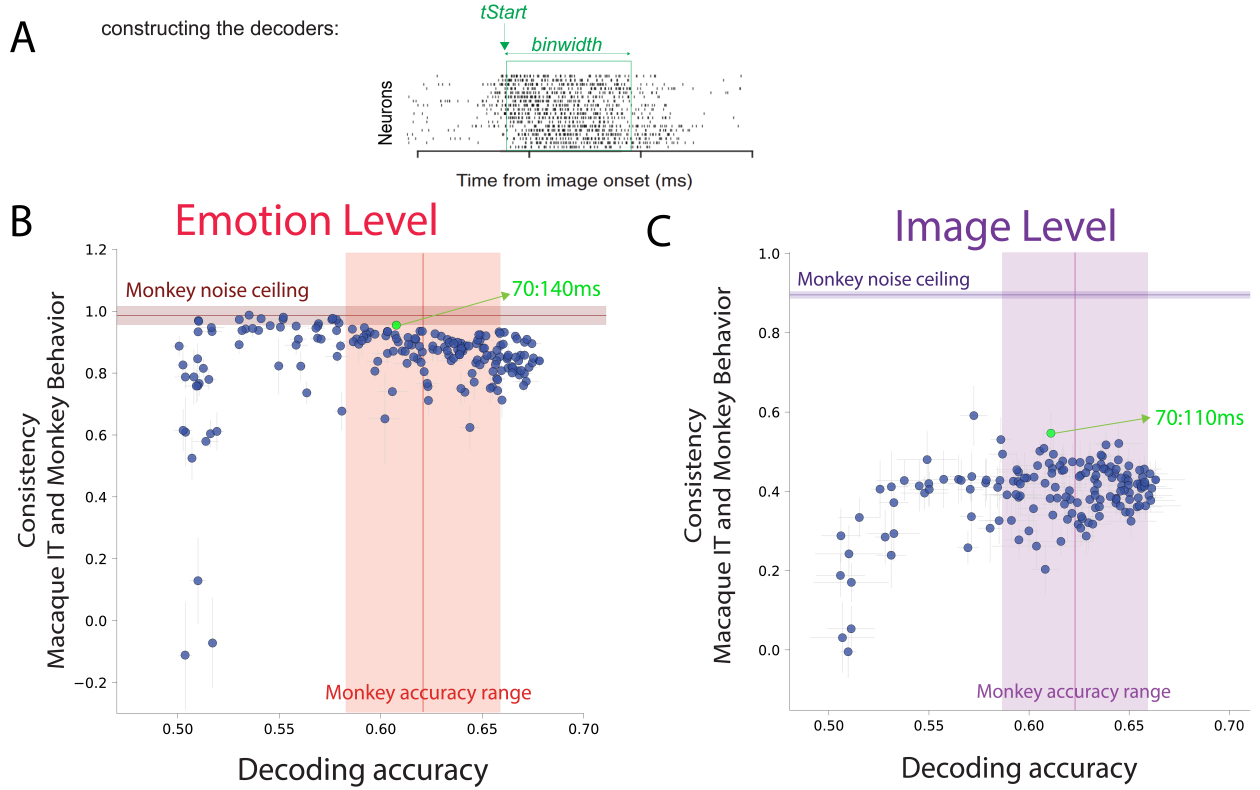


Figure 3.4: **Consistency of Monkey IT population with Monkey Behavior for different time bins** **A.** Illustration of neuronal activity over time from image onset, highlighting the different time bins utilized for constructing various decoders. **B.** The decoding accuracy at the emotion level for each time bin is plotted on the x-axis, while the consistency between the monkey IT population and monkey behavior (corrected for internal consistency) is plotted on the y-axis. The optimal time bin for constructing the decoder, within the monkey behavioral accuracy range and exhibiting the highest consistency (70-140 ms), is indicated in green. **C.** Similarly, decoding accuracy at the image level is analyzed, with the optimal time bin for predicting image-level metrics in monkey behavior identified as 70-110 ms, shown in green.

Upon identifying the optimal time bin for predicting the consistency of each metric (we found 70-140 ms for the emotion-level and 70-110 ms for the image-level metric), we proceeded to develop the most effective decoder for each respective metric. Utilizing these tailored decoders, we predicted the accuracy again and conducted a comparative analysis of the monkey's behavior, considering both emotion-level and image-level metrics. Our findings

indicate a stronger correlation at the emotion-level metric compared to the image-level metric. This observation aligns with the expectation that correlation diminishes when transitioning from a coarser metric to a finer-grained metric, as illustrated in Figure 3.5.

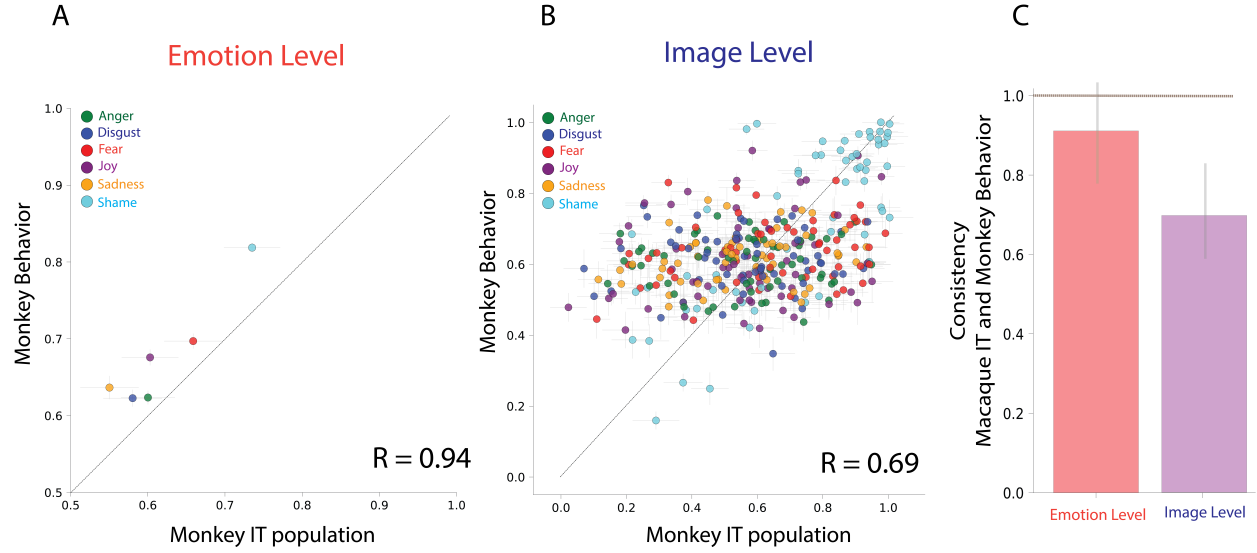


Figure 3.5: Consistency between Monkey IT population and Monkey Behavior for the most consistent time bins **A.** Scatter plot showing the consistency between the monkey IT population and monkey behavior at the emotion level. The analysis was performed using the 70-140 ms time bin and resulted in a corrected Pearson R-value of 0.94 (p -value = 0.016). **B.** Scatter plot depicting the consistency between the monkey IT population and monkey behavior at the image level. This analysis was performed using the 70-110 ms time bin and resulted in a corrected Pearson R-value of 0.69 ($p = 5.6e-12$). **C.** Bar plot summarizing the consistency metrics at both the emotion and image levels, highlighting the higher consistency at the emotion level compared to the image level.

By this analysis, we found that the accuracy of categorization by the monkey IT population is highly and significantly correlated with monkey behavior. The confirmation provides a robust foundation to proceed with testing the next hypothesis which aims to determine whether the accuracy of categorization made by the monkey IT population is also consistent with human behavior. By establishing this consistency, we can further validate the use of the monkey model for understanding the neural basis of social cognitive processes in humans.

3.4.3 Comparison of neural models with Human Behavior

Following the methodology established in the previous section, we extended our analysis to compare the monkey IT population predictions with human behavior. This approach allows us to test the main Hypothesis, which aims to determine whether the predictions made by the monkey IT population are consistent with human behavior.

We employed the same systematic exploration of time bins to optimize the decoder’s performance, but this time focusing on its consistency with human behavioral data. As illustrated in Figure 3.6, we analyzed all possible time bins to identify those that maximize both decoding accuracy and consistency with human behavior.

For the emotion-level analysis (Figure 3.6.B), we found that the optimal time bin for constructing the decoder that best aligns with human behavioral accuracy and exhibits the highest consistency was 110-170 ms post-stimulus onset. This time bin is slightly later than the one identified for monkey behavior (70-140 ms), which may suggest subtle differences in the temporal dynamics of emotion processing between species.

At the image-level (Figure 3.6.C), the optimal time bin for predicting human behavior was identified as 70-170 ms. This broader time window encompasses the one found for monkey behavior (70-110 ms), indicating a potentially more extended period of relevant neural activity when predicting finer-grained human responses.

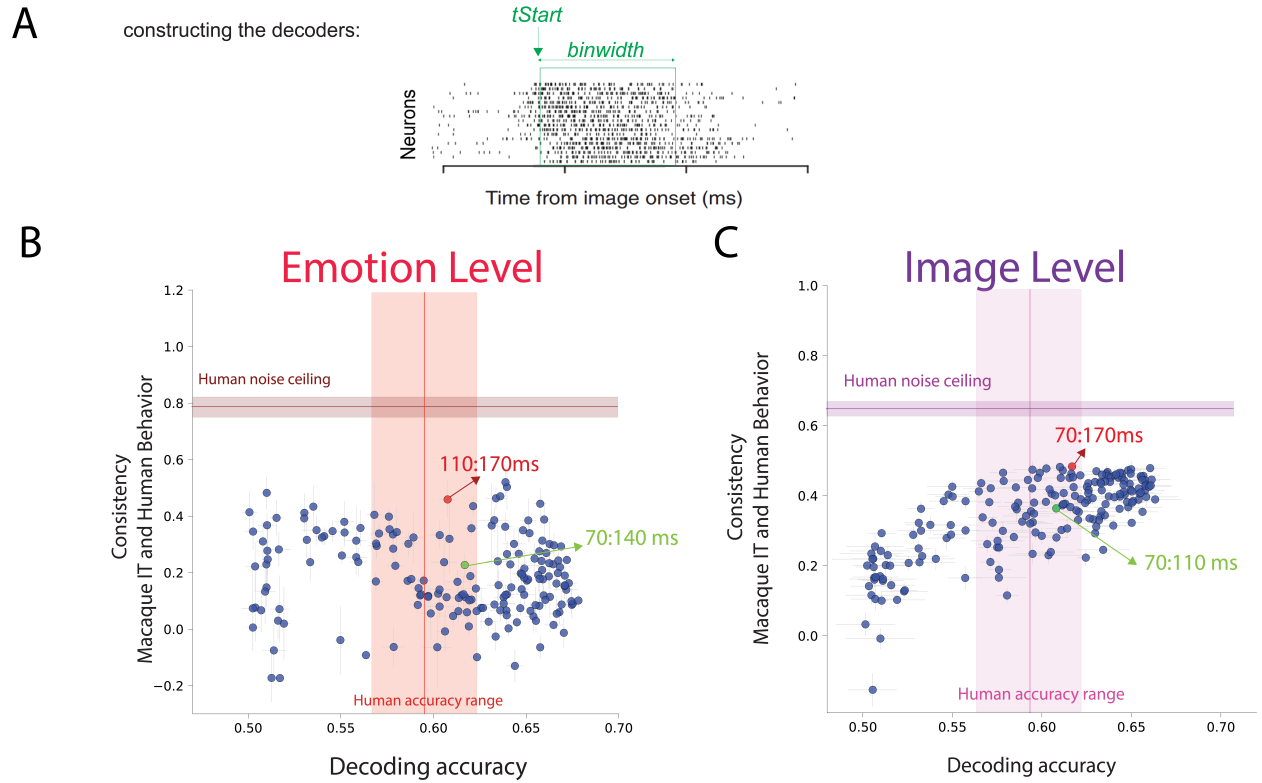


Figure 3.6: Consistency of Monkey IT population with Human Behavior for different time bins **A.** Illustration of neuronal activity over time from image onset, highlighting the different time bins utilized for constructing various decoders. **B.** The decoding accuracy at the emotion level for each time bin is plotted on the x-axis, while the consistency between the monkey IT population and human behavior (corrected for internal consistency) is plotted on the y-axis. The optimal time bin for constructing the decoder, within the human behavioral accuracy range and exhibiting the highest consistency (110-170 ms), is indicated in green. **C.** Similarly, decoding accuracy at the image level is analyzed, with the optimal time bin for predicting image-level metrics in human behavior identified as 70-170 ms, shown in green.

Using these optimized time bins, we then analyzed the consistency between the monkey IT population predictions and human behavior for both emotion-level and image-level metrics (Figure 3.7). Our findings reveal significant correlations at both levels, though with some notable differences compared to the monkey behavior results:

Emotion-level consistency: Using the 110-170 ms time bin, we observed a corrected

Pearson R-value of 0.49 ($p = 0.027$) between the monkey IT population predictions and human behavior. While statistically significant, this correlation is lower than that observed with monkey behavior ($R = 0.94$). Image-level consistency: Analysis using the 70-170 ms time bin yielded a corrected Pearson R-value of 0.43 ($p = 2.7\text{e-}09$). This correlation is also significant but lower than that seen with monkey behavior ($R = 0.69$).

The bar plot in Figure 3.7 summarizes these consistency metrics, highlighting the overall consistency observed at each level. Notably, unlike in the monkey behavior comparison, we do not see a substantial difference in consistency between the emotion-level and image-level metrics when compared with human behavior.

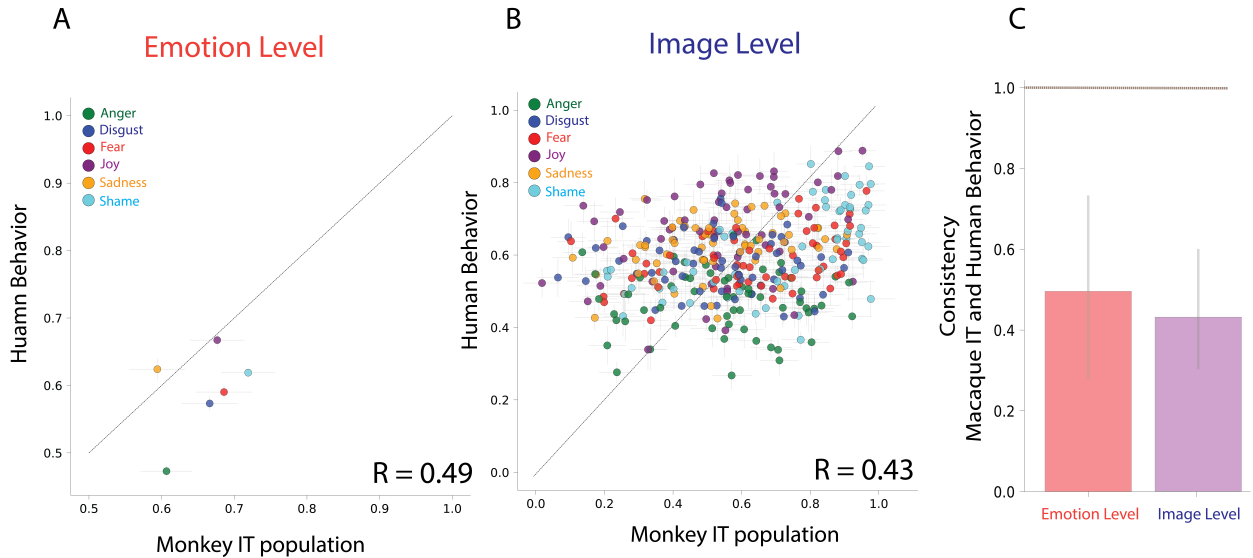


Figure 3.7: Consistency between Monkey IT population and Human Behavior for the most consistent time bins **A.** Scatter plot showing the consistency between the monkey IT population and human behavior at the emotion level. The analysis was performed using the 110-170 ms time bin and resulted in a corrected Pearson R-value of 0.49 ($p = 0.027$). **B.** Scatter plot depicting the consistency between the monkey IT population and human behavior at the image level. This analysis was performed using the 70-170 ms time bin and resulted in a corrected Pearson R-value of 0.43 ($p = 2.7\text{e-}09$). **C.** Bar plot summarizing the consistency metrics at both the emotion and image levels, highlighting the overall consistency observed at each level.

These results provide evidence to reject the null hypothesis, demonstrating that the monkey IT population accuracy is consistent with human behavior to a significant degree.

The consistency between monkey neural data and human behavior, albeit lower than with monkey behavior, underscores the value of the macaque model in understanding the neural basis of social cognitive processes in humans. It suggests a shared neural substrate for facial emotion recognition across primates.

This consistency in neural predictions and behavior not only validates the use of the monkey model but also provides a bridge for translating findings from animal studies to human cognition.

3.5 Inferences

The findings presented in this chapter provide valuable insights into the neural mechanisms underlying facial emotion recognition (FER) in primates, with significant implications for both basic neuroscience and clinical applications. Through the analysis of large-scale, chronic multi-electrode recordings in the macaque ventral visual cortex, we have uncovered several key points that shed light on the neural basis of emotion processing and its conservation across species.

Our results demonstrate that population activity in the macaque inferior temporal (IT) cortex can indeed categorize facial emotions with accuracy significantly above chance level. This finding supports the crucial role of the IT cortex in facial emotion processing, aligning with previous research highlighting its importance in object and face recognition.

The analysis of different time bins revealed optimal decoding periods for emotion recognition, with slight variations between monkey and human behavior predictions. For monkey behavior, the optimal time bin was 70-140 ms post-stimulus onset for emotion-level decoding and 70-110 ms for image-level decoding. In contrast, for human behavior predictions, the optimal time bins were 110-170 ms and 70-170 ms for emotion-level and image-level decoding, respectively. These temporal differences suggest subtle variations in the processing speed and integration of emotional information between species. The later optimal time bin for human behavior predictions might indicate a more complex or nuanced processing of emotional cues in humans, possibly reflecting the influence of higher-order cognitive processes.

A striking finding of this study is the high consistency between macaque IT population activity and monkey behavioral responses in FER tasks. The strong correlations observed, particularly at the emotion level ($R = 0.94$), provide robust evidence that IT cortex activity is a reliable predictor of facial emotion recognition behavior in macaques. This close alignment between neural activity and behavior underscores the IT cortex's central role in translating visual inputs into meaningful emotional categories that guide behavioral responses.

Perhaps one of the most intriguing aspects of this study is the significant correlation found between macaque IT population activity and human behavior in FER tasks. Although the consistency was lower than that observed with monkey behavior ($R = 0.49$ for emotion-level and $R = 0.43$ for image-level), these correlations remain statistically significant. This finding suggests a considerable degree of conservation in the neural mechanisms underlying facial emotion recognition. This consistency validates the use of macaque models in studying human

emotion processing and provides a foundation for translating insights from animal studies to human cognition.

The varying levels of consistency observed between neural predictions and behavior at different levels of analysis (emotion-level vs. image-level) hint at a hierarchical processing of emotional information within the visual system. The generally higher consistency at the emotion level compared to the image level suggests that the IT cortex may be more involved in categorical representations of emotions rather than fine-grained, image-specific details. This hierarchical organization aligns with current theories of visual processing, where higher-order regions like the IT cortex are thought to represent increasingly abstract and behaviorally relevant information.

The success in decoding emotional categories from neural activity and predicting behavior across species has significant implications for computational models of emotion processing. These findings provide a biological benchmark against which artificial neural networks and other computational models can be evaluated and refined. By aligning computational predictions with observed neural patterns and behavior, we can develop more accurate and biologically plausible models of facial emotion recognition.

The identification of neural biomarkers for facial emotion recognition in the IT cortex opens up new possibilities for clinical applications. These biomarkers could potentially be used to develop early diagnostic tools for disorders affecting social cognition, such as autism spectrum disorders or schizophrenia. Moreover, understanding the precise neural dynamics involved in emotion processing could guide the development of targeted interventions or

brain-computer interfaces aimed at improving emotional recognition and social functioning in affected individuals.

While this study provides valuable insights, it also highlights areas for future research. The differences in consistency between monkey and human behavior predictions suggest that additional factors may influence human emotion recognition, such as cultural context or higher-order cognitive processes not fully captured by IT cortex activity alone. Future studies could explore the contribution of other brain regions, such as the amygdala or prefrontal cortex, to emotion recognition processes. Additionally, investigating how these neural representations change with learning or in different social contexts could provide a more comprehensive understanding of the dynamic nature of emotion processing in the brain.

In conclusion, this chapter presents compelling evidence for the central role of the IT cortex in facial emotion recognition. These findings not only advance our understanding of the neural basis of social cognition but also pave the road for transnational research aimed at developing novel diagnostic and therapeutic approaches for disorders affecting emotional processing. As we continue to unravel the complex neural mechanisms underlying facial emotion recognition, we move closer to a comprehensive understanding of how the brain interprets and responds to the emotional world around us.

Chapter 4

Discussion

The primary aim of this study was to build a non-human primate model to dissect the neural mechanisms of facial emotion processing. One of the main reasons we chose facial emotion expressions is that this relatively automatic visuocognitive feature, which neurotypical adults take for granted, shows significant differences in children and adults with autism [23]. By identifying the neural biomarkers of facial emotion recognition, we can inspire a range of experiments to help autistic individuals.

4.1 Limitations of the current study

While this study provides valuable insights into the neural mechanisms of facial emotion recognition across species, it is important to acknowledge several limitations that could impact the interpretation and generalizability of the results:

Sample size: The study involved only two macaque subjects for neural recordings. While

this is common in non-human primate research due to ethical and practical constraints, it limits the generalizability of the findings. A larger sample size would provide more robust and representative data.

Limited emotion set: The study focused on six basic emotions (anger, disgust, fear, joy, sadness, and shame). This may not capture the full spectrum of human emotional experiences, particularly more complex or nuanced emotions. Future studies could expand to include a broader range of emotions or emotional blends. Static images: The use of static facial images, while allowing for controlled experimental conditions, may not fully represent real-world dynamic facial expressions. This could limit the ecological validity of the findings, as emotion recognition in natural settings often involves dynamic cues.

Comparison limitations: While the study found significant correlations between macaque neural activity and human behavior, there are inherent limitations in directly comparing them. Differences in cognitive processes, social experiences, and evolutionary adaptations between humans and macaques may influence emotion recognition in ways not fully captured by this study.

Potential confounds in human data collection: The human behavioral data was collected through online platforms (Amazon Mechanical Turk) and limited lab testing. This method, while efficient, may introduce variability due to uncontrolled testing environments and potential differences in display quality or viewing conditions across participants.

Focus on IT cortex: While the IT cortex is crucial for facial processing, emotion recognition involves a network of brain regions. The study's focus on the IT cortex, while in-

formative, may not capture the full complexity of emotion processing in the brain, particularly higher-order cognitive influences.

Temporal limitations: Although the study analyzed various time bins, the relatively short presentation time (100 ms) and analysis window may not capture longer-term temporal dynamics involved in emotion processing, especially for more subtle or complex emotions.

Potential for overfitting: The process of selecting optimal time bins for decoding, while methodologically sound, could potentially lead to some degree of overfitting. This risk should be considered when interpreting the strength of correlations between neural activity and behavior.

Limited exploration of individual differences: The study focused on group-level analyses and did not extensively explore individual differences in emotion recognition abilities or neural representations, which could provide additional insights into the variability of emotion processing.

Absence of clinical populations: The study was conducted on neurotypical subjects and macaques. Including clinical populations, such as individuals with autism spectrum disorders, could provide valuable comparative data and enhance the translational potential of the findings.

Addressing these limitations in future research would further strengthen our understanding of the neural basis of facial emotion recognition and its conservation across species. Despite these constraints, the current study provides a solid foundation for future investigations in this important area of social neuroscience.

4.2 Clinical Implications

Our research has significant clinical implications, particularly for adults with Autism Spectrum Disorder (autistic individuals referred to as ASD). We found a higher correlation between monkey behavioral data and neurotypical human data than with ASD data in the in-lab data collected previously at MIT [68]. This trend was also observed in neural data, where the IT neural population in the macaque IT cortex correlated more with neurotypical behavior than ASD behavior. This suggests that monkeys serve as better models for neurotypical individuals. By developing the Non-Human Primate model, we can improve the diagnosis of neurotypical individuals and predict ASD behavior more accurately, aiding in the diagnosis of autistic subjects.

4.2.1 Improving Diagnosis

The integration of artificial neural networks (ANNs) with behavioral data from individuals with autism spectrum disorder (ASD) offers a promising approach to enhancing diagnostic accuracy and understanding the underlying mechanisms of facial emotion recognition deficits. A recent study employs pre-trained ANNs to model and predict happiness scores, comparing these predictions between control and ASD groups [77]. By leveraging these models, we aim to identify specific image-level representations that elicit different responses between neurotypical individuals and those with ASD.

The ANN-driven image selection method was particularly effective in highlighting the differences in facial emotion recognition between the two groups. As depicted in Figure

4.1, the ANN-driven selection (green bars) resulted in a more pronounced difference in the probability of choosing a "happy" emotion between control and ASD participants compared to random image selection (gray bars). This suggests that ANNs can effectively identify images that accentuate behavioral discrepancies, providing a more sensitive diagnostic tool.

The inset in Figure 4.1 further reinforces this finding, showing a significant increase in the differential happiness score ($\Delta[\text{Ctrl} - \text{ASD}]$) when using ANN-selected images. This highlights the potential of ANN-driven methodologies to refine diagnostic criteria and enhance the sensitivity of behavioral assessments.

A significant correlation was shown between different ANN models and the behavioral performance of control versus ASD groups. Notably, models such as CLIP and ViT, which have high correlation values with control participants, also exhibit significant differences in $\Delta[\text{Ctrl} - \text{ASD}]$. This indicates that certain ANN architectures are more adept at capturing the nuanced facial emotion recognition patterns seen in neurotypical individuals compared to those with ASD.

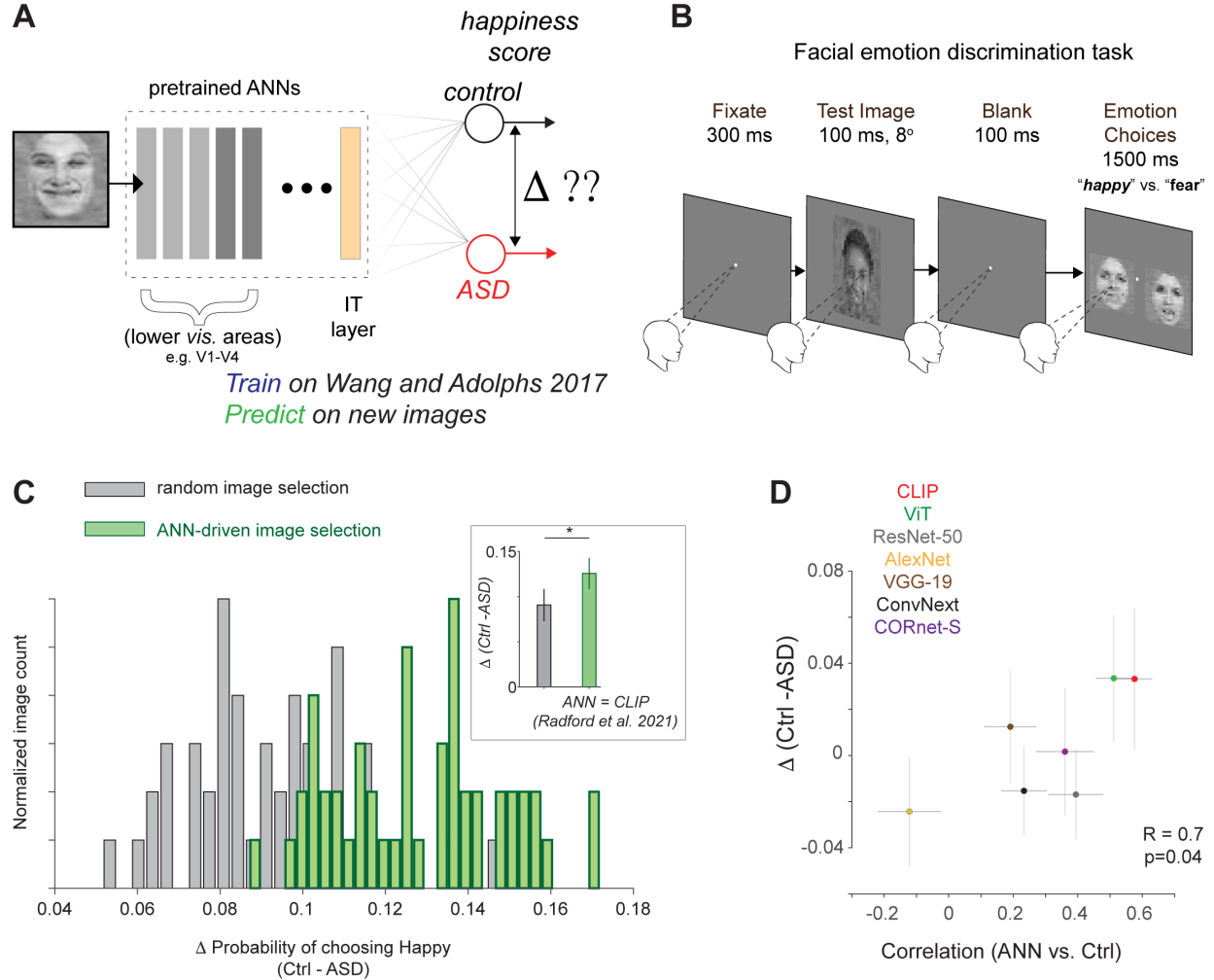


Figure 4.1: **A.** Our method involves fine-tuning artificial neural networks (ANNs) using data from both ASD and NT participants from Wang and Adolphs (2017) to predict the Δ behavior for a new image set. **B.** For the new image set ($n=80$), human participants, including both NT (Control; $n=13$) and autistic individuals (ASD; $n=12$), viewed a Test image (face) for 100 ms within their central ~ 8 degrees, followed by a choice screen displaying two images: one extremely happy and one extremely fearful face. Participants were required to press a key to identify the emotion in the Test image (fearful or happy?). **C.** The distribution of Δ score differences between NT (Ctrl) and autistic (ASD) adults for two image sets is shown. The green bars represent 24 images selected based on ANN efficiency, while the gray bars represent 24 randomly selected images. The inset indicates that the average difference is statistically significant. **D.** ANN diagnostic efficiency (the difference between the NT vs. ASD gap for the diagnostic (ANN-predicted) image set and the randomly selected image set) is presented as a function of the ANN models' original predictivity of NT behavior. The figures are adapted from K Mukherjee, et al [77]

The findings from this study underscore the potential for integrating ANN-driven image selection and prediction models into clinical practice. By identifying specific images that elicit differential responses, clinicians can develop more targeted and accurate diagnostic tools. This approach moves beyond traditional broad-category descriptors and focuses on individual image-level analysis, offering a more granular and precise assessment of facial emotion recognition deficits.

To build on these findings, future research should explore the application of ANN models to a broader range of emotions and contexts. Expanding the dataset to include more diverse and subtle emotional expressions will help in capturing the full spectrum of facial emotion recognition capabilities and deficits. Additionally, integrating longitudinal studies will provide insights into the developmental trajectory of these deficits and the effectiveness of targeted interventions over time.

Moreover, investigating the neural correlates of ANN-driven predictions through techniques such as functional MRI or electrophysiological recordings in both human and non-human primates will further elucidate the underlying neural mechanisms. This multi-faceted approach, combining behavioral, computational, and neural data, will enhance our understanding of ASD and improve diagnostic and therapeutic strategies.

4.2.2 Symptom Mitigation

In this section, we discuss how artificial neural network (ANN) models can be used to mitigate symptoms of autism spectrum disorder (ASD) by controlling visual perception phenotypes.

We can illustrate an innovative approach where ANNs are employed to synthesize new images that minimize behavioral differences between individuals with ASD and neurotypical (NT) individuals. This process can be achieved through a series of iterative updates and model training phases.

Figure 4.2 outlines a method where an original image undergoes modifications through a pre-trained Generative Adversarial Network (GAN), specifically referred to as GANmut. This GAN is used to adjust the latent space coordinates (θ, ρ) of the image. The adjustments are made using gradient descent, a technique to iteratively minimize the difference between the predicted responses of individuals with ASD and NT individuals. The synthesized images are evaluated using a pre-trained vision model to predict the behavioral differences.

Latent Space Manipulation: The latent space coordinates of the original image are adjusted using gradient descent to produce new images (iterations 1 and 2) that gradually reduce the predicted behavioral differences between ASD and NT responses.

Pretrained Vision Model: These synthesized images are passed through a pre-trained vision model to evaluate and predict the response probabilities for both ASD and NT individuals. The goal is to minimize the predicted differences (Δ) in responses.

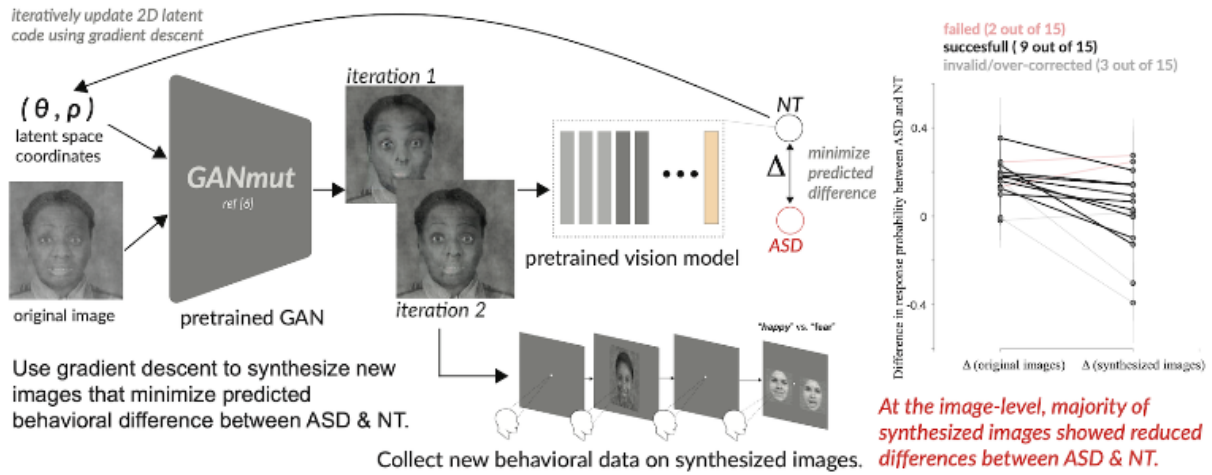


Figure 4.2: This figure illustrates the process of using a pre-trained Generative Adversarial Network (GAN) to modify original images and minimize behavioral differences between individuals with ASD and neurotypical (NT) individuals. The latent space coordinates (θ, ρ) of the original image are iteratively updated using gradient descent to produce new images. These synthesized images are then evaluated using a pre-trained vision model to predict response probabilities for both ASD and NT individuals. The objective is to minimize the predicted difference (Δ) in responses. The results, shown on the right, indicate that the majority of synthesized images demonstrated reduced differences in response probabilities between ASD and NT groups, highlighting the potential of this approach in mitigating ASD-related perceptual differences. Out of 15 cases, 9 were successful, 2 failed, and 3 were invalid or over-corrected.

Behavioral Data Collection: The new synthesized images are then used to collect behavioral data, focusing on minimizing the differences in response probability between ASD and NT individuals for binary emotion discrimination tasks, such as distinguishing between “happy” and “fear.”

The results displayed in Figure 4.2 demonstrate the effectiveness of this approach. The majority of synthesized images showed reduced differences in response probabilities between ASD and NT individuals, indicating a successful minimization of behavioral differences. Specifically:

Success Rate: Out of 15 cases, 9 were successful in reducing the difference.

Failure and Over-correction: 2 cases failed, and 3 cases resulted in invalid or over-corrected responses.

The image-level majority of synthesized images exhibited a significant reduction in the behavioral differences between ASD and NT groups, thereby indicating a promising direction for mitigating ASD-related perceptual differences.

This methodology suggests that ANN models, specifically GANs, can be effectively used to adjust visual stimuli in a way that mitigates the perceptual differences experienced by individuals with ASD. By iteratively refining images to minimize predicted behavioral differences, it is possible to create visual stimuli that are perceived more similarly by ASD and NT individuals, potentially easing some of the social communication challenges faced by those with ASD. This approach not only enhances our understanding of the neural and perceptual mechanisms underlying ASD but also opens up new avenues for developing targeted interventions and supports for individuals with autism. Future work should focus on refining these models and exploring their applicability across a broader range of visual and emotional stimuli to fully harness their potential in clinical settings.

4.3 Future Direction

The findings of this study open up several exciting avenues for future research in the field of facial emotion recognition and its neural underpinnings. Based on our results and the potential applications of artificial neural networks (ANNs) in this domain, we propose the

following directions for future investigations:

1. Leverage ANNs to screen and select facial image stimuli that maximize predicted behavioral differences between neurotypical individuals and those with autism spectrum disorders (ASD). Develop deep image synthesis methods to create optimized stimuli for behavioral testing and diagnosis, as well as images that minimize differences between neurotypical and ASD responses for potential therapeutic applications.

2. Continue developing and refining ANN models to better match primate brain architecture and function, particularly focusing on the integration of feedback connections and temporal dynamics. Investigate how these improved models can provide more accurate predictions of both neurotypical and atypical facial emotion processing.

3. Conduct macaque electrophysiology experiments to test the specific neural hypotheses generated by ANN models, particularly focusing on the inferior temporal (IT) cortex. Explore the temporal components of IT population responses in facial emotion signal processing, which may not be fully captured by current ANN models.

4. Perform simultaneous neural recordings in the IT cortex and amygdala to directly test the hypothesis that higher-level visual cortices drive facial affect signals observed in the amygdala. Investigate the functional influence of the IT cortex on value-coding (VS) and face-selective (VF) neurons in the amygdala, exploring their complementary coding schemes.

5. Design and conduct finer-grained causal perturbation experiments to elucidate the role of the IT cortex in facial emotion processing and its influence on amygdala activity.

6. Examine how IT cortical activity is modulated by feedback projections from the

amygdala and other regions such as the ventrolateral prefrontal cortex (PFC).

7. Estimate noise levels in sensory cortical signal variability in children with autism to test the hypothesis of increased sensory noise affecting behavior and learning. Investigate how sensory noise interacts with learning new tasks in both neurotypical individuals and those with ASD.

8. Utilize ANN models to generate more diagnostic image sets for future studies, simulating various levels and types of noise under different learning regimes. Use these model-driven hypotheses to guide experimental designs and predict effect sizes, optimizing the efficiency and power of future studies.

9. Conduct longitudinal studies to track the development of facial emotion recognition abilities and corresponding neural patterns in neurotypical individuals and those with ASD. Investigate how early interventions based on ANN-derived insights might influence the trajectory of facial emotion processing development.

10. Develop and test potential cognitive therapies based on the insights gained from ANN models and neural recordings. Explore the use of ANN-based tools for early diagnosis and intervention in ASD and other disorders affecting social cognition.

11. Extend the research to diverse cultural contexts to examine the universality and cultural specificity of facial emotion recognition processes and their neural correlates.

12. Investigate how facial emotion recognition integrates with other sensory modalities (e.g., voice, body language) at both behavioral and neural levels. Develop multimodal ANN models that capture the complexity of real-world emotion recognition.

By pursuing these future directions, researchers can build upon the foundation laid by this study to further our understanding of facial emotion recognition, its neural basis, and its alterations in conditions like ASD. This knowledge has the potential to significantly impact both basic neuroscience and clinical applications in social cognition and emotional processing.

Bibliography

- [1] Albert Mehrabian. “Communication without words”. In: *Communication theory*. Routledge, 2017, pp. 193–200.
- [2] Kathrin Kaulard et al. “The MPI facial expression database—a validated database of emotional and conversational facial expressions”. In: *PloS one* 7.3 (2012), e32321.
- [3] Byoung Chul Ko. “A brief review of facial emotion recognition based on visual information”. In: *sensors* 18.2 (2018), p. 401.
- [4] Lisa Feldman Barrett et al. “Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements”. In: *Psychological science in the public interest* 20.1 (2019), pp. 1–68.
- [5] Alexander Todorov et al. “Understanding evaluation of faces on social dimensions”. In: *Trends in cognitive sciences* 12.12 (2008), pp. 455–460.
- [6] Mark Williams et al. “Look at me, I’m smiling: Visual search for threatening and nonthreatening facial expressions”. In: *Visual cognition* 12.1 (2005), pp. 29–50.
- [7] Maximilian Riesenhuber and Tomaso Poggio. “Hierarchical models of object recognition in cortex”. In: *Nature neuroscience* 2.11 (1999), pp. 1019–1025.
- [8] Kohitij Kar and James J DiCarlo. “The Quest for an Integrated Set of Neural Mechanisms Underlying Object Recognition in Primates”. In: *arXiv preprint arXiv:2312.05956* (2023).
- [9] John G Keller et al. “Object recognition based on human saccadic behaviour”. In: *Pattern Analysis & Applications* 2 (1999), pp. 251–263.
- [10] KE Ameen-Ali, MJ Eacott, and Alexander Easton. “A new behavioural apparatus to reduce animal numbers in multiple types of spontaneous object recognition paradigms in rats”. In: *Journal of neuroscience methods* 211.1 (2012), pp. 66–76.
- [11] Ayhan Şık et al. “Performance of different mouse strains in an object recognition task”. In: *Behavioural brain research* 147.1-2 (2003), pp. 49–54.
- [12] Abdelkader Ennaceur. “One-trial object recognition in rats and mice: methodological and theoretical issues”. In: *Behavioural brain research* 215.2 (2010), pp. 244–254.

- [13] William Overman et al. “Object recognition versus object discrimination: comparison between human infants and infant monkeys.” In: *Behavioral Neuroscience* 106.1 (1992), p. 15.
- [14] Rishi Rajalingham, Kailyn Schmidt, and James J DiCarlo. “Comparison of object recognition behavior in human and monkey”. In: *Journal of Neuroscience* 35.35 (2015), pp. 12127–12136.
- [15] Rishi Rajalingham et al. “Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks”. In: *Journal of Neuroscience* 38.33 (2018), pp. 7255–7269.
- [16] Olivier Pascalis et al. “Development of face processing”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 2.6 (2011), pp. 666–675.
- [17] Winrich Freiwald, Bradley Duchaine, and Galit Yovel. “Face processing systems: from neurons to real-world social perception”. In: *Annual review of neuroscience* 39.1 (2016), pp. 325–346.
- [18] Lisa A Parr. “The evolution of face processing in primates”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1571 (2011), pp. 1764–1777.
- [19] Minh Nui Nguyen et al. “Neuronal responses to face-like stimuli in the monkey pulvinar”. In: *European Journal of Neuroscience* 37.1 (2013), pp. 35–51.
- [20] Yousif Khairuddin and Zhuofa Chen. “Facial emotion recognition: State of the art performance on FER2013”. In: *arXiv preprint arXiv:2105.03588* (2021).
- [21] Henrik Kessler et al. “Alexithymia and facial emotion recognition in patients with eating disorders”. In: *International Journal of Eating Disorders* 39.3 (2006), pp. 245–251.
- [22] Ana Idalina de Paiva-Silva et al. “How do we evaluate facial emotion recognition?” In: *Psychology & neuroscience* 9.2 (2016), p. 153.
- [23] Ralph Adolphs, Lonnie Sears, and Joseph Piven. “Abnormal processing of social information from faces in autism”. In: *Autism*. Routledge, 2013, pp. 126–134.
- [24] Michael K Yeung. “A systematic review and meta-analysis of facial emotion recognition in autism spectrum disorder: The specificity of deficits and the role of task characteristics”. In: *Neuroscience & Biobehavioral Reviews* 133 (2022), p. 104518.
- [25] Vicky Tsang. “Eye-tracking study on facial emotion recognition tasks in individuals with high-functioning autism spectrum disorders”. In: *Autism* 22.2 (2018), pp. 161–170.
- [26] Doris Y Tsao, Sebastian Moeller, and Winrich A Freiwald. “Comparing face patch systems in macaques and humans”. In: *Proceedings of the National Academy of Sciences* 105.49 (2008), pp. 19514–19519.
- [27] N Kanwisher, J McDermott, and MM Chun. “The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. 10.[Europe PMC free article][Abstract][Google Scholar]”. In: (1997).

- [28] Lisa A Parr and Matthew Heintz. “Facial expression recognition in rhesus monkeys, *Macaca mulatta*”. In: *Animal behaviour* 77.6 (2009), pp. 1507–1513.
- [29] Doris Y Tsao et al. “A cortical region consisting entirely of face-selective cells”. In: *Science* 311.5761 (2006), pp. 670–674.
- [30] David H Hubel, Torsten N Wiesel, et al. “Receptive fields of single neurones in the cat’s striate cortex”. In: *J physiol* 148.3 (1959), pp. 574–591.
- [31] David H Hubel and Torsten N Wiesel. “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: *The Journal of physiology* 160.1 (1962), p. 106.
- [32] Hubel Dh and WIESEL TN. “Shape and arrangement of columns in cat’s striate cortex.” In: *The Journal of Physiology* 165 (1963), pp. 559–568.
- [33] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. “How does the brain solve visual object recognition?” In: *Neuron* 73.3 (2012), pp. 415–434.
- [34] Daniel J Felleman and David C Van Essen. “Distributed hierarchical processing in the primate cerebral cortex.” In: *Cerebral cortex (New York, NY: 1991)* 1.1 (1991), pp. 1–47.
- [35] Michael H Herzog and Aaron M Clarke. “Why vision is not both hierarchical and feedforward”. In: *Frontiers in computational neuroscience* 8 (2014), p. 135.
- [36] Esther Peterhans and Rüdiger von der Heydt. “Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps”. In: *Journal of Neuroscience* 9.5 (1989), pp. 1749–1763.
- [37] Anitha Pasupathy and Charles E Connor. “Shape representation in area V4: position-specific tuning for boundary conformation”. In: *Journal of neurophysiology* (2001).
- [38] Brad C Motter. “Neural correlates of attentive selection for color or luminance in extrastriate area V4”. In: *Journal of Neuroscience* 14.4 (1994), pp. 2178–2189.
- [39] Roozbeh Kiani et al. “Object category structure in response patterns of neuronal population in monkey inferior temporal cortex”. In: *Journal of neurophysiology* 97.6 (2007), pp. 4296–4309.
- [40] Rufin Vogels and Guy A Orban. “Coding of stimulus invariances by inferior temporal neurons”. In: *Progress in brain research* 112 (1996), pp. 195–211.
- [41] Nuo Li et al. “What response properties do individual neurons need to underlie position and clutter “invariant” object recognition?” In: *Journal of neurophysiology* 102.1 (2009), pp. 360–376.
- [42] Chou P Hung et al. “Fast readout of object identity from macaque inferior temporal cortex”. In: *Science* 310.5749 (2005), pp. 863–866.
- [43] James V Haxby, Elizabeth A Hoffman, and M Ida Gobbini. “The distributed human neural system for face perception”. In: *Trends in cognitive sciences* 4.6 (2000), pp. 223–233.

- [44] Frank Tong et al. “Response properties of the human fusiform face area”. In: *Cognitive neuropsychology* 17.1-3 (2000), pp. 257–280.
- [45] Ralph Adolphs. “What does the amygdala contribute to social cognition?” In: *Annals of the New York Academy of Sciences* 1191.1 (2010), pp. 42–61.
- [46] Ralph Adolphs et al. “Recognition of facial emotion in nine individuals with bilateral amygdala damage”. In: *Neuropsychologia* 37.10 (1999), pp. 1111–1117.
- [47] John S Morris et al. “A differential neural response in the human amygdala to fearful and happy facial expressions”. In: *Nature* 383.6603 (1996), pp. 812–815.
- [48] Paul J Whalen et al. “Human amygdala responsivity to masked fearful eye whites”. In: *Science* 306.5704 (2004), pp. 2061–2061.
- [49] Peter Mende-Siedlecki et al. “Robust selectivity for faces in the human amygdala in the absence of expressions”. In: *Journal of cognitive neuroscience* 25.12 (2013), pp. 2086–2106.
- [50] Itzhak Fried, Katherine A MacDonald, and Charles L Wilson. “Single neuron activity in human hippocampus and amygdala during recognition of faces and objects”. In: *Neuron* 18.5 (1997), pp. 753–765.
- [51] Indre V Viskontas, Rodrigo Quiñan Quiroga, and Itzhak Fried. “Human medial temporal lobe neurons respond preferentially to personally relevant images”. In: *Proceedings of the National Academy of Sciences* 106.50 (2009), pp. 21329–21334.
- [52] Giulia Mattavelli et al. “Neural responses to facial expressions support the role of the amygdala in processing threat”. In: *Social cognitive and affective neuroscience* 9.11 (2014), pp. 1684–1689.
- [53] Shuo Wang et al. “Neurons in the human amygdala selective for perceived emotion”. In: *Proceedings of the National Academy of Sciences* 111.30 (2014), E3110–E3119.
- [54] Runnan Cao et al. “A neuronal social trait space for first impressions in the human amygdala and hippocampus”. In: *Molecular psychiatry* 27.8 (2022), pp. 3501–3509.
- [55] Richard Passingham. “How good is the macaque monkey model of the human brain?” In: *Current opinion in neurobiology* 19.1 (2009), pp. 6–11.
- [56] Nikolaus Kriegeskorte et al. “Matching categorical object representations in inferior temporal cortex of man and monkey”. In: *Neuron* 60.6 (2008), pp. 1126–1141.
- [57] Dante Mantini et al. “Interspecies activity correlations reveal functional correspondence between monkey and human brain areas”. In: *Nature methods* 9.3 (2012), pp. 277–282.
- [58] Guy A Orban, David Van Essen, and Wim Vanduffel. “Comparative mapping of higher visual areas in monkeys and humans”. In: *Trends in cognitive sciences* 8.7 (2004), pp. 315–324.
- [59] Doris Y Tsao et al. “Faces and objects in macaque cerebral cortex”. In: *Nature neuroscience* 6.9 (2003), pp. 989–995.

- [60] Livingstone MS, Freiwald WA, Tsao DY. “A face feature space in the macaque temporal lobe.” In: *Nat Neurosci* (2009). DOI: 10.1038/nn.2363.
- [61] Lisa A. Parr and Matthew Heintz. “Facial expression recognition in rhesus monkeys, *Macaca mulatta*”. In: *Animal Behaviour* 77.6 (2009), pp. 1507–1513. ISSN: 0003-3472. DOI: <https://doi.org/10.1016/j.anbehav.2009.02.024>. URL: <https://www.sciencedirect.com/science/article/pii/S0003347209001274>.
- [62] Daniel P Kennedy and Ralph Adolphs. “Perception of emotions from facial expressions in high-functioning adults with autism”. In: *Neuropsychologia* 50.14 (2012), pp. 3313–3319.
- [63] Shuo Wang and Ralph Adolphs. “Reduced specificity in emotion judgment in people with autism spectrum disorder”. In: *Neuropsychologia* 99 (2017), pp. 286–295.
- [64] Madeline B Harms, Alex Martin, and Gregory L Wallace. “Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies”. In: *Neuropsychology review* 20 (2010), pp. 290–322.
- [65] Hugo D Critchley. “Neural mechanisms of autonomic, affective, and cognitive integration”. In: *Journal of comparative neurology* 493.1 (2005), pp. 154–166.
- [66] Robert T Schultz et al. “Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome”. In: *Archives of general Psychiatry* 57.4 (2000), pp. 331–340.
- [67] Kevin A Pelphrey et al. “Perception of dynamic changes in facial affect and identity in autism”. In: *Social cognitive and affective neuroscience* 2.2 (2007), pp. 140–149.
- [68] Kohitij Kar. “A computational probe into the behavioral and neural markers of atypical facial emotion processing in autism”. In: *Journal of Neuroscience* 42.25 (2022), pp. 5115–5126.
- [69] MG Beaupré, N Cheung, and U Hess. “The Montreal set of facial displays of emotion”. In: *Montreal, Quebec, Canada* (2000).
- [70] Verena Willenbockel et al. “Controlling low-level image properties: the SHINE toolbox”. In: *Behavior research methods* 42 (2010), pp. 671–684.
- [71] Christopher Giverin, Hamid Ramezanpour, and Kohitij Kar. “Low-cost, portable, easy-to-use kiosks to facilitate home-cage testing of non-human primates during vision-based behavioral tasks”. In: (Jan. 2023). DOI: 10.31219/osf.io/csdzv.
- [72] Pascal Fries and Eric Maris. “What to Do If N Is Two?” In: *Journal of Cognitive Neuroscience* 34.7 (June 2022), pp. 1114–1118. ISSN: 0898-929X. DOI: 10.1162/jocn_a_01857. eprint: https://direct.mit.edu/jocn/article-pdf/34/7/1114/2028478/jocn_a_01857.pdf. URL: https://doi.org/10.1162/jocn%5C_a%5C_01857.

- [73] Kohitij Kar and James J DiCarlo. “Fast recurrent processing via ventral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition”. In: *BioRxiv* (2020), pp. 2020–05.
- [74] Rishi Rajalingham et al. “Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks”. In: *Journal of Neuroscience* 38.33 (2018), pp. 7255–7269.
- [75] Najib J Majaj et al. “Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance”. In: *Journal of Neuroscience* 35.39 (2015), pp. 13402–13418.
- [76] Kenneth O Johnson, Steven S Hsiao, and Takashi Yoshioka. “Neural coding and the basic law of psychophysics”. In: *The Neuroscientist* 8.2 (2002), pp. 111–121.
- [77] Kushin Mukherjee et al. “Leveraging Artificial Neural Networks to Enhance Diagnostic Efficiency in Autism Spectrum Disorder: A Study on Facial Emotion Recognition”. In: ().