THE UNIVERSITY OF CHICAGO

NEURAL ACTIVITY AND DYNAMICS UNDERLYING VISUAL CATEGORIZATION AND LEARNING

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON COMPUTATIONAL NEUROSCIENCE

BY
KRITHIKA MOHAN

CHICAGO, ILLINOIS
DECEMBER 2019

To Bootsy, my unwitting accomplice

And Chin, my beloved inspiration

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I am incredibly lucky and grateful to have had the opportunity to study Neuroscience at the University of Chicago. Central to my rewarding experience at UChicago is my advisor, David Freedman. Dave's enthusiasm for creative scientific work, together with his unmatched patience and generosity makes him an exceptional mentor. Under his guidance, I have truly grown as a student of science and perhaps more importantly, as a human being.

I would also like to thank Dave for cultivating a collaborative and supportive research environment through which I have met some of the most brilliant and outstanding people I have ever worked with. I thank Guilhem Ibos for teaching me everything about electrophysiology and Nicolas Masse for being my analysis guru. I could not have asked for better comrades-in-arms than Jeff Johnston and Barbara Peysakhovich, whose humor and company I will dearly miss. I was incredibly lucky to collaborate with Jeff, he is a phenomenal scientist and has contributed immensely towards my appreciation for theory and computation in neuroscience. Barbara and I are kindred spirits and I thank her for being an enthusiastic partner in our attempts to learn things we knew nothing about. I would also like to thank Stephanie Thomas, who besides teaching me everything monkey, did a lot towards making my initial years in the lab fun. I also thank all the FreedmanLab members - both past and present - for the laughter during the lulls and the camaraderie during the catastrophes.

I am eternally grateful to my committee members - John Maunsell, Daniel Margoliash, and Nicolas Brunel - for their constructive criticism and challenging me to think critically at every step in the way. Nicolas, especially, has instilled in me an appreciation for abstract theory, complementing that with concrete ideas and experiments. Through the collaboration with Nicolas's group, I am also thankful to have interacted with Ulises Pereira who was an absolute pleasure to work with. I am extremely grateful to Yali Amit in Stastistics for his exceptional patience in teaching me the nuts and bolts of statistical modeling.

I thank SP Arun, my Master's thesis advisor, at the Indian Institute of Science, Bangalore for nurturing my interest in research and getting me hooked onto primate vision. I am also thankful to my undergraduate alma mater IISER, Pune and my advisor, Aurnab Ghose without whose guidance I would not be pursuing neuroscience at all!

I would like to thank my friends who have played a pivotal role in my happiness throughout graduate school. I would like to thank The Herd for being fantastic virtual companions through my PhD. They have encouraged, sustained, inspired and tolerated me, and are the greatest group of friends I could have ever asked for. Life away from India, away from home was made so much easier by my friends in Chicago. I am extremely thankful to have had many hours of conversations over many cups of *chai* with Vishwas, Kushal, Swati, Poojya and Piyush. You guys are the closest I have to family in a foreign land - thank you for always having my back.

I am deeply grateful to my parents for everything in my life so far. My father has always been a tireless champion of education. Any iota of discipline, work ethic, and integrity that I have is a direct result of his upbringing and for all of those, I thank him. My mother is an exceptional human being, and by example, she continues to inspire me to be patient, kind, and caring without which all the science in the world is pointless. She is a constant source of comfort and has single-handedly contributed to my sanity by ensuring I eat, sleep and relax on time, all from 8000 odd miles away. Thank you, amma! My sister Bhargavi, a doctor, has been a tremendous role model and my love for science and math can be easily traced back to our study sessions as kids. She has also perhaps taught me the most important thing that I ever need to learn - to not take myself too seriously - something she and my family regularly demonstrate by making endless fun of my existence. The objective function of all actions in my life can easily be approximated to ”do anything that impresses my sister”, and my PhD is a small step in that direction. Finally, I thank my little niece Chinmayi, who occupies a disproportionately large amount of my cortical space compared to the physical space she occupies in this world. During the time it has taken me to finish my fairly insignificant PhD, she has done many, many significant things. She has been born, she has mastered vision, motor control, speech and her explorations of the world constantly makes me gasp in wonder,

”How does the brain do it all?” - a question I aspire to think about all my life.

ABSTRACT

Humans have a remarkable ability to sense visual stimuli, adapt decisions and respond to rapid changes in the environment. This ability is thought to rely on learning-dependent plasticity in neural circuits. While some visual behaviors are learned in a supervised manner through explicit experience, some others are innate and develop in an unsupervised manner. In this thesis, we investigated neuronal representations of behavior formed through both supervised (visual categorization) and unsupervised learning (novel/familiar recognition). The central goal is to understand the neural activity and dynamics underlying visual categorization and learning in the dorsal and ventral visual streams of the cortical hierarchy. Our experimental approach involved electrophysiological recordings from populations of neurons in awake, behaving monkeys that were trained to perform complex, cognitive tasks.

First, we investigated how flexible, categorical decisions are mediated by the dorsal stream (specifically, the lateral intraparietal cortex (LIP) in PPC). Prior studies have than shown that LIP encodes abstract learned visual categories of motion directions learned over months of supervised training. Visual motion categorization has only been studied in the context of the delayed match-to-category task with specific, fixed task-demands such as delayed decisions and match/non-match computations. So, we do not know whether LIP's role in categorization with stable, category rules generalizes to more flexible tasks that require animals to rapidly adapt their decisions to changes in context. To explore whether LIP plays a generalized role in category-based decision making, we developed a task-switching paradigm in which monkeys categorized the same random-dot motion stimuli in two different tasks. The tasks differed in the specific behavioral demands involved in making decisions, the effector used for reporting the decision (eye vs. hand movement) and in the sequence of task events upon which decisions were based. We found that neural populations showed remarkably similar categorical encoding in both tasks, with greater category selectivity in the task with memory and matching functions. These findings demonstrate that parietal neural circuits contribute to flexible task switching by dynamically changing information

coding in a task-dependent manner.

Second, we examined how unsupervised learning is mediated in the ventral visual pathway (area IT) as monkeys learned to recognize and become familiar with novel visual stimuli. Humans and other primates have an exceptional ability to learn to recognize familiar visual items, and to discriminate them from novel items. This ability is thought to rely on experience dependent changes in neuronal representations in the ventral visual pathway, particularly IT cortex. However, how neuronal familiarity encoding develops during learning and the link between neuronal familiarity effects and behavior is not well understood. To date, experimental studies have primarily sampled from the extremities of the familiarity spectrum, comparing highly familiar to highly novel images without capturing the dynamics in between. To probe this gap in knowledge, we familiarized rhesus macaques with multiple sets of novel images for up to 8 days each, while monitoring responses of IT neuronal ensembles during the familiarization process - both over long and short time scales. We found that familiarity encoding due to learning in IT cortex develops slowly over the course of several days, and familiarity-related changes in neuronal representations are minimal over short time scales of a single recording session. Our results give insight into how experience affects visual representations in IT across multiple timescales.

Together, our results provide insights into how the primate brain learns to represent flexible behavior in the dorsal and ventral visual hierarchy. In the dorsal visual stream, our results demonstrate the task-dependence and flexibility of visual, category, and choice encoding in LIP, and reveal distinct roles for LIP in the two categorization tasks in the task switching paradigm. Through detailed comparisons between fundamentally similar categorization tasks with tractable differences, our contributions provide compelling evidence that dynamical formatting of relevant information provides a mechanistic basis for LIP to govern flexible, categorical decisions in the dorsal visual stream. In the ventral visual stream, our results demonstrate that familiarity-dependent plasticity in neural activity develops slowly over the course of several days, with a modest impact of familiarity on representations within

a single recording session. Our results suggest that experience dependent plasticity occurs at multiple timescales in IT object representations.

CHAPTER 1

INTRODUCTION

The ultimate goal of the nervous system is to sense, process, and act on stimuli in the environment. Over evolution, nervous systems have evolved from simple nerve nets that perform both sensory and motor functions to complex central nervous systems which possess distinct circuits for sensation and action. Further, specialized neural circuits have also emerged for the detailed processing of sensory inputs before routing to motor outputs. This sophistication in the neural circuitry for processing has been accompanied by a dramatic increase in the capacity for flexible behavior.

Humans and advanced animals excel in a wide range of complex, learned behaviors that go beyond simple, reflex actions. Almost all of these complex, behaviors are produced via learning. While some behaviors are innate or learned over development, many of them are flexibly learned through experience - by observation or instruction. As an example of an innate bias, humans and infants instinctively look at novel objects and explore them more than objects that have already been encountered before. On the other hand, recognizing object categories (such as ripe vs. unripe fruits) usually depends on experiencing multiple exemplars from each category and explicitly learning category labels. Understanding how visual categories are learned in supervised and unsupervised scenarios could provide key insights into the general mechanisms of learning in the brain.

In this thesis, we investigate how supervised and unsupervised learning influences neural encoding in cortex, and how learning-dependent changes in neuronal representations serve to store behaviorally relevant information in long term memory. In particular, this research focuses primarily on the parietal cortex in the dorsal stream and the temporal cortex in the ventral stream.

1.1 Two-streams hypothesis

An influential theory proposed over 30 years ago, the two-streams hypothesis, still serves to anchor our understanding of how the brain processes visual information (Figure 1.1). According to this proposal, our visual experience is constructed by the concerted activity of two parallel pathways - the dorsal "where" pathway and the ventral "what" pathway (Belkacem et al., 2014; Mishkin and Ungerleider, 1982; Ingle et al., 1967). The dorsal visual stream projects from the primary visual cortex through areas MT and MST to the posterior parietal cortex. The dorsal stream is involved in the encoding of space, motion, spatial attention, and visually guided movements such as eye movements and hand movements. Damage to the dorsal pathway disrupts the ability to make spatial judgments and precise, visually-guided movements with profound effects on the ability to attend to locations in space. The ventral visual stream extends from the primary visual cortex to the inferotemporal cortex through areas V2 and V4. The ventral stream is critical for object recognition and specializes in the detailed processing of objects, faces, texture, and color. When areas in this stream are damaged, object perception and recognition is disrupted. To capture these functional differences, the dorsal stream is referred to as "vision for action" and the ventral stream as "vision for perception" (Belkacem et al., 2014). Although the dorsal and ventral streams perform seemingly independent functions, both streams interact and integrate visual information to guide behavior (Sereno and Maunsell, 1998; Hong et al., 2016; DiCarlo and Maunsell, 2003).

The goals of this research are to elucidate the neuronal mechanisms of supervised and unsupervised learning. In order to accomplish this, we focus on two abstract behaviors that are functionally associated with the two visual streams. First, in the realm of supervised learning, we examine the flexibility and context-dependence of category representations that have been learned over months of explicit training. Second, we aim to understand the neuronal dynamics of unsupervised learning by examining how neuronal representations of familiarity develop over multiple time scales.

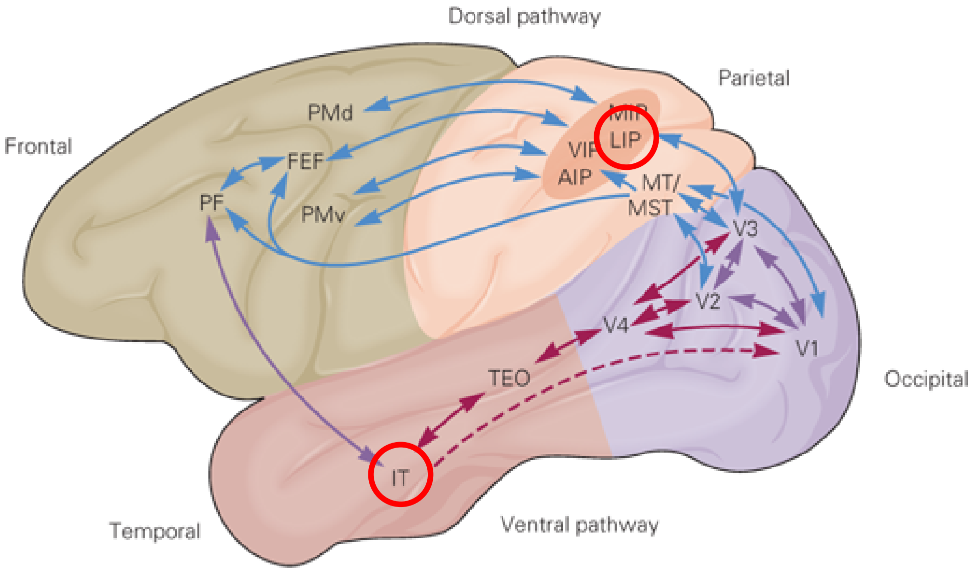

Figure 1.1: Two-streams hypothesis.
Lateral view of a macaque brain highlighting the dorsal and ventral visual pathways. Red circles represent the cortical areas relevant for the studies presented here. Adapted from Kandel ER, et al., Principles of Neural Science. 5th ed.

1.2 Supervised learning: visual categorization in the dorsal visual stream

Visual categorization is a cognitive behavior in which we group together objects that share common properties, regardless of their physical differences. The neuronal basis of categorization has been studied in many brain areas, including areas in the dorsal (LIP, 7a) visual stream, ventral visual stream (IT) and the prefrontal cortex (PFC).

1.2.1 Shape categorization

Early studies on categorization were performed with shape stimuli that were images of real-world, complex objects (like categories of cars or birds) or reduced versions of complex objects that were constructed by changing parametric visual features in a discrete manner. For example, Sigala and Logothetis (2002) trained monkeys to perform categorization with two sets of parametric shape stimuli. Each stimulus was defined by a set of 4 varying features (such as eye height, nose length) but only two of them for relevant for categorization and others were irrelevant. After training, single neurons in IT cortex showed an enhanced neuronal representation of relevant, but not irrelevant, category-specific features.

These early experiments on categorization were confounded by high physical similarity between members of the same category (for example, all faces are physically similar to each other) and high physical dissimilarity between members from different categories (faces and cars are not physically similar to each other). Hence, if stimuli from the same category produced similar neuronal responses, it remained unclear if this was because of visual similarity or conceptual similarity. This confound was overcome in a study that investigated category representations in PFC during a delayed match-to-category (DMC) task with ”cat” and ”dog” categories that were generated through a continuous morphing system. The authors controlled for physical similarity by ensuring that some visually dissimilar items belonged to the same category and some visually similar items belonged to different categories (Freedman et al., 2001). During the DMC task with cat/dog categories, the activity of PFC neurons reflected the category of the visual stimuli by responding with similar firing rates to stimuli within a category but not for stimuli in different categories. A subsequent study compared category representations in IT and PFC and found that while PFC represents behaviorally relevant categories in a near-binary manner, IT represents visual features rather than categorical features of stimuli (Freedman et al., 2003).

1.2.2 Motion categorization

Later studies explored the neural correlates of categories based on visual features other than shape such as motion (Freedman and Assad, 2006), spatial location (Masse et al.; Antzoulatos and Miller, 2011), and speed (Ferrera et al., 2009) - features encoded in the dorsal visual stream. Visual, motion-based categories were first examined in two brain areas in the dorsal stream - middle temporal (MT) area, an upstream sensory area that encodes motion direction and lateral intraparietal cortex (LIP), a downstream area involved in sensorimotor and higher cognitive functions (Freedman and Assad, 2006). Neurons from LIP and MT were recorded as monkeys performed a delayed match-to-category task with motion stimuli. Motion stimuli comprised random-dot kinetograms with dots moving in a range of 0-360°, divided into two categories based on an arbitrary category boundary such that directions on either side of the boundary belonged to different categories. In this task, monkeys indicated whether a test stimulus matched in category to a previously presented sample stimulus. The two motion stimuli, sample and test, were separated by a delay period during which the category of the sample stimulus had to be maintained in short term-memory. After learning, neurons in LIP reflected the category membership of stimuli, whereas neurons in MT reflected motion direction of stimuli (Figure 1.2). Notably, in this task, LIP neurons encoded reflected categorical activity was strongest in the delay period. Thus, LIP neurons reflect learned, abstract, motion-based, categorical decisions in a task with memory and matching computations.

Since the first report of motion category selectivity in LIP, a host of studies have compared the neuronal activity and dynamics of category representations during the DMC task in multiple regions in the dorsal and frontal visual stream. In addition to LIP, category signals have been observed in PFC (Swaminathan and Freedman, 2012) and MIP (Swaminathan et al., 2013) with category encoding appearing with a shorter latency in LIP over MIP and PFC. Whereas MIP strongly reflected motor aspects of the match/non-match decision, PFC reflected the abstract match/non-match status of test stimuli, separate from the motor movement associated with match trials. A recent study found that reversible inactivation of LIP impairs performance in motion discrimination and categorization tasks, suggesting that LIP is causally involved in categorical decision-making (Zhou and Freedman, 2019). The results from inactivation suggested that LIP's involvement in perceptual decisions might be more pronounced in tasks with flexible rather than fixed sensorimotor associations.

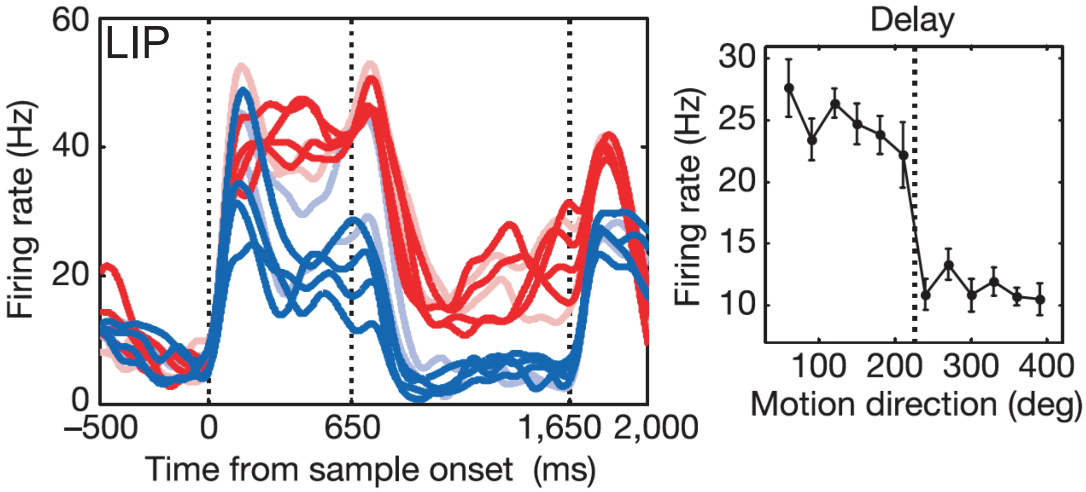

Figure 1.2: Example of a category-selective neuron.
Average activity to 12 sample directions for one LIP neuron. The red and blue traces correpond to directions in the two categories (red, category 1; blue, category 2), and pale traces indicate the directions closest to the boundary. The three vertical dotted lines indicate sample onset at 0 ms, sample offset at 650 ms, and test onset at 1650 ms. The plot on the right shows activity (mean ± s.e.m) for the 12 directions during the delay period. Adapted from Freedman and Assad (2006).

1.2.3 Task-dependence of category representations

Although reversible inactivation and comparisons between different brain areas reveal LIP as a candidate brain region in category-based decisions, whether LIP circuits show generalized representations for categories beyond the DMC task remains unknown. As described above, in the DMC task, monkeys decide whether two motion stimuli, separated by a delay period, belong to the same category or not. The DMC task is associated with behavioral demands such as short-term memory and stimulus comparison. Does LIP's role in categorical decisions perhaps reflect its role in these cognitive functions like working memory or matching? For instance, in the context of working memory, neurons in a number of cortical areas in the dorsal stream including LIP, display spatially selective, persistent activity during the delay period of a memory-guided saccade task (Chafee and Goldman-Rakic, 1998) (Gnadt and Andersen, 1988). In addition to space, LIP neurons have also been shown to encode non-spatial forms of short-term memory, such as shape in a delayed match-to-sample task (Sereno and Maunsell, 1998). Further, a recent study compared memory representations for motion stimuli in a delayed match-to-sample (DMS) task and a delayed match-to-category (DMC) task (Sarma et al., 2015). The authors found that whereas LIP populations did not encode motion direction during the delay period in the DMS task, they showed robust encoding of abstract motion categories during the delay period in the DMC task. Overall, these results suggest that learning results in the appearance of persistent activity in the delay period. Additionally, LIP neurons show strongest category selectivity in the delay period rather than the sample period of the DMC task. In the example neuron shown in Figure 1.2, the tuning curve on the right shows binary-like encoding of categories with similar response magnitudes to all directions within a category. These observations raise the question of whether binary category representations arise because of the need to maintain task-relevant information in short-term memory. By examining how neuronal representations depend on task context and cognitive demands, we can gain insight into the functional role of LIP in flexible, categorical decision-making.

1.3 Unsupervised learning: Familiarity learning in the ventral visual stream

Both humans and monkeys preferentially attend to and look at novel compared to familiar images. Additionally, our ability to remember images that we have encountered is impressive - we correctly determine whether we have seen an image before even after viewing thousands of images. The neural processes that support this kind of unsupervised visual familiarity memory are thought to be implemented in the same brain areas that support visual object recognition - the temporal cortex.

1.3.1 Visual object recognition

The ventral visual stream originates in V1 which projects to extrastriate visual areas V2 and V4 and finally terminates in the inferotemporal cortex (IT) (HUBEL and WIESEL, 1965). Starting from cells with small receptive fields selective to simple visual features like orientation in V1, neurons along the ventral stream show an increase in receptive field size and selectivity to increasingly complex features (Brincat and Connor, 2004). For example, whereas simple cells in V1 respond preferentially to oriented bars, cells in V4 encode visual curvature, which could be envisioned as an integral over many oriented bars (Hubel and Wiesel, 1962; Pasupathy and Connor, 2002). At the top of this hierarchy, in the inferotemporal cortex, neurons respond with exquisite selectivity to objects, faces, textures, and color (Gross et al., 1972). Combined with this specialized visual selectivity, IT neurons also exhibit invariant responses to changes in low-level features of visual stimuli such as scale, position, and sometimes, view (Logothetis et al., 1995). In addition to encoding object identity, the tuning of IT neurons also depends on visual experience (Li et al., 1993; Fahy et al., 1993).

1.3.2 Supervised learning in the inferotemporal cortex

Most learning studies in IT have focused on supervised learning in which monkeys perform a recognition, discrimination, or search task. For example, (Sakai and Miyashita, 1991) found that single neurons in IT cortex acquire selectivity for visual patterns through associative learning. In a delayed pair-association task, IT neurons showed similar responses to shapes associated as a pair. They also found that when single neurons strongly responded to an image during the sample period, they also exhibited increasing delay period activity when the paired image was presented as a sample, suggesting that the delay activity enabled image recall. In addition to associative learning experiments, studies also examined the impact of perceptual learning in IT. Sigala and Logothetis (2002) studied how neuronal selectivity changes in IT after training on a face categorization task. The results showed that selectivity for features relevant for the categorization task, what they call diagnostic features, was significantly larger than those irrelevant for the task, indicating an enhancement in selectivity as a result of training. The tuning of neurons in anterior IT can also be modulated by visual experience, as neurons acquire specificity for certain views or luminance dependent on training. For example, Logothetis et al. (1995) et al. trained monkeys to perform an object recognition task with isolated views of novel three-dimensional objects ('paperclips'). When recording from IT, they found that the great majority of neurons that showed selectivity for the trained objects were tuned to the specific view of the object that they had experienced. Although monkeys were performing a recognition task with feedback on every trial during training, it is tricky to determine if the view tuning arose as a result of trial-by-trial feedback or as a result of multiple, passive exposures to the same object in a particular view. One could argue that the process by which neurons get tuned more and more specifically to a particular value of a given feature (for e.g. view) is unsupervised and does not require explicit feedback. However, the distinction between supervised and unsupervised learning can be tricky.

1.3.3 Unsupervised learning in the inferotemporal cortex

Unsupervised learning has been relatively unexplored in IT and in other brain areas, in general. One aspect of object recognition - recognizing objects effortlessly despite changes in object position, scale and pose or tolerance - is an example of unsupervised learning (Tanaka, 1996). We rarely look at every object in every possible configuration, yet we correctly recognize them. A previous study examined how position tolerance arises in IT representations by altering the natural temporal structure of natural experience. Specifically, they hypothesized that retinal images that appear close together in time are associated with each other (Li and DiCarlo, 2008, 2010). When natural temporal structure of visual experience was disturbed in a systematic way, position tolerance in single IT neurons also changed rapidly after an hour of exposure in the altered visual world.

Behavioral studies in both humans and non-human primates show that visual stimuli that are novel or unexpected preferentially attract attention and eye movements compared to previously viewed ”familiar” objects. This is shown during preferential looking tasks in which subjects spend more time viewing novel compared to familiar stimuli (Jutras and Buffalo, 2010). The ability to recognize previously viewed stimuli as familiar is believed to involve IT. Neurons in IT are strongly influenced by experience and learning. IT neurons show a strong experience-dependent encoding of whether stimuli are novel (never seen before) or familiar (viewed many times in the past), with a pronounced decrease in average IT activity, and a sharpening of selectivity for familiar images.

1.3.4 Impact of neuronal familiarity on behavior

What is the relationship between neuronal encoding of familiarity and behavior? This question has mostly been unexplored until recently, since most studies in IT cortex have employed passive viewing in which monkeys view images presented at the fovea without performing a behavioral task. On comparing IT responses between a passive viewing task and a delayed match-to-category task, a recent study found similar IT responses to the same visual stimuli, suggesting that IT representations primarily provide highly processed visual information in a task-independent manner (McKee et al., 2014; DiCarlo and Maunsell, 2000). To study how novel images become familiar across single image exposures, a recent study trained monkeys to detect exactly one repetition of an image and report whether an image was novel (never seen before) or familiar (seen exact once) with an upward or downward eye movement, respectively (Meyer and Rust, 2018). The authors found that familiarity suppression explained behavior in the visual memory task, since suppression was more pronounced when the image was detected as a repeat than when it was not. Thus, it has been hypothesized that the reduction of the population response could enable better detection of familiar images. It has also been posited that the reduction of the population response could facilitate better discrimination of familiar objects. Support for this idea comes from the observation that familiarity suppression is more evident for non-preferred images and this results in sharper neuronal tuning and sparser and more selective population representations for familiar images than novel images (?Woloszyn and Sheinberg, 2012). However, behavioral support for this idea is mixed since improved processing has been observed only during explicit training in some tasks and not during passive viewing (Rainer and Miller, 2000; Rainer et al., 2004). Finally, reduced responses for familiar images might reflect reduced salience of familiar images as compared with novel images (Jutras and Buffalo, 2010; Ghazizadeh et al., 2016). Monkeys, like humans and infants, preferentially foveate novel images more than familiar images, but no study has yet linked neural familiarity suppression to image salience or gaze patterns to novel/familiar images.

1.3.5 Neural dynamics of familiarity learning

Studies of familiarity in IT have compared neuronal responses to stimuli in two discrete categories - entirely novel images and highly familiar images. This provides only an asymptotic view of learning with no insight into the dynamics and time course of how initially novel objects become familiar. The goal of this research is to investigate the dynamics of familiarity learning over multiple time scales by sampling continuously across different levels of familiarity. To investigate the effects of repeated presentations of initially novel stimuli, we examined changes in neuronal encoding as initially novel stimuli became increasingly familiar over the course of a recording session, and across multiple daily recording sessions. These results will provide crucial insights into the neural mechanisms responsible for visual familiarity learning and memory.

CHAPTER 2

INTERACTION BETWEEN NEURONAL ENCODING AND DYNAMICS DURING CATEGORIZATION TASK SWITCHING IN PARIETAL CORTEX

2.1 Abstract

Primates have a remarkable ability to make flexible categorical decisions. Visual categories are encoded by neural populations in multiple cortical areas, yet the format of the underlying neural code (e.g. discrete vs. continuous) differs between areas. Furthermore, it remains unclear how neuronal encoding generalizes between categorization tasks with different cognitive demands. We trained monkeys to flexibly categorize visual motion stimuli in two tasks which varied in their effectors and mnemonic demands. In a one-interval categorization (OIC) task, monkeys rapidly reported their categorical decisions by a saccade, while a delayed match-to-category task required monkeys to compare sequentially presented stimuli and report their match/non-match decisions with a manual response. Neural populations in parietal cortex exhibited similar patterns of category selectivity, but showed more binary and persistent categorical encoding in the DMC task and more graded direction encoding in the OIC task. We hypothesize that more binary category encoding could arise by compression of stimulus representations due to attractor dynamics underlying delay activity. Together, our findings suggest that parietal circuits contribute to flexible categorization task switching, and that the format of categorical encoding is markedly influenced by attractor dynamics underlying persistent activity during working memory-based computations.

2.2 Introduction

Primates flexibly process incoming sensory stimuli by assigning them to groups, a powerful cognitive process called categorization. We learn to group stimuli into behaviorally relevant categories such as 'fruit' or 'furniture' and use these categorical decisions to guide appropriate motor responses. During categorization, the brain transforms continuous sensory information into discrete classes, representing categories in an abstract, intermediate stage between sensory input and motor output. Indeed, neural responses correlated with abstract, visual categories have been observed in multiple brain regions of the macaque visual system, including the prefrontal (Freedman et al., 2001; Swaminathan and Freedman, 2012) cortex, parietal cortex (Freedman and Assad, 2006) and inferotemporal cortex (Sigala and Logothetis, 2002; Freedman et al., 2003; Meyers et al., 2008). Yet, the detailed dynamics of these category representations differ between brain regions, with neural populations encoding categories in a discrete or continuous format.

Previous work has shown that single neurons in the prefrontal (Freedman et al., 2001; Swaminathan and Freedman, 2012) and parietal cortices (Swaminathan and Freedman, 2012; Freedman and Assad, 2006) encode abstract categories in a discrete format with distinct responses to stimuli in different categories but similar responses to all stimuli within the same category. In this encoding scheme, neurons produce nearly identical firing patterns for all stimuli within a category and thus, firing rate variability within a category is thought to collapse to a stable neural state. Single neurons in the inferotemporal cortex (Freedman et al., 2001, 2003) appear to follow an alternative category encoding scheme representing categories in a continuous format with distinct responses between categories but with graded responses to stimuli within a category. Whereas discrete formats of categorical encoding yield binary-like tuning curves with sharp transitions across the category boundary, continuous formats yield gaussian-like tuning curves that taper off near the boundary. These disparate forms of category encoding directly control the extent of sensory (feature) vs. decision (category) information in neural populations, yet the mechanisms that produce these distinctions between formats of category representations remain unknown. Do different formats of category encoding correspond to different cognitive processes or computations?

The neural correlates of abstract categorical decisions have been examined in behavioral tasks in which categorical decisions were dissociated from sensory stimuli and motor responses. Most often, this has been accomplished by introducing a memory delay period between sensory and motor aspects of the task. For example, the neural correlates of motion categorization were investigated using delayed match-to-category (DMC) tasks in which monkeys stored category information in working memory and compared sequentially shown sample and test stimuli to identify category matches (Freedman and Assad, 2006). After learning, single neurons in multiple cortical areas, including LIP and PFC, show binary-like encoding of learned motion categories, and LIP activity in particular correlates with perceptual and categorical choices on a trial-by-trial basis (Swaminathan and Freedman, 2012). Indeed, recent work has demonstrated a causal role for LIP in categorization during both perceptual and categorical decisions about visual motion stimuli (Zhou and Freedman, 2019). Yet, binary category signals have mostly been observed in categorization tasks that have employed delayed matching tasks as a way to study abstract decisions. Whereas including a delay period enables the careful study of abstract decisions, it however imposes additional cognitive demands, such as maintaining and manipulating information, which are orthogonal to the specific categorization aspect of such tasks. This raises the possibility that the binary-like categorical signals that have been observed are perhaps a consequence of the specific set of task-demands that accompany categorization tasks. However, little is known about how different cognitive demands affect the underlying neural codes and mechanisms adopted by the brain.

Here we examine whether LIP plays a generalized role in categorical decisions, or whether its role is more related to task-specific functions such as memory, motor response selection, or comparison. To identify whether different neural mechanisms are engaged by different cognitive functions, we devised a cued task switching paradigm in which monkeys alternated between motion categorization tasks that varied in the behavioral demands required to make and report categorical decisions. The task structure required flexibility in several domains including: - (i) the timing of monkeys' decisions (e.g. rapid vs. delayed), (ii) whether the task involved memory and comparison between multiple stimuli, and (iii) the motor response modality (e.g. eye vs. hand movement) used to report decisions. Monkeys alternated in blocks between two motion categorization tasks, both of which required grouping 360° of motion directions into two categories according to an arbitrary, learned category boundary. In the first task, a one-interval categorization (OIC) task, monkeys rapidly reported the category of a stimulus with an eye movement to a red/green target to indicate category 1/category 2. In the second task, the DMC task, monkeys compared two sequentially presented stimuli and reported whether they matched in category by releasing a touch bar. The key differences between tasks are that, in the OIC task, monkeys made rapid categorical decisions without relying on memory or matching functions, while in the DMC task, monkeys made delayed decisions which involved both functions.

In both tasks, neurons in LIP exhibited diverse responses and multiplexed encoding of behaviorally relevant variables motion direction, category, and motor responses. Despite heterogeneity in single-neuron responses, LIP activity contained a representation of stimulus categories which generalized across both tasks. While shared circuits represented motion categories, category representations exhibited distinct encoding schemes in each task. First, category selectivity was stronger in the DMC task as neural activity was nearly binary with little separation between neural responses to directions within a category, but displayed more heterogeneous responses among directions within a category in the OIC task. Second, categorical representations were persistent and stable in the DMC task, but more transient and dynamic in the OIC task. We hypothesize that binary categorical encoding observed in the DMC task is a consequence of attractor dynamics required to generate persistent activity in working memory. Indeed, recurrent neural networks trained to perform categorization task switching revealed attractor-like dynamics in the DMC, but not in the OIC task. Thus, the binary category-correlated activity observed in DMC tasks may not be a general mechanism underlying the categorization process, but may instead arise, either in part or in full, from the network dynamics underlying working memory. More broadly, our approach of incorporating flexibility in behavioral tasks facilitates the understanding of how activity patterns in the brain support generalized cognitive computations, and also distinguishes between distinct patterns of neural encoding which support stimulus categorization and working memory.

2.3 Results

2.3.1 Categorization task switching

Monkeys alternated between the OIC and DMC tasks in which they indicated the category of motion stimuli with a saccade (OIC) or their category match/non-match decisions (DMC) with a hand movement (Figure 2.1A). In the OIC task, monkeys' categorical decisions were associated with saccades to either a red or a green target for category 1 or category 2 respectively. Importantly, the target position varied randomly between two possible locations, one in each hemifield, ensuring a dissociation between the decision and the choice report. In the DMC task, monkeys indicated whether a test stimulus matched in category to the previously presented sample by releasing a touch bar. On trials in which the first test stimulus was not a categorical match to the sample, the monkey held the touch bar until a second test stimulus appeared (which was always a categorical match and required a touch bar release). In both tasks, sample stimuli were presented within LIP neurons' receptive fields (RF) assessed by RF mapping earlier in that session (see Methods). In the OIC task, one saccade target was always presented inside the RF and the other target was placed diametrically opposite to the RF. In the DMC task, test stimuli were always presented at the same location as the sample stimulus (i.e. in the RF). Based on a colored cue presented at the start of the trial, monkeys switched between the OIC and DMC tasks in blocks of 20 or 30 correct trials in Monkey M or B respectively. Monkeys successfully switched between the two tasks with mean behavioral performance of 87.6% correct in OIC (Monkey M: 88.9%; Monkey B: 86.4%) and 84.3% correct in DMC (Monkey M: 85.7%; Monkey B: 83.1%) (Figure 2.1C). Monkeys' performance was significantly higher in the OIC task, especially for the most difficult motion directions that were 10° away from the boundary (p-value of unpaired t-test: 10e-08).

2.3.2 LIP neurons are preferentially engaged in the OIC task

As monkeys alternated between the OIC and DMC tasks, we recorded from 100 LIP neurons from two monkeys (monkey M, N=56 single units; monkey B, N=44 single units). We compared LIP responses in the OIC and DMC tasks to determine whether LIP neurons were preferentially engaged in one task over the other. We hypothesized that LIP's well-known role in spatial attention and saccadic eye movements would be reflected in the firing rates in the OIC task which requires a saccadic report. Alternatively, LIPs role in spatial functions might be distinct from its role in visual motion categorization, as suggested by a previous study in which saccade signals and category signals were independently encoded when they were not behaviorally linked (Rishel et al., 2013). We focused on the shared sample period which was identical in both tasks during which monkeys were fixating and stimulus presentation was exactly alike. During the sample period, the mean firing rate of 73.2% (60/82) (unpaired two-sample t-test, p¡0.01) of individual LIP neurons was significantly greater in the OIC than DMC task (Figure 2.2D). This was evident even for some individual neurons, in which block-wise firing rates of single neurons were higher in the OIC than DMC task (Figure 2.2C). We fit a linear model to predict OIC firing rates from DMC firing rates and the best-fitting model had a slope of 1.13 (linear model, p=10e-42) and an intercept of 2.72 (linear model, p=0.10). This indicates that increased OIC firing rates could be predicted by a positive multiplicative gain applied to DMC firing rates.

Higher average firing rates in the OIC than DMC task could be a result of many factors that vary between the two tasks, such as differences in: motor response modality, task difficulty or task demands such as memory or comparison. Because of LIP's known involvement in saccadic eye movements (Snyder et al., 1997; Snyder et al.), we hypothesized that LIP responses could be modulated by prior information about the motor modality used to report decisions. Even before trial onset, the context cue indicated which effector (eye or hand) is

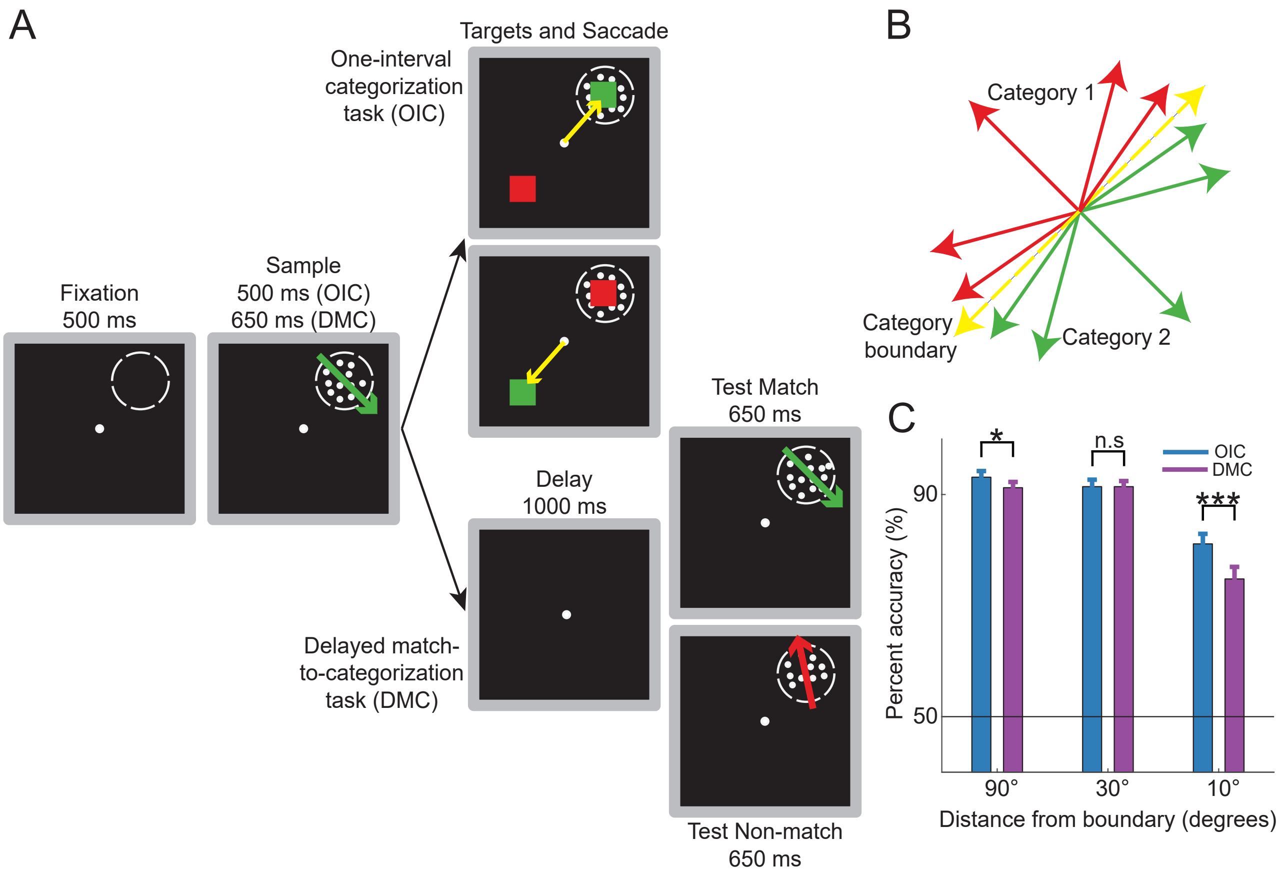

Figure 2.1: Categorization task switching.

(A) Monkeys were trained to alternate between a one-interval categorization (OIC) task (top panel) and a delayed match-to-category (DMC) task (bottom panel). Each trial started with a 500 ms fixation period, followed by a sample period (500 ms in OIC, 650 ms in DMC) during which motion stimuli appeared in the RF of the recorded neuron. In OIC, colored saccade targets (red and green) previously associated with the two motion categories appeared at 500 ms from sample onset. The locations of the targets were randomly assigned to one of two possible locations inside the RF and diametrically opposed to the RF. At target appearance, monkeys made a rapid eye movement to the red/green target to indicate category 1/category 2. In DMC, the sample stimulus was followed by a 1 s delay, which was followed by a test stimulus. If the sample and test stimulus matched in category, monkeys released a lever, else they held the lever till a second test stimulus appeared, which was always a categorical match. Both the sample and test stimuli were placed in the RF of the recorded neuron. (B) The stimulus space consisted of motion stimuli moving in one of 10 possible motion directions, divided into 2 categories (red and green colors) by an arbitrary category boundary (yellow dashed line). (C) Monkeys performed categorization task switching successfully at 87.63% correct in OIC, 84.38% correct in DMC. Performance declined at directions close to the boundary in both tasks. OIC was easier than DMC for the easiest (90° away from bound) and hardest (10° away from bound) motion directions. Error bars indicate s.e.m. * $p<0.05$, *** $p<0.001$, n.s not significant, unpaired two-sample t-test.

relevant during that block. If LIP neurons were modulated by prior knowledge about the motor modality, we would expect fixation period firing rates to be higher in the OIC than DMC task. Indeed, the mean firing rate of 68.4% (39/57) of neurons was greater in the OIC task during the fixation period. Further, a population-level linear classifier decoded task context (OIC or DMC) at 91.93% accuracy using population firing rates during the fixation period. Thus, LIP neurons are preferentially engaged with higher firing rates in the OIC task when the task requires a saccadic vs. manual report.

2.3.3 LIP neurons simultaneously encode multiple task-relevant variables

We next asked how encoding of task-relevant variables differed between tasks. We quantified and compared the fractions of cells that encoded shared task variables - motion direction and category. In the OIC task, 80% (80/100) of recorded neurons were selective for motion direction (one-way ANOVA with 10 motion directions, p¡0.01) during the sample period and 69% (69/100) were selective during the saccade epoch. Similarly, during the DMC task, 80% of recorded neurons were motion direction selective in the sample period and 53% (53/100) were selective during the delay period. In addition to encoding motion direction, LIP neurons also encoded motion category during various task epochs. While 57% (57/100) of cells were category selective (one-way ANOVA with 2 motion categories, p¡0.01) during the OIC sample period, a significantly greater fraction (72/100) of cells were category selective during the DMC sample period (chi-square test, χ^2=4.91, p=0.02). In both tasks, many single neurons differentiated between categories by responding with similar firing rates and dynamics to directions within a category and showed distinct activity for directions in different categories (example neurons in Figure 2.2A and 2.2B). Finally, LIP neurons also encoded choice e.g. saccade to/away from the RF in OIC and match/non-match in DMC. 64% (64/100) of neurons encoded saccadic direction in OIC and 61% (61/100) of neurons encoded match/non-match in DMC. Thus, LIP activity multiplexes a variety of behaviorally relevant sensory (motion direction), decision (motion category), and motor (saccade direction and match/non-

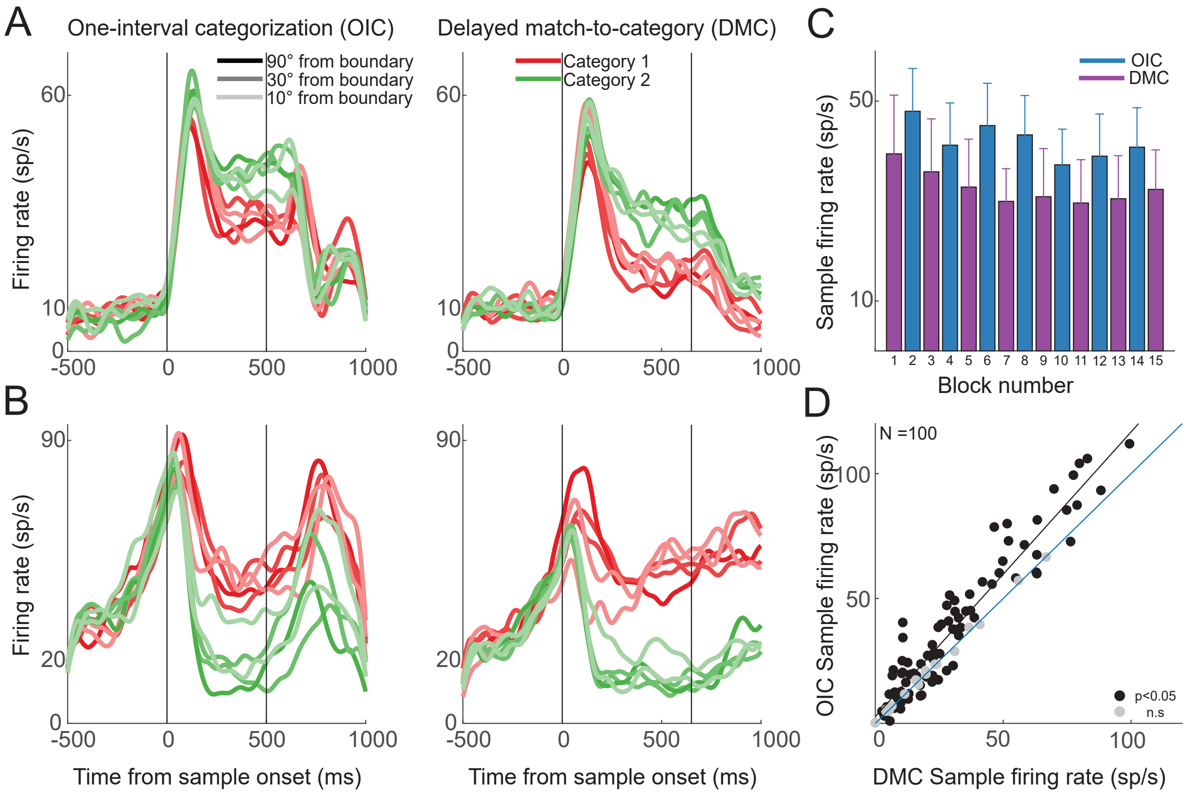

Figure 2.2: Single-neuron LIP responses in OIC and DMC.

(A) Peri-stimulus time histogram (PSTH) of a single example LIP neuron during OIC (left) and DMC (right). Spike times were convolved with a gaussian distribution (σ=20 ms). Colors represent categories and shade represents distance from boundary, lighter shades refer to motion directions closer to boundary. Vertical lines at 0, 500 (OIC) and 650 ms (DMC) indicate sample onset, saccade target onset and delay period onset, respectively. (B) PSTH of another example neuron, details as in (A). (C) Block-wise comparison of firing rates during the shared sample period for the single neuron in (A). Error bars represent s.e.m. (D) Scatter plot comparing OIC and DMC sample firing rates of the population. Black line is the regression line for all the significant neurons (black points, $p<0.05$, two-sample t-test; grey points, not significant (n.s)), blue line is the reference unity line (slope=1).

match) variables.

2.3.4 Neural populations show shared decision representations in both tasks

While single neurons encoded seemingly complex combinations of task parameters in both tasks, we wondered whether population level category representations were encoded similarly in both tasks, since both relied on the same category rules. Specifically, we asked whether the strength and timing of category signals differed between the OIC and DMC tasks by quantifying category selectivity in two ways (i) a ROC-based category tuning index (rCTI) applied to single neurons(Swaminathan et al., 2013), and (ii) population-level linear decoders (Sarma et al., 2015) to classify motion category on a trial-by-trial basis.

As used previously (Swaminathan et al., 2013), the rCTI index quantifies category selectivity for each neuron by comparing neuronal discriminability between pairs of motion directions within the same category (within category difference or WCD) vs. different categories (between category difference or BCD). rCTI values can range from -0.5 to +0.5 with more positive values indicating stronger category selectivity. In both tasks, mean sample rCTI values were shifted towards positive values (indicating categorical tuning) and were significantly greater than values during the fixation-epoch (Figure 2.3A; sample rCTI (p-value of paired t-test) = 0.035 (0.0067) in OIC, 0.06 (4e-05) in DMC; fixation rCTI = 0.003 (0.98) in OIC; -0.0005 (0.08) in DMC). We found that 63% (41/65) of neurons showed significantly greater rCTI values in the DMC than OIC task, although mean rCTI values in the OIC and DMC tasks were not statistically different (paired t-test, p=0.18).

Next, we computed the time-course of rCTI values which revealed that mean rCTI values were greater in the DMC than OIC task, reaching significance during the late sample period (Figure 2.3C). Mean rCTI values were greater in the DMC than OIC task in the motor-response-aligned condition too, in which firing rates were aligned to response onset instead of sample onset (Figure 2.3C, saccade onset in OIC and touch bar release for match trials in DMC). Interestingly, the time-course of response-aligned rCTI monotonically increased to reach its maximum value just before movement onset in both OIC and DMC tasks, showing that categorical tuning peaks just prior to the motor response. In sum, the examination of individual neurons' tuning by the rCTI measure reveals that category selectivity is higher in the DMC task, notably in the late sample period. We also used a population-level decoding approach to confirm our findings from single-neuron rCTI measures (Sarma et al., 2015). We quantified the amount of category information in the LIP population using SVM classifiers trained to classify whether trial-by-trial population responses belonged to category 1 or category 2. The mean and time-course of category decoding accuracies were significantly greater in the DMC than OIC task, particularly during 220-350 ms of the sample period. (Figure 2.4A, category decoder: DMC 95.9%, OIC 85.5%, bootstrap, $p<0.001$).

Greater category selectivity in the DMC task was unexpected because firing rates in the DMC task were lower than in the OIC task. To understand what aspects of tuning accounted for greater category selectivity in the DMC than OIC task, we separately examined the BCD and WCD measures used to construct the category tuning index. This revealed that, while BCD values were comparable in both tasks, WCD values were greater in the OIC than DMC task, especially in the late sample period (Figure 2.3B; bootstrap, $p<0.05$). This indicates that discriminability between directions in different categories was similar between tasks, but that selectivity among directions in the same category was greater in the OIC than DMC task. Hence, the apparent increase in rCTI during the late sample of the DMC task is a direct consequence of more uniform firing rates among the within-category directions resulting in a tendency for more binary-like category selectivity in the DMC task. Using direction decoders devised to classify motion directions within a category, we confirmed greater within-category direction information in the OIC than DMC task (Figure 2.4B, direction decoder: DMC 65.1%, OIC 73.9%, bootstrap, p¡0.01). The time-course of direction decoding also showed greater accuracies for within-category direction information in the OIC than DMC task achieving significance in the late sample period (250-500 ms, Figure 2.3D), suggesting that stronger category tuning in the DMC task is apparently due to tighter clustering of

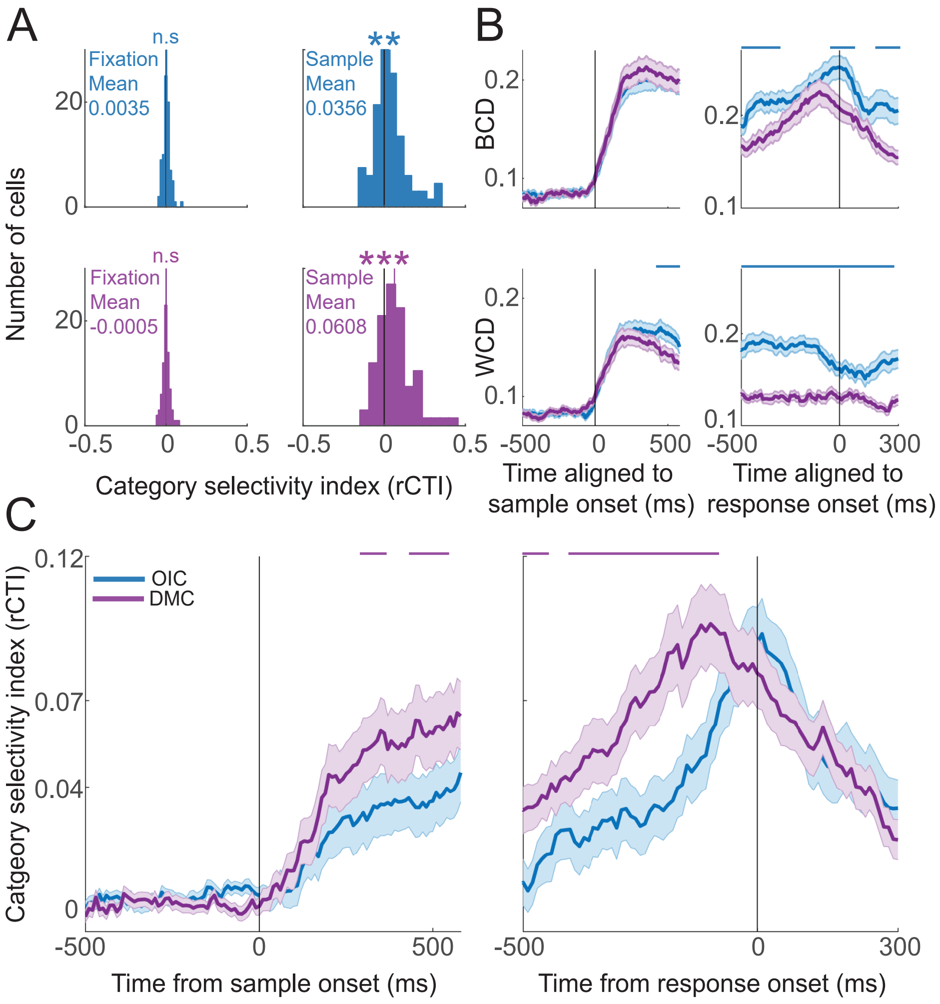

Figure 2.3: Comparing direction and category selectivity in OIC and DMC.
(A) Histograms of rectified category tuning index (rCTI) values during fixation period (left panel) and shared sample period (right panel) in OIC (top panel) and DMC (bottom panel). rCTI values range from -0.5 to 0.5 with -0.5 indicating high direction selectivity and +0.5 indicating high category selectivity. The black line represents zero and the colored line represents the mean of the population. Mean values are significant greater than 0 during sample, but not fixation, ** $p<0.001$, *** $p<0.0001$, n.s not significant,(B) Time course of BCD (top panel) and WCD (bottom panel) computed by aligning firing rates to sample onset (left panel) and response onset (right panel) in OIC and DMC (200 ms sliding window, stepped by 10 ms). Response onset is defined as lever release in match trials in DMC and saccade initiation towards the RF in OIC. Vertical lines indicate sample onset and response onset respectively. Blue line indicates time points where values for OIC are greater than DMC (bootstrap, $p<0.05$), shaded lines indicate s.e.m. (C) Time course of rCTI computed by aligning firing rates to sample onset (left panel) and response onset (right panel) in OIC and DMC. Purple line indicates time points where values for DMC are greater than OIC (bootstrap, $p<0.05$).

responses to directions within each category and larger differences between categories.

2.3.5 Population readout of category information is modulated by task-specific demands

Our results so far reveal similar neuronal category representations in both OIC and DMC tasks with higher category selectivity in the DMC task. But it remains unclear whether functionally shared or distinct neuronal sub-populations mediate category computations in the two tasks. In other words, are the same neurons contributing towards categorical judgments in a similar manner in both tasks? We employed a ”cross-task” population decoding approach to directly compare the extent of overlap in population level representations in OIC and DMC tasks. We trained category decoders on population firing rates from one task (e.g. OIC) and instead of testing it using trials from the same task used to train the decoder (as is usual), we tested the decoder using trials from the other task (e.g. DMC). The cross-task decoder reliably decoded category well above chance, during the sample period in both the OIC and DMC tasks (Figure 2.4a, category cross-decoder: DMC 77.65%, OIC 81.40%) indicating that the LIP population encoded categorical information in a similar manner using highly overlapping pools of neurons in the two tasks. For both tasks, the time-course of category decoding was nearly identical to the case in which the decoder was trained and tested on the same task (i.e. a ”within-task” decoder) in the early sample period (0-250 ms), but diverged in the late sample (250-500 ms) (Figure 2.4E, 2.4F). These findings suggest that shared aspects in both tasks such as category computations are indeed mediated by shared sub-populations of neurons.

We sought to understand in more detail how representations differed in the late sample period. The late sample period is contextually distinct between the two tasks. In the OIC task, it is followed by a saccade for reporting the stimulus category. In the DMC task, it is followed by a 1 second memory delay period, one or more test stimuli and a manual release of the touch bar. Previous studies found that category representations in the delay

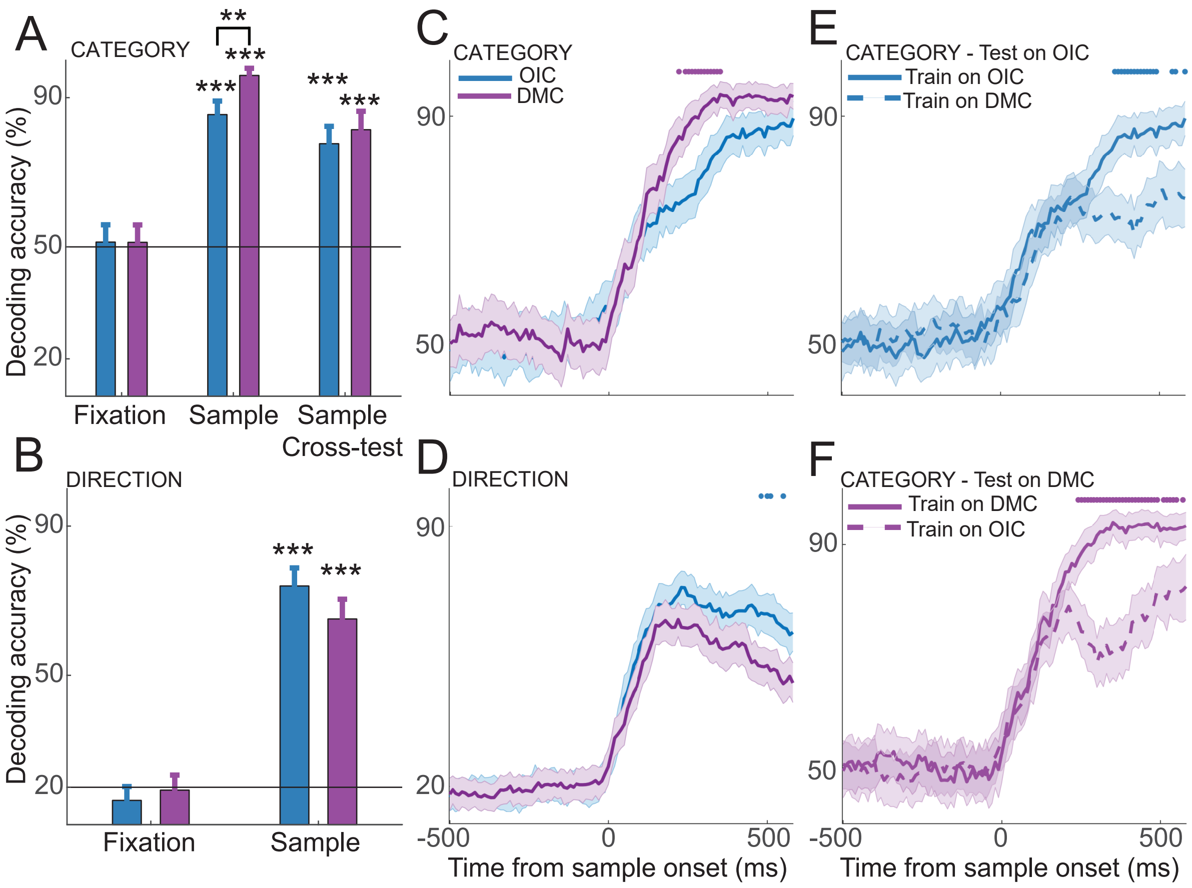

Figure 2.4: Population readout of shared and distinct information in OIC and DMC. (A) Performance of a category decoder during fixation and shared sample period. Error bars indicate s.e.m. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, bootstrap. (B) Performance of a direction decoder during fixation and shared sample period. Error bars indicate s.e.m. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, bootstrap. (C) Time course of performance of a category decoder trained and tested on DMC and OIC during the sample period. DMC has more category information than OIC. Asterix (purple dots) above indicate time points at which curves are significantly different from each other ($p<0.05$, bootstrap). Purple dots represent time points where DMC accuracy is significantly greater than OIC. (D) Time course of performance of a direction decoder trained and tested on DMC and OIC during the sample period. OIC has more direction information than DMC. Asterix (blue dots) above indicate time points at which curves are significantly different from each other ($p<0.05$, bootstrap). Blue dots represent time points where OIC accuracy is significantly greater than DMC. (E-F) Time course of performance of a category decoder tested on one task, but trained on data from the other task. In (E), the category cross-decoder is trained on OIC or DMC, but tested on OIC. In (F), the category cross-decoder is trained on OIC or DMC, but tested on DMC. Asterix represent time points at which the cross-task decoder was significantly worse than the within-task decoder. Shaded lines represent s.e.m. * $p<0.05$, bootstrap.

period are governed by persistent activity as neuronal populations stably maintain their firing rates even after the stimulus is no longer present (Chafee and Goldman-Rakic, 1998; Funahashi et al., 1989). Consequently, we wondered whether category signals diverged in the late sample period because of this greater temporal stability in the DMC than OIC task. To test this, we evaluated temporal stability of category decoding using SVM decoders that were trained at one time point and then tested at all other time points to ascertain if categories were stored in a stable or dynamic format. The cross-temporal decoders revealed greater stability in the DMC task than OIC task in the late sample epoch, as shown by higher decoding accuracies in the DMC task when decoders were trained and tested at different time points (block-like structure in late sample period in Figure 2.5 in the DMC task (right panel)). These observations highlight two findings: first, neuronal activity and representations supporting sensory stimulus evaluation are similar between different task structures (OIC vs. DMC) in the early sample; second, neuronal population representations diverge following the initial evaluation of the sample stimulus, as the two tasks begin to differ in their behavioral demands. Notably, the population-level representations are governed by a stable, persistent categorical code in the DMC task, but a dynamic, transient categorical code in the OIC task.

2.3.6 RNN models recapitulate distinct, task-specific neural codes for category representations

Our analyses on LIP neuronal data established greater category selectivity in the DMC task that manifested as binary-like encoding - compressing variability among directions within a category which accentuates differences between categories. Further, category representations in the DMC task were characterized by stronger persistent activity resulting in a stable categorical code across time. In contrast, during the equivalent time epoch, OIC category representations exhibited graded responses to directions within a category and were characterized by transient activity resulting in a dynamic categorical code. What explains

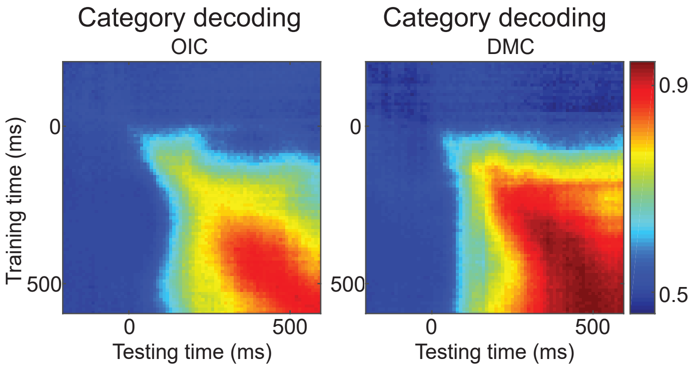

Figure 2.5: Stabiliy and dynamics of category decoding in OIC and DMC.
The stability of category selectivity across time was determined by training the classifier at one time point (y axis) and testing at a second time point (x axis) during the shared sample period. The heat map indicates accuracy of the category decoder at coordinates (training time y, testing time x). The category stability for OIC is shown on the left, and for DMC on the right. Category representations are stable in both tasks with higher levels of persistent activity and temporal stability in the DMC task during the late sample period - evident by the block like structure of high category decoding accuracies indicating better generalization over time.

the differences between the underlying neural codes in the OIC and DMC tasks? We hypothesized that cognitive demands differentially modulate neural activity and dynamics to produce distinct neural codes. Because the sample category needs to be maintained during the delay period in the DMC task, attractor dynamics underlying persistent activity might reformat or compress category-related information to a simpler, binary format by collapsing all directions within a category toward a single scalar value. Specifically, we hypothesized that persistent activity can more easily encode and maintain a binary variable with two attractor states (one for each category) as opposed to a graded variable with multiple attractor states (one for each direction in each category). Inspired by recent approaches using artificial neural network models to understand neural dynamics (Chaisangmongkon et al.) (Masse et al., 2019), we tested whether recurrent neural networks trained to perform the OIC and DMC tasks reflected attractor-like dynamics characterized by persistent, binary category activity in tasks with a mnemonic delay period (i.e. DMC task).

We trained 10 recurrent neural networks to solve the OIC and DMC tasks separately

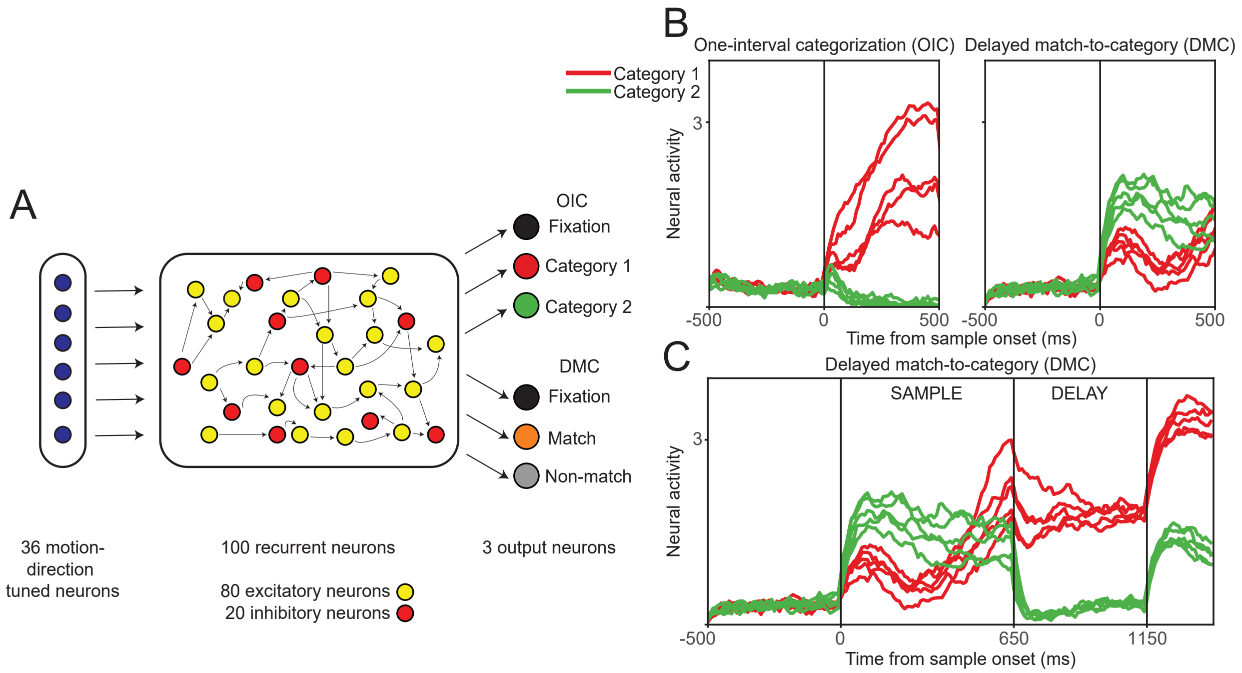

Figure 2.6: RNN model architecture and artificial unit responses in OIC and DMC. (A) Trained RNN models consisted of 36 motion-direction tuned input neurons which projected to 100 hidden neurons (80 excitatory+20 inhibitory) which projected to 3 output units. In networks trained on OIC, the output units corresponded to fixation, category 1 and category 2. In networks trained on DMC, the output units corresponded to fixation, match and non-match. (B) Peri-stimulus time histogram (PSTH) of 2 example RNN hidden units during OIC (left) and DMC (right). Colors represent categories. Vertical lines at 0 indicate sample onset. (C) PSTH of the same example unit shown in B for DMC plotted for the entire trial. Vertical lines at 0, 650, 1150 ms indicate sample onset, delay onset, and delay offset respectively.

with the sequence of task events matched to the tasks performed by the behaving animals. The models received input from 36 motion-tuned neurons that projected to one recurrent layer consisting of 100 hidden units (80 excitatory + 20 inhibitory units) (Figure 2.6A). The excitatory neurons in the recurrent layer projected to 3 output units. The first is a fixation unit which was trained to remain active until response time. The networks were trained so that its other 2 output units in the OIC task were activated for category 1 or category 2 stimuli and its 2 outputs in the DMC task were activated for match or non-match test stimuli, respectively. Trained networks successfully learned to optimize the desired output function and the responses of the hidden layer neurons reproduced features of the neuronal data (Figure 2.6B). Akin to LIP neuronal populations, both OIC and DMC model neurons represented category with similar responses to directions within a category and dissimilar responses to directions between categories. Population-level category decoders classified sample category at near-perfect accuracies in both OIC and DMC networks. To test whether the underlying categorical code was discrete or continuous, we assessed the amount of direction information within each category. Using category-independent direction decoders, we found near-perfect decoding of direction in the OIC network, whereas direction information in the DMC network was significantly lower than the OIC network (Figure 2.7A, direction decoder: DMC 39.7%, OIC 59.77%, bootstrap, p¡0.001). This suggests that the underlying formats of the neural codes reflect task-specific differences with discrete representations in DMC and continuous representations in OIC because of differences in delay-dependent cognitive demands.

To understand whether different mechanisms govern the different neural codes, we analyzed model dynamics through the analysis of fixed points in state-space (Golub and Sussillo, 2018). Fixed points (small red/black dots) correspond to neural activity patterns that are stable when external sensory inputs are turned off. A recent study performed fixed point analysis on RNN models trained to perform the DMC task and found that the dynamics consisted of stable states at the end of the sample period, followed by a tunnel-like trajectory during the delay period (Chaisangmongkon et al.). Similarly, we also found stable states associated with each category at the end of the sample period, in addition to a stable state associated with the fixation epoch. This dynamical analysis revealed that in a delay-based categorization task, stable states (”attractors”) emerge as sample categories need to be maintained in working memory. Further, at the end of the sample period, population activity for all motion directions within a category narrows towards the stable state, thus resulting in compression of stimulus representations in anticipation of the upcoming delay period. In contrast, category-based stable states were not present in networks trained to

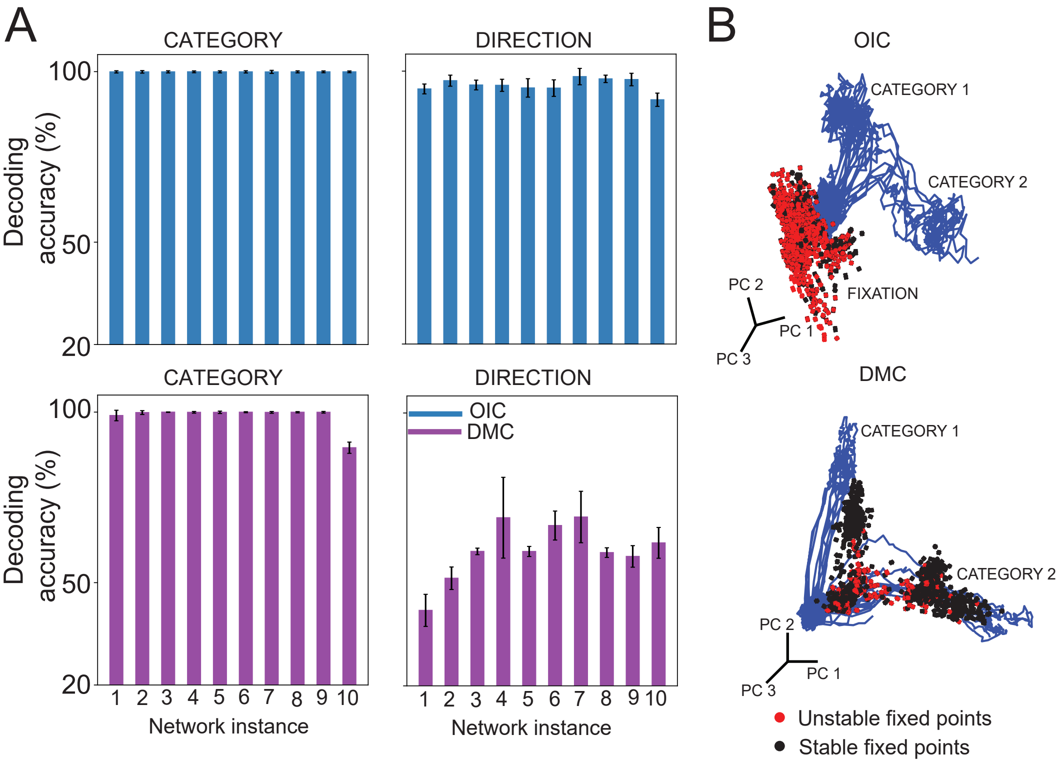

Figure 2.7: RNN decoding and dynamics in OIC and DMC.
(A) Average performance of category (left panel) and direction (right panel) decoders trained and tested on RNNs performing OIC (top panel) and DMC (bottom panel). Decoding accuracies are plotted for 10 independent networks trained each on OIC and DMC. Whereas category decoding is close to 100% in both OIC and DMC networks, we see higher within-category direction information in OIC networks than DMC networks. Error bars represent s.d. (B) Neural trajectories of one model trained on OIC (left panel) and DMC (right panel) separately. Blue traces indicate PCA trajectories of the hidden layer neurons during the shared sample period. Black and red dots indicate unstable and stable fixed points of the networks computed on the hidden layer neural activity.

perform the OIC task in which there are no demands on working memory. These results are consistent with our previous observations that DMC representations were more persistent and temporally stable than OIC representations (Figure 2.7B). Taken together, the results from the modeling work support our hypothesis that attractor-like dynamics underlying working memory compresses stimulus representations.

2.4 Discussion

We investigated the neural correlates of abstract categories during categorization task switching as we flexibly varied the behavioral demands in each task context. Monkeys alternated between two categorization tasks, both of which required that they categorize visual motion stimuli according to the same category boundary. However, the tasks varied in the sequence of task events upon which decisions were based, in the effector used to report decisions, and in the cognitive demands unique to each context. We investigated whether neuronal circuits in LIP are involved in categorical decisions in both task contexts, or whether they are specialized for the cognitive processes specific to each task such as motor response modalities, working memory, or sequential comparison. We found that LIP activity robustly encodes motion category in both tasks with similar category representations evident in the early sample period but with distinct patterns of encoding in the late sample period. The differences in the late sample period were explained by a difference in population representations of the categorical code in both tasks. In the delayed matching task, category encoding showed greater levels of persistent activity and temporal stability, presumably in preparation for the imminent delay period. In contrast, encoding during the OIC task was more transient, presumably as stimulus and decision information was rapidly routed to motor circuits with little demands on working memory. In sum, our results suggest that LIP circuits mediate flexible task switching through generalized category representations which are rapidly reconfigured and specialized for working memory-based computations.

Decades of research in LIP has focused on perceptual decision-making with most studies employing behavioral tasks with relatively simple stimulus-response (S-R) structures in which noisy sensory stimuli are mapped directly to motor responses (**?**Antzoulatos and Miller, 2011). More recent studies have investigated flexibility beyond these simple S-R tasks. Some tasks have included more complex sensorimotor mappings which dissociate decisions from actions (Bennur and Gold, 2011; Gold and Shadlen, 2003). Other studies employed multiple response modalities to report the same decisions (de Lafuente et al., 2015), or modulated behavioral demands in both cued (Kumano et al., 2016; Mante et al., 2013; Siegel et al., 2015) and uncued (Sarafyazd and Jazayeri, 2019; Asaad et al., 1998) scenarios. Task flexibility has also been studied in PFC and PPC by training animals to alternate between tasks with different sensory stimuli, different decisions, or different motor responses in order to understand how circuits route information according to shifting behavioral demands (Merchant et al.; Roy et al.; Miller and Cohen, 2001). Here, we studied flexible decisions by extending the S-R task structure to complex sensorimotor associations. In both the OIC and DMC tasks, the category decision is decoupled from the eventual motor action, thus allowing us to delineate the role of LIP neuronal populations in decision-making and response selection. In contrast to recent work finding no impact of LIP inactivation on a motion discrimination task during simple S-R decisions (Katz et al.), a study from our group found that inactivation of LIP neurons indeed impairs behavioral performance in a flexible motion discrimination task (Zhou and Freedman, 2019). These findings suggest that incorporating flexibility in classical task structures can provide a more nuanced understanding of the circuits that mediate decision-making. However, a potential issue with incorporating flexibility in task structures is that it might engage new and task-specific neural and behavioral mechanisms that might not generalize more broadly to other task scenarios.

In concert with its involvement in perceptual decision-making, it is well established that LIP plays an important role in spatial functions such as directing spatial attention (Gnadt and Andersen, 1988; Bisley and Goldberg; Colby et al.; Herrington and Assad) and eye movements (Snyder et al.) to behaviorally relevant locations in the visual field. More recently, studies have highlighted a distinct role of PPC in non-spatial cognitive functions such as task context, rules (Stoet and Snyder, 2004), and abstract categories of motion (Freedman and Assad, 2006), spatial locations (Masse et al.), and shape (Fitzgerald et al., 2011). A recent study directly compared spatial and abstract category encoding in LIP as monkeys were trained to make saccades toward or away from RF during the delay period of the DMC task8. Importantly, the direction of the eye movement had no relationship with the monkeys category judgments, but was expected to modulate activity due to LIPs well-known spatial encoding. This revealed that spatial and cognitive (i.e. category) representations are independently encoded in LIP neuronal activity. In the current study, the OIC task combined LIPs spatial and category-related functions to ascertain whether category encoding generalizes beyond the DMC task. This revealed that similar category representations emerge in both the OIC and DMC tasks, arguing for a generalized participation of LIP in encoding abstract cognitive variables during diverse tasks. A caveat is that the category signals observed in the OIC task may have emerged as a result of prior training on the DMC task, before the monkeys learned the OIC task. However, we posit that it is more likely that the brain employs similar mechanisms and computations to solve similar tasks. Thus, we predict similar category signals would be evident even when animals are trained only on the OIC task.

Abstract, categorical signals have been observed in multiple brain regions including the prefrontal cortex (Freedman et al., 2001; Swaminathan and Freedman, 2012), parietal cortex (Swaminathan and Freedman, 2012; Freedman and Assad, 2006), and the inferotemporal cortex Freedman et al. (2003); Meyers et al. (2008). Early studies found that neurons in prefrontal cortex encode shape categories with binary responses, especially in the delay period of a DMC task1. In contrast, neurons in IT cortex encoded shape categories in a different format single neurons showed partial category encoding (mixed with visual feature representations) in the sample period with a lack of binary delay activity (Freedman et al., 2001, 2003). These observations led to the conclusion that PFC neurons encoded behaviorally relevant categories in their responses, but IT neurons were engaged more in high-level visual processing. However, later studies showed that IT populations encoded stimulus category in a dynamic format by which categories could be decoded, but the format was not binary or discrete (Meyers et al., 2008). This work suggested that IT neurons encode differences between categories but they also encode feature information through within-category differences, thus encoding categories in a continuous format. Further, the lack of binary delay activity in IT cortex could potentially reflect the limited role of IT cortex in working memory.

Previous studies on categorization have proposed that the abstract, binary category signals that have been observed are the output of a neural computation which transforms sensory representations into category signals. In other words, the binary signals represent categorical information about the stimulus. However, our results from this study provide an alternative interpretation. We put forward the idea that the specific format of category signals binary or discrete signals observed in the DMC task may be a consequence of the short-term memory demands of delayed matching tasks. This idea is consistent with the observations of near-binary categorical encoding in both PPC and PFC, areas that encode stimulus information during working memory delays. Further, a recent study found that LIP neurons encode shape pair associations in a discrete manner in a delayed pair association matching task with shape stimuli (Fitzgerald et al., 2011). Interestingly, the same neurons that showed strong selective delay activity for motion categories also showed strong separation between three different shape pair associates, again consistent with the idea that categorical delay activity is related to maintenance of task-relevant information in working memory (and also that such effects can extend to three groups or attractor states). Finally, a recent study found binary-like separation between abstract rules (match vs. non-match) in PFC during the delay period as cue information (both auditory and visual) was maintained during the delay before the test stimulus appeared (Wallis et al., 2001). This suggests that the requirement to maintain information in working memory leads to reformatting of stimulus representations, specifically by compressing variability within an abstract set of objects (category, rules or associations) to a single neural state or attractor.

Our results suggest that parietal circuits support flexible task switching through generalized categorical encoding in multiple categorization tasks with different cognitive demands. Further, behavioral demands such as working memory reorganize the format of neural category encoding to a more discrete or binary format in delayed matching tasks, which is likely mediated by network attractor dynamics which likely underlie persistent neural activity during tasks requiring short term working memory. This provides novel insight into the interplay between neuronal encoding underlying stimulus representations, categorical decisions, and short-term memory, as well as insight into how flexible task switching influences these dynamics.

2.5 Methods

2.5.1 Behavioral tasks and stimulus display

Two male monkeys (*Macaca mulatta*) were trained to alternate between a one-interval categorization (OIC) task and a delayed match-to-category (DMC) task in blocks of 20 trials (monkey M) or 30 trials (monkey B) based on a colored cue presented at the start of a trial. In both tasks, monkeys were trained to categorize the same random-dot motion stimulus into one of two categories. Stimuli were circular patches (5 diameter) of 100 high-contrast dots that moved with 100% coherence at a speed of 12/s. Motion stimuli were created by dividing 360° of motion directions into two categories based on an arbitrary learned category boundary. Six evenly spaced motion directions (60° apart - 15°, 75°, 135°, 195°, 255°, 315°) were used as sample and test stimuli in addition to four directions that were 10° away (35°, 55°, 215°, 235°) from the category boundary. In OIC, monkeys indicated the category membership of a sample stimulus by making an eye movement to a colored target associated with the category. In DMC, monkeys indicated whether sequentially presented sample and test stimuli matched in category by releasing/holding a manual lever for match/non-match trials respectively.

In the OIC block, every trial started with a yellow colored 0.2° fixation point at the center of the display. After gaze fixation was maintained for 500 ms (within a 2° fixation window), a sample stimulus was presented for 500 ms in the receptive field (RF) of the recorded neuron. At 500 ms after sample onset, two colored targets appeared one in the RF of the recorded neuron and the other 180° opposite to the RF. After targets appeared, the monkeys were trained to make a saccade and maintain fixation for 300 ms on the red/green target to indicate the category of the sample stimulus as category 1/category 2, in order to receive juice reward. Critically, on every trial, the red and green saccade target locations were counterbalanced between the two possible locations (inside RF or outside RF). This ensured that the saccade directions were not correlated with the category of the sample stimulus, thus, allowing us to disambiguate sensory and/or decision signals from motor signals in the sample period.

In the DMC block, every trial started with a white colored 0.2° fixation point at the center of the display. After gaze fixation and lever press was maintained for 500 ms (within a 2° fixation window), a sample stimulus was presented for 650 ms in the receptive field of the recorded neuron, followed by a 1000 ms delay and a 650 ms test stimulus. If the sample stimulus and the test stimulus matched in category, the monkeys released the manual lever using a hand movement, in order to receive a juice reward. If the sample and the test stimulus did not match in category, a second test stimulus appeared, which was always a category match to the sample, and the monkeys were required to release the lever.

Gaze positions were measured using an Eyelink 1000 optical eye tracker (SR Research) at a sampling rate of 1 KHz and stored for offline analysis. Task events, stimulus display, timings and reward delivery were controlled via a MATLAB-based toolbox, MonkeyLogic. Stimuli were displayed on a 21-inch color CRT monitor (1280x1024 resolution, 75 Hz refresh rate, 57 cm viewing distance).

2.5.2 Electrophysiological recordings

Two male monkeys were implanted with a headpost and a recording chamber. The recording chamber was implanted over the intraparietal sulcus centered 3 mm posterior to the interaural line, 12 mm lateral from the midline. Stereotaxic coordinates were determined from anatomical MRI scans obtained prior to headpost and chamber implantation. All surgical procedures were in accordance with the University of Chicago's Animal Care and Use Committee and US National Institutes of Health guidelines. LIP recordings were conducted using 75μm tungsten microelectrodes (FHC), a dura-piercing 27 Ga guide tube and an electronic micromanipulator (NAN Instruments). Neurophysiological signals were amplified, digitized, and stored for offline spike sorting (Plexon) to verify the quality and stability of neuronal isolation within a recording session.

2.5.3 RF mapping and stimulus placement

In every recording session, we tested and recorded the activity of well-isolated neurons during the delayed memory-guided saccade. Neurons were considered to be in LIP if they showed spatially selective visual, delay, and/or peri-saccadic activity or were located in between such neurons. The spatial location on the screen which elicited the highest activity was chosen as the receptive field (RF). Stimuli in the OIC and DMC tasks were always presented in the RF of the recorded neuron(s). The eccentricity of stimulus placement for LIP recordings ranged from 5-9°. The depth of LIP recordings ranged from 4-11 mm from the surface of the dura.

2.5.4 Delayed memory-guided saccade task

In this task, every trial started with a white colored 0.2° fixation point at the center of the display. After gaze fixation was maintained for 500 ms (within a 2° fixation window), a visual target was presented for 300 ms at a fixed eccentricity in one of 8 possible peripheral locations. The brief target flash was followed by a 1000 ms delay after which the fixation point disappeared and cued the monkey to make a saccade to the remembered location.

2.5.5 Data Analysis

All our analyses were conducted on correct trials, excluding trials with incorrect responses, fixation breaks, and early responses. Unless otherwise stated, all our results were qualitatively and quantitatively similar in both monkeys. Thus, we combined data sets from both monkeys for all analyses.

Epoch-based analysis

The fixation period was a 500-ms epoch prior to sample onset in both the OIC and DMC tasks The sample period was a 500-ms epoch in the OIC task and a 650-ms epoch in the DMC task, both epochs offset by 80 ms to account for neuronal response latency. In the OIC task, the saccade/choice period was a 250-ms epoch starting at saccade target onset which was 500 ms after sample stimulus onset. In the DMC task, the delay period was a 100-ms epoch beginning at sample stimulus offset. Early and late sample periods were 250-ms epochs beginning 80 ms and 330 ms after stimulus onset. In the DMC task, the match/non-match or choice period was a 250-ms epoch starting at test stimulus onset.

ROC-based category tuning index (rCTI)

Category selectivity was quantified using a ROC-based metric for individual neurons that compared neuronal response distributions for motions directions in the same vs. different categories. The rCTI index was computed as the difference between two quantities the between category difference (BCD) and the within category difference. BCD was measured as the average value of the area under the curve (AUC, rectified to a range of 0.5-1) calculated for firing rate distributions for pairs of motion directions in different categories. Similarly, WCD was measured as the average AUC for firing rate distributions for pairs of motion directions within the same category. To account for the inherent motion direction tuning of LIP neurons, we considered only pairs of directions for which the angular difference was equal in BCD and WCD. Out of 25 (5x5 directions per category) possible direction pairs for BCD and 20 (2 categories x ${}^{5}C_{2}$) possible direction pairs for WCD, the angular difference of 12 direction pairs were equal in BCD and WCD. Further, we also equalized the number of direction pairs in BCD and WCD for each angular difference value. For example, there were 2 pairs of directions with an angular difference of 20° in BCD, and 4 pairs of directions with the same angular difference of 20° in WCD, we multiplied the rectified AUC value of BCD by 2 to ensure that equal numbers of pairs contributed towards the ROC comparison. rCTI values ranged between -0.5 (directions within a category are much more distinct from each other compared to directions between categories) to +0.5 (directions in different categories are much more distinct from each other compared to directions within categories).

Population decoding

Population decoders were constructed as SVM classifiers that were trained and tested to classify task relevant variables such as motion direction, category, and task. In all population decoding analyses, a four-fold cross-validation approach was used in which 75% of the data was used for training and the remainder for testing.

We measured direction information that was independent of category information by constructing separate direction decoders for each category and averaging performance across both decoders. We randomly sampled with replacement 20 trials for each direction in a category and used an SVM multi-class classifier to classify one among 5 directions within a category (chance performance = 20%).

We measured category information that was independent of direction information by constructing category decoders in which training and testing sets consisted of motion directions with equal angular distance from the boundary in both categories. This resulted in 4 category decoders with non-overlapping sets of motion directions for training and testing. For each motion direction, we randomly sampled 20 trials with replacement and trained the classifier to classify test response vectors as responses for category 1 or category 2 (chance performance = 50%). For the cross-task category decoding analysis, we constructed category decoders that were trained on trials from one task (say, OIC) and tested on trials from the other task.

Temporal stability analysis

We extended the category decoding approach to compare the amount of category information encoded at different time points in the trial. To assess if category information is maintained in population activity over time, we trained a category decoder at one time point in the trial and tested it at all time points in the trial. If the category decoders were able to decode category information at different time points in the trial, this indicated that the patterns of neural activity were similar over time. A block-like structure of high decoding accuracy indicates stable patterns of category selectivity over the time-course of a trial.

Statistics

Selectivity for various task-relevant variables such as motion direction, motion category, saccade direction, and match/non-match were computed using a one-way ANOVA with significance measured at p¡0.05. Motion direction selectivity of individual neurons was computed using a one-way ANOVA (10 motion directions, $p<0.05$). Motion category selectivity of individual neurons was measured using a one-way ANOVA (2 categories, $p<0.01$). Saccade direction selectivity was measured using a one-way ANOVA (2 saccade directions, inside RF vs. outside RF) and match/non-match selectivity was also measured using a one-way ANOVA (match vs. non-match). Comparisons between proportions of direction selective or category selective neurons were conducted using a chi-square test.

The statistical significance of differences in rCTI, BCD, and WCD was computed through a shuffle analysis. To obtain a null distribution, we shuffled the direction labels, calculated rCTI and repeated this process 1000 times for both the OIC and DMC tasks. The rCTI values were significantly different from 0 if it was greater than 95% of values from the null distribution.

The statistical significance of differences in decoding accuracies for both direction and category information in OIC and DMC were determined using a bootstrap analysis. For the epoch-based decoding analysis, we constructed a null distribution by shuffling the labels for motion direction and then calculated decoding accuracies. Average classification accuracy was considered significantly greater than chance if the value was greater than 95% ($p<0.05$) or 99% ($p<0.01$) of the values from the null distribution. For time-course based decoding analysis, we used a bootstrap analysis to compare classification accuracies between OIC and DMC at each time point. Average classification accuracy was considered to be significantly different between OIC and DMC if 97.5% ($p<0.05$) or 99.5% ($p<0.01$) of values from one task were greater than the other task.

Network models

The details of the recurrent neural network modeling used in this study have been described previously (Masse et al., 2019) and are described in brief in the following sections.

Neural networks were trained and simulated using the machine learning framework Tensorflow. In both tasks, the stimuli were visual motion stimuli moving in one of 10 directions. Networks trained on both OIC and DMC tasks consisted of 36 motion-direction tuned input neurons. Networks trained on the OIC task also contained 2 target units that specified the location of response targets (left/right) and 4 fixation neurons that were active until 500 ms from sample onset. Networks trained on the DMC task contained 4 fixation neurons that remained active until the end of the delay. These input neurons projected to a hidden layer consisting of 100 recurrently connected neurons (80 excitatory + 20 inhibitory), which in turn projected to 3 output units. Networks trained on the OIC task contained 3 output neurons 1 fixation, and 2 neurons that corresponded to category 1/category 2. Networks trained on the DMC task also contained 3 output neurons 1 fixation, and 2 neurons for match/non-match responses. 10 separate networks were trained on the OIC and DMC tasks each with task events matched to the tasks performed by behaving animals. The DMC networks were trained to maintain neural activity over a 500 ms delay period instead of the 1000 ms delay used in the neurophysiological experiments. The networks were trained with the following parameters which were kept constant for both tasks: learning rate: 0.02, neuron time constant: 100 ms, time step: 10 ms, standard deviation of input noise: 0.1, standard deviation of recurrent noise: 0.5. gradient batch size: 256, number of batches: 2000. The networks were trained through stochastic gradient descent (Adam version). All trained networks had accuracy greater than 95

Fixed-point analysis

For finding fixed points in the trained networks, we implemented FixedPointFinder (Golub and Sussillo, 2018), an open-source Tensorflow toolbox for finding fixed points and linearized dynamics in trained RNNs. Hidden layer activity of trained RNNs was used to find fixed points separately for the OIC and DMC networks. The fixed points were calculated for 1024 initial points that were randomly sampled across trials, neurons, and time within a trial. Joint optimization was utilized to find all fixed points simultaneously.

CHAPTER 3

VISUAL IMAGE FAMILIARITY LEARNING AT MULTIPLE TIMESCALES IN THE INFEROTEMPORAL CORTEX

3.1 Abstract

We have an exceptional ability to learn to recognize familiar visual items and to discriminate them from novel ones. This ability is thought to rely on experience dependent changes in neuronal representations in the ventral visual pathway, particularly the inferotemporal cortex (IT). IT neurons differentiate between novel and familiar stimuli, with greater average activity for novel images. Yet, we do not understand how these representations arise, largely because changes in neuronal activity have not been measured during long-term familiarity learning. Here, we examine the dynamics of learning at two distinct time scales, while recording from IT neuronal populations both over long-term (across days to weeks) and short-term (within a single daily session) learning. In long-term learning, monkeys were familiarized with initially novel images that were presented many times over multiple days, while in short-term learning, they were familiarized with multiple repeated presentations of initially novel images within a single session. Across days, we observed a graded reduction in average firing rates as familiarity increased. Overall, neuronal responses were greatest for initially novel, followed by intermediately familiar, and the weakest for highly familiar stimuli. This mean reduction in response developed over several days and was best estimated by an exponential function with a time constant of 4 days (or 135 views). In contrast, within a single day, firing rate distributions of single neurons remained mostly stationary as responses for most neurons did not significantly change on this time scale, while a significant minority showed a uniform decrease or uniform increase from the beginning until the end of the session. Together, our results give insight into how experience affects visual representations in IT across multiple timescales.

3.2 Introduction

Humans rapidly recognize and recall previously seen visual stimuli with remarkable detail and accuracy. We also readily distinguish familiar stimuli from never-before seen novel stimuli without explicit instruction. This ability is thought to arise from learning dependent synaptic changes in the inferotemporal cortex (IT), an area in the ventral visual stream critical for visual object perception and recognition (Gross et al., 1972; Tanaka, 1996). Indeed, neurons in macaque IT cortex discriminate between novel and familiar stimuli by responding more strongly to novel images and less strongly to familiar images that have been presented many hundreds of times (Li et al., 1993; Anderson et al., 2008). Previous studies have typically compared IT responses to highly novel images that were never seen before with highly familiar images that were previously viewed hundreds of times. While these studies provide a static window into how IT populations change after long-term experience, we do not yet understand how these representations arise dynamically during familiarity learning.

How do initially novel stimuli become familiar? Early studies have observed that repeated presentations of initially novel visual stimuli lead to a reduction in the average response of most neurons(Li et al., 1993; Xiang and Brown, 1998). A recent study demonstrated that even a single exposure of an image results in a decrease in the IT population response and a weighted linear readout of IT predicted behavioral performance in a recognition memory task (Meyer and Rust, 2018). While this study employed a behavioral task in which monkeys indicated with an eye movement whether an image was novel (never viewed before) or familiar (previously viewed exactly once in that session), most studies of familiarity in IT have used passive fixation to present images. Response reductions in IT neurons have been observed during active tasks and passive exposure. Further, comparisons between visual responses to novel and familiar stimuli have demonstrated that responses to familiar stimuli are more selective than responses to novel stimuli as neuronal populations respond more sparsely to familiar stimuli. Thus, long-term familiarity leads to a decrease in neuronal responses to non-preferred images, but an increase in the responses to neurons' most preferred images.

Accordingly, a recent study showed that in putative excitatory neurons, but not in inhibitory neurons, maximum responses (responses to the most preferred image) were higher for familiar stimuli than novel stimuli, suggesting that image familiarization sharpens neuronal dynamics for familiar stimuli (Woloszyn and Sheinberg, 2012).

Previous studies have employed a wide range of time scales for familiarity with hundreds or thousands of exposures over a period of weeks to months Freedman et al. (2006); Meyer et al. (2014). Novel stimuli, on the other hand, have been presented once per session or few times (5-20 views) within a session. This has led to comparisons between neural representations at discrete stages, highly novel stimuli on the one end with highly familiar stimuli on the other, thereby providing little insight into how learning continually progresses through tens, hundreds, and thousands of views. A recent study examined the neural correlates of familiarity in V2 as monkeys repeatedly viewed a large set of images (Huang et al., 2018). They assessed the activity of V2 populations through semi chronic multielectrode arrays during familiarization as initially novel images were repeatedly presented around 300 times. In addition to 25 unique novel images, 25 initially novel images comprised the familiarization set and were presented many times over multiple days in a passive fixation task and they found that familiarity suppression occurs in V2 after several days of training. However, no prior study has directly compared the dynamics of long-term familiarity learning in IT across multiple time scales.

In this study, we investigated dynamics over multiple time scales of learning and familiarization as initially novel stimuli became familiar. We trained monkeys to perform a dimming detection task as we repeatedly presented a large database of images at different time scales of familiarity to the monkeys. During familiarization, we recorded simultaneously from small populations of neurons in IT cortex. In the first experiment, we examined the impact of long-term familiarity with stimuli at three different familiarity levels - novel stimuli (tens of times), intermediately familiar stimuli (hundreds of times), and highly familiar stimuli (thousands of times) - and presented them multiple times over multiple days.

In the second experiment, we examined the impact of short-term familiarity by repeatedly presenting novel stimuli and highly familiar stimuli ∼30-40 times within a single recording session.

We found that familiarity over both long- and short-time scales has a pronounced effect on neuronal response selectivity and dynamics in IT. Over long time scales, as stimulus familiarity increased, we observed a gradual decrease in the average population activity of IT neurons. The firing rate decrease developed over several days and the time course of this effect could be explained by an exponential function with a rate constant of 4 days or 150 views. Over short time scales, as stimulus familiarity increased, average firing rates to novel stimuli showed a slight decreasing trend whereas responses to familiar stimuli remained constant. Further, IT neurons also exhibited familiarity-independent changes in neuronal responses at different time scales within a session - i) Over 1-2 hours, average baseline firing rates increased within a session, and ii) Over the course of a trial, average firing rates were suppressed around image onset, suggesting that multiple, rapid, image presentations modulate dynamics of IT responses. Taken together, our results suggest that the neuronal impact of familiarity develops slowly over many repeated presentations of images across days with a minor influence of short-term familiarity.

3.3 Results

3.3.1 RSVP Dimming Detection

To investigate the timescales of visual learning in IT, we first familiarized two monkeys with a large library (∼500 images) of images of natural and man-made objects and scenes. During familiarization, monkeys were trained to perform a dimming detection task in which they had to detect a decrease in luminance in a series of rapidly presented images which varied in their degree of familiarity (Figure 3.1). The images were presented at central fixation for 400 ms each with a 150 ms inter-stimulus interval. On each trial, the number of images presented before the dimmed image ranged from 1 to 5. On half of the trials, one of the stimuli decreased in luminance and the monkeys were rewarded for releasing a lever when they detected the dimming. On the remaining half of the trials, none of the images presented in the sequence dimmed and the monkeys were rewarded for withholding a response. The familiarization phase lasted for 3-5 months during which each image was viewed at least 1000-2000 times by both monkeys.

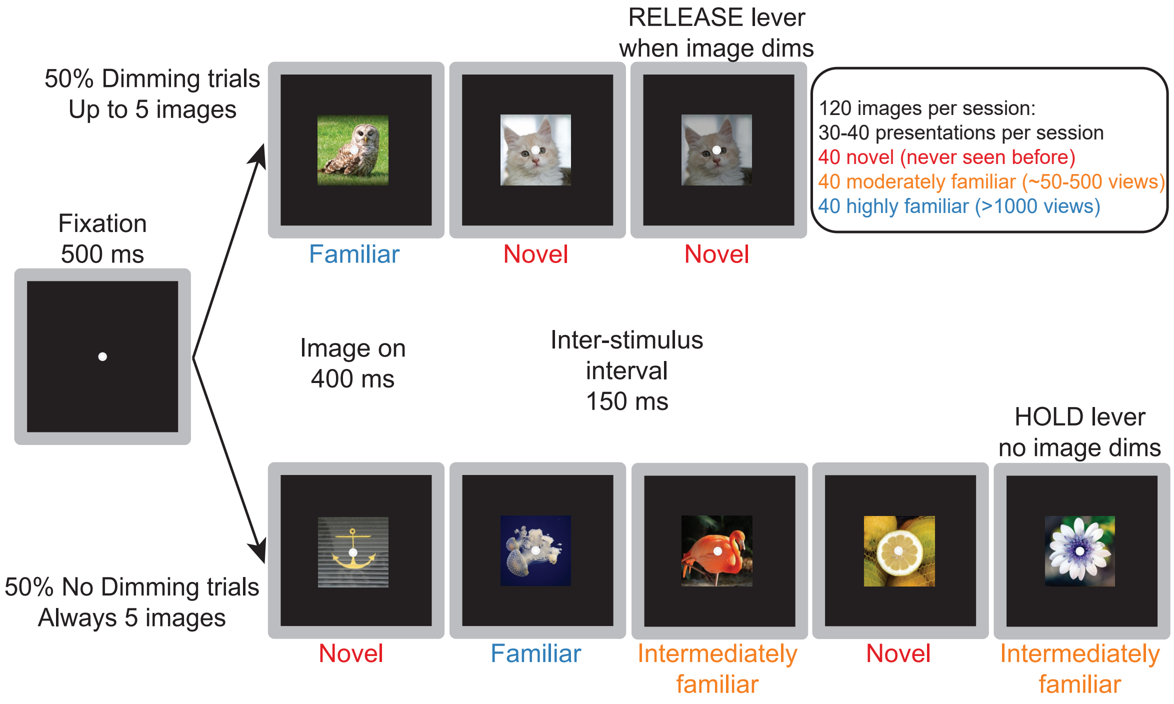

Figure 3.1: RSVP Dimming Detection task.
Monkeys were trained to release a lever when they detected a decrease in the luminance of an image (i.e. dimming). Each trial started with a 500 ms fixation period, followed by image presentation for 400 ms with an inter-stimulus interval of 150 ms. On 50% of the trials, up to 5 images were shown and dimming occurred on any of of the five images, after which the trial ended. On the rest of the trials, 5 images were always presented and no dimming occurred and monkeys were rewarded to hold the lever through all the images. The images varied in their levels of familiarity to the monkey - novel, intermediately familiar, and highly familiar images - and were presented in an interleaved manner.

After familiarization, in every recording session, we monitored the activity of 2-24 IT neurons simultaneously using linear arrays (N=442 from monkey B, N=407 from monkey S) during the dimming detection task. In each recording session, 120 unique (40 novel, 40 intermediately familiar, 40 highly familiar) stimuli were presented ∼30 times. The images varied in their familiarity to the monkey as follows - novel (∼1-30 times, viewed on the same day), intermediately familiar (∼50-800 times, viewed over a week), and highly familiar (∼1000s times, viewed previously over 3-5 months during training). The intermediately familiar stimuli were retained in the image set over several recording sessions until they were presented ∼800 times (∼10 days), while the novel stimuli were replaced with a new set of novel images in each recording session. Consequently, we assessed the emergence of familiarity-related changes in neuronal responses as a function of long-term (across days to weeks, experiment 1) and short-term (within a single daily session, experiment 2) familiarization.

3.3.2 Impact of long-term familiarity on neuronal encoding

We compared neuronal responses between novel, intermediately familiar, and highly familiar stimuli. Consistent with previous reports, we observed a decrease in the average neuronal response with familiarity, both within and across sessions. While the average transient response was comparable for all stimuli, the average sustained response was highest for novel, followed by intermediately familiar, and the lowest for highly familiar stimuli (Figure 3.2A). This suggests that familiarity was accompanied by a gradual decrease in mean firing rates over several daily sessions.

To determine the time-course over which firing rates decreased over multiple days, we compared visual responses of neurons to novel and intermediately familiar images presented on the same day. On a given day, we averaged the visually evoked responses of all neurons across all novel and intermediately familiar images and tested for a relative decrease in firing rates for intermediately familiar images with respect to novel images presented on the same day. Although we recorded simultaneously from multiple neurons on a given day, our recordings were acute, targeting slightly different IT locations from day-to-day. Because of this, the neuronal populations recorded from day-to-day were different. However, our

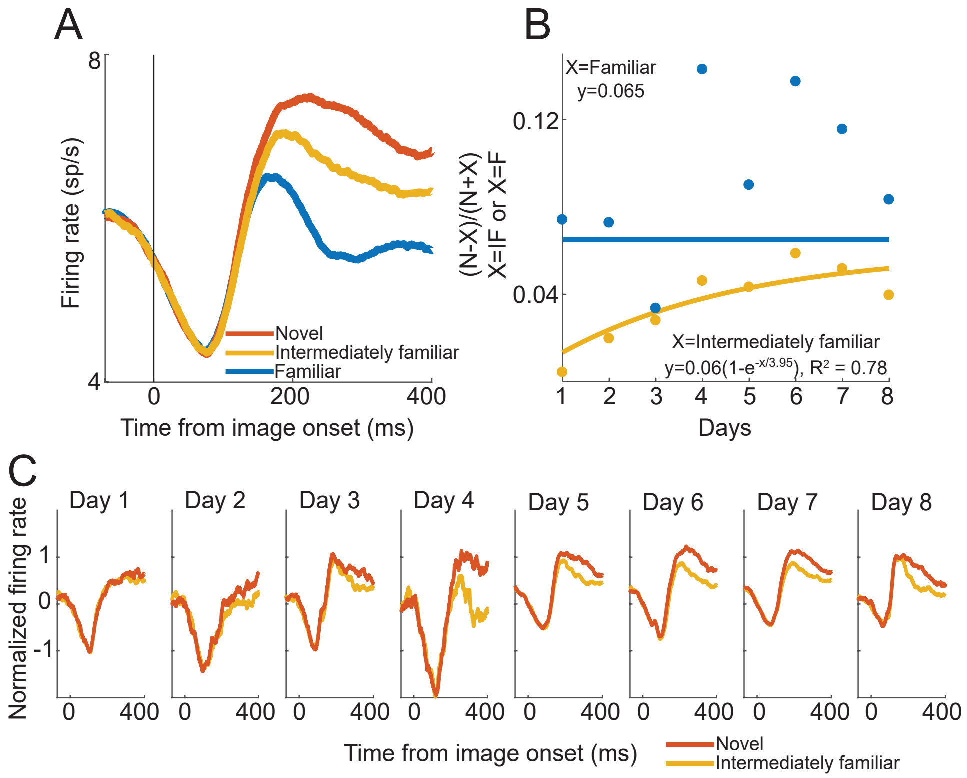

Figure 3.2: Impact of long-term familiarity on neuronal encoding.

(A) Peri-stimulus time histogram (PSTH) of visually selective IT neurons ($N=539$) to novel (seen $\sim$10s of times), intermediately familiar (seen $\sim$100s of times), and highly familiar images (seen $\sim$1000s of times). Firing rates were highest for novel, followed by intermediately familiar, and weakest for highly familiar images. Firing rates were smoothed using a 20-ms causal boxcar kernel. (B) Normalized firing rate difference for intermediately familiar (orange) and familiar (blue) images with respect to novel images as a function of number of days each IF or F image had been viewed before the recording day. Positive values of the index indicate that novel responses are higher than familiar responses. For intermediately familiar images, the normalized difference increases over days and the rate of change in this difference is best estimated by an exponential function (orange curve with the functional form of the equation). For familiar images, the normalized difference remained constant indicating a point of saturation. The rate of change in this difference is best estimated by a line with positive intercept and zero slope (blue line with the constant line equation). N refers to novel, IF refers to intermediately familiar, and F refers to highly familiar, X refers to a variable that can take values of F or IF. (C) Normalized PSTHs to N and IF images on each day. The difference between N and IF images increases over days, reaching statistical significance first on day 3 and staying significant till day 8 ($p<0.05$, GLM). For each day, the curves represent the PSTH averaged across all neurons recorded on that day, averaged over all images in a given category (N and IF). Firing rates on each day were normalized by subtracting the baseline at $t=0$ and dividing by peak minus baseline, with both peak and baseline measured on mean population activity combined across all images in a category. Firing rates were smoothed using a 20-ms causal boxcar kernel.

recordings from multiple neurons on a given day (min 2-max 24 neurons per day) and the large image-set allowed us to minimize the influence of variations in image selectivity within a day. We controlled for the day-to-day variations in the neuronal populations by comparing changes in the response strength of intermediately familiar images with respect to that of novel images recorded on the same day. Upon visually inspecting the population histograms for novel and intermediately familiar images, we discerned that the familiarity-related decrease in neuronal response emerged over the course of several days of familiarization (Figure 3.2C). We observed a progressive increase in the difference between responses to novel and intermediately familiar images in the first few days, after which the firing rate difference saturated.

We quantified neuronal selectivity for familiarity by computing a normalized difference index which measured the difference between mean firing rates for novel and intermediately familiar stimuli divided by their sum, calculated separately for all neurons and all images on a given day. To assess the learning rate for the development of novelty/familiarity selectivity, we tracked the evolution of this normalized difference index across time and found that the difference increased over days, indicating that familiarity suppression occurred in all sessions conducted after the monkey had viewed each image at least 60 times (∼2 days). Fitting an exponential function to the data, the best fit model had an asymptote of 0.05 with a rate constant of 3.95 days (Figure 3.2B). We repeated the same analysis by comparing the familiarity difference index across views rather than days and found similar results (exponential fit, asymptote = 0.06, rate constant = 135 views).

In addition to presenting novel and intermediately familiar images every day, we also presented highly familiar images (viewed ∼1000s of times prior to recording). As a control, we recomputed the normalized familiarity difference index for highly familiar images, instead of intermediately familiar images. We found that the difference between novel and highly familiar images remained constant across days (or views) and was best estimated by a linear function with no slope, but a significant, positive intercept (linear model, intercept=0.065, p-value=0.05), indicating that the asymptotic difference between firing rates for novel and highly familiar images remained constant (F test vs. constant model, p-value = 0.29, F = 1.34). This means that once images are highly familiar, further exposure does not produce familiarity suppression, suggestive of a ceiling effect on familiarity dependent learning.

We modeled the time course of familiarity learning in more detail using a generalized linear model that accounted for response heterogeneity due to neurons and images. We regressed trial-averaged spike counts of single neurons for novel and intermediately familiar images with the following predictors - day (days 1-8), image category (novel or intermediately familiar), neuron number and an interaction between day and image category. The day regressor was a categorical variable with 8 levels and signified the number of days the monkeys had viewed each intermediately familiar image before the session. Image category and neuron numbers were also categorical variables that indicated whether spike counts belonged to a novel or intermediately familiar image and a scalar number that corresponded to each neuron's identity, respectively. The interaction between day and image category was significant from day 3 onwards (interaction of day3 x NF, estimate = -0.07, p-value=0.03) and the weights of the interaction terms increased over days, before saturating to a constant value, confirming that familiarity learning is slow as it takes multiple days for IT activity to reflect image familiarity.

3.3.3 Impact of short-term familiarity on neuronal encoding

Our analyses so far have focused on the effects of familiarity on intermediately familiar images that were presented hundreds of times over multiple recording sessions, spanning many days. To investigate the effects of familiarity within a single recording session, we presented monkeys with a unique set of ~40 novel stimuli every day that were repeatedly presented 30-40 times during each session. Along with the novel stimuli, we also presented ~40 highly familiar stimuli as a control image set. The evolution of the average visual responses (averaged over all stimulus-selective neurons, all stimuli, and time) across multiple

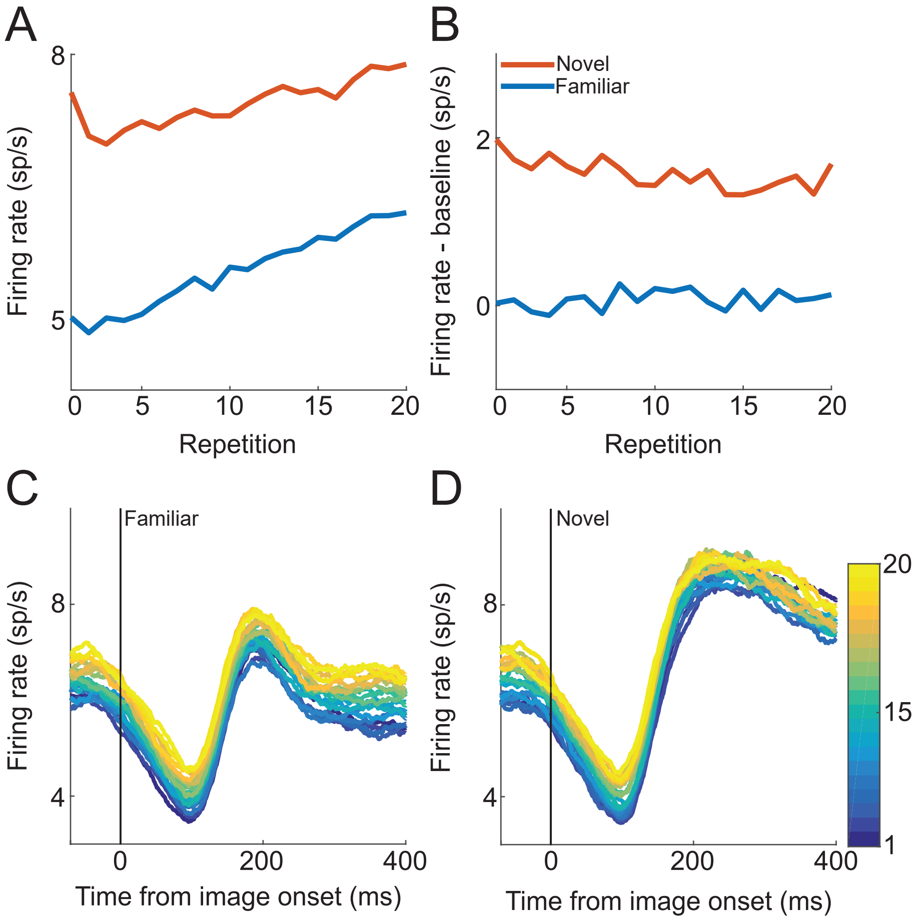

Figure 3.3: Impact of short-term familiarity on neuronal encoding.
(A) The average activity to novel and familiar stimuli during image presentation period (80 to 480 ms) across repetitions within a session for all visually selective units that were present from the start of the session. Firing rates increased to both novel and familiar stimuli. (B) Same as A but after subtracting baseline firing rates (-70 to 0 ms) for each image. On subtracting baseline firing rates, firing rates remained unchanged for both novel and familiar images. Note that scales are different for (A) and (B). (C) Time course of averaged neuronal responses to familiar stimuli colored by repetition number. Firing rates increased with repetition. Firing rates were smoothed using a 20-ms causal boxcar kernel. (D) Same as C, but for novel stimuli. Color bar represents repetition within a session.

repetitions showed that the responses increased with repeated presentations of both initially novel and highly familiar stimuli (Figure 3.3A). The time course of the average responses to novel and highly familiar stimuli also showed an increase in firing rates with repetition (Figure 3.3C and 3.3D). Notably, the baseline firing rates, prior to stimulus onset, also increased with repetition within a session for both novel and highly familiar stimuli. On subtracting baseline firing rates in a time window of -70 to 0 ms prior to stimulus onset for all stimuli, the average responses to initially novel stimuli showed a decreasing trend over repeated presentations, while the responses to highly familiar stimuli did not change significantly within a session (Figure 3.3B), suggesting that average responses to novel stimuli changed modestly at short time scales.

We wondered whether familiarity with stimuli results in increased selectivity within a single recording session. Neurons in IT respond selectively to complex images and tend to be driven by a small set of stimuli. To determine whether selectivity increased within a session, we compared maximum firing rates to novel and familiar stimuli instead of mean firing rates. Specifically, for each neuron, instead of averaging responses over all stimuli, we determined one novel image and one familiar image that elicited the highest firing rates. Maximum firing rates (averaged over all stimulus-selective neuron and time) did not change significantly with repetitions of novel and familiar stimuli (Figure 3.4A). Similarly, the time course of maximum firing rates also showed that they did not significantly change with repetitions of novel or familiar stimuli (Figure 3.4C and 3.4D). The time courses revealed that the baseline firing rates, prior to stimulus onset, also increased slightly with repetition within a session for both novel and highly familiar stimuli. On subtracting baseline firing rates in a time window of -70 to 0 ms prior to stimulus onset for all stimuli, the maximum responses to both novel and familiar stimuli did not change significantly over repeated presentations (Figure 3.4B), indicating that, at short time scales of minutes to hours, selectivity remained unchanged with increasing familiarity.

To gain insight into the single neurons that composed the neuronal population, we further

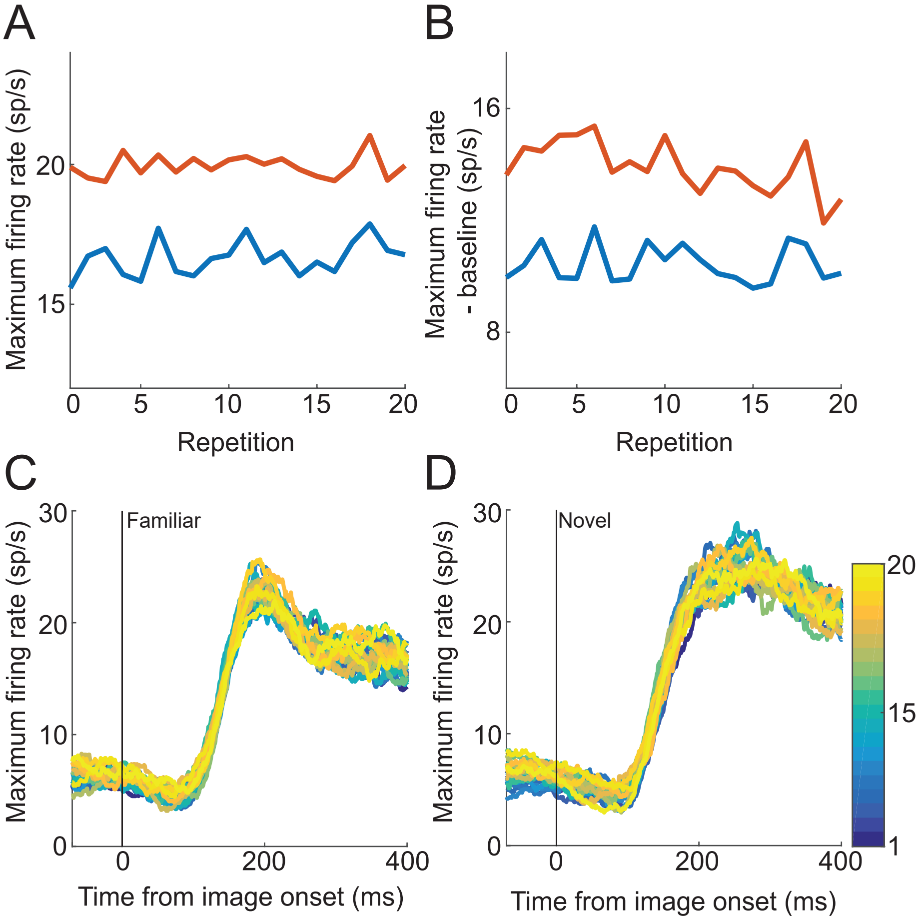

Figure 3.4: Impact of short-term familiarity on neuronal selectivity.
(A) The maximum activity to novel and familiar stimuli during image presentation period (80 to 480 ms) across repetitions within a session for all visually selective units that were present from the start of the session. Maximum firing rates remained unchanged to both novel and familiar stimuli. (B) Same as A but after subtracting baseline firing rates (-70 to 0 ms) for each image. On subtracting baseline firing rates, maximum firing rates showed a decreasing trend for novel objects, whereas they stay constant for familiar images. Note that scales are different for (A) and (B). (C) Time course of maximum neuronal responses to familiar stimuli colored by repetition number. Maximum firing rates did not change with repetition. Firing rates were smoothed using a 20-ms causal boxcar kernel. (D) Same as C, but for novel stimuli. Color bar represents repetition within a session.

analyzed learning effects on firing rates of individual neurons within a session. To quantify firing rate changes, for each neuron, we tested whether the evolution of firing rates with repetition of novel and familiar stimuli could be explained best by a linear model with zero slope, positive slope or negative slope. A neuron with zero slope denoted no change in firing rates with repetition, whereas a neuron with positive or negative slope denoted increasing or decreasing firing rates with repetition, respectively. Our analysis revealed that most neurons (70% (182/258) for novel, 78% (202/258) for familiar) did not change their responses significantly over 30 repetitions (2 hours in session length). 20% (50/258) of neurons showed a gradual decrease in firing rates with repetition of initially novel stimuli, while 11% (28/258) of neurons showed a gradual decrease in firing rates with repetitions of highly familiar stimuli. In comparison, 10% (25/258) of neurons showed increasing firing rates with repetition of initially novel stimuli, while 11% (28/258) of neurons showed decreasing firing rates with repetitions of highly familiar stimuli. Thus, a significant minority of neurons either showed a uniform decrease, or a uniform increase, from the beginning till the end of the session.

3.3.4 Familiarity-independent modulation of neuronal representations

Although the impact of familiarity within a session was modest, neuronal firing rates were modulated by other aspects unrelated to familiarity. First, we observed that baseline firing rates increased from the start to the end of a session. We calculated baseline firing rates for each image (-70 to 0 ms before image onset) and averaged over neurons and images and found that the average showed an increasing trend for all images independent of familiarity (Figure 3.5A). This was also evident in the individual neurons as we found 22% (57/258) of neurons showed an increase in baseline firing rate across time within a session. We also calculated baseline firing rates for the most preferred novel and familiar images for each neuron and averaged over all neurons and confirmed that maximum baseline firing rates for both novel and familiar images also increased (Figure 3.5B). Because baseline firing rates computed for each image could be contaminated by off-responses of the previous image presented in the RSVP sequence, we measured baseline firing rates for each trial (-500 to 0 ms before onset of first image in sequence), instead of all images in a trial. As the session progressed, we found that baseline firing rates computed for each trial, also showed an increase over time (Figure 3.5C). An increase in the baseline firing rates could result from neurons with poor isolation quality. On computing baseline firing rates for cells with very good isolation quality, we confirmed that baseline firing rates increased even for well-isolated cells with high signal-to-noise ratio.

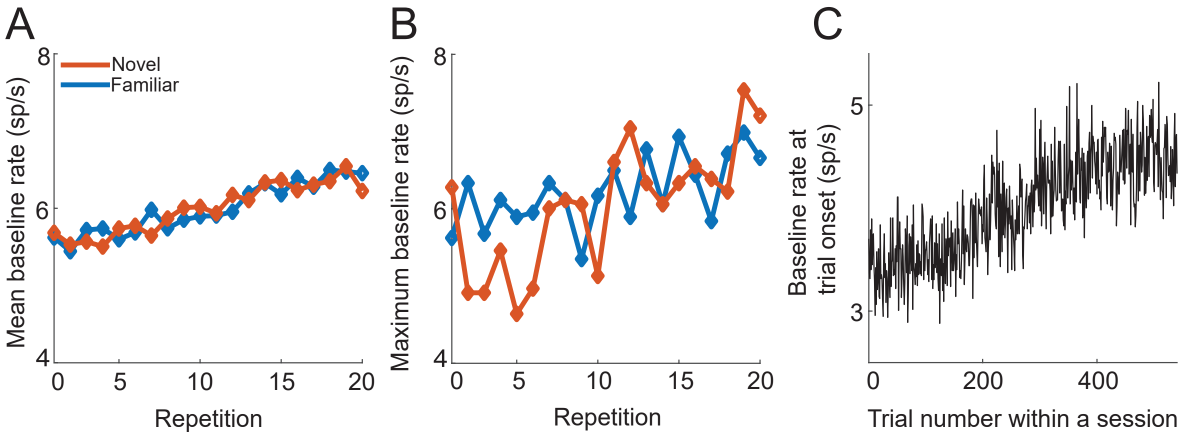

Figure 3.5: Familiarity-independent effects on baseline firing rates.
(A) Average baseline firing activity to novel and familiar stimuli, prior to stimulus onset, (-70 to 0 ms before stimulus onset) across repetitions within a session for all visually selective units that were present from the start of the session. Firing rates were averaged over all images and all neurons. Baseline firing rates increased for both novel and familiar stimuli. (B) Maximum baseline firing activity to the best novel and the best familiar stimulus across repetitions within a session for all visually selective units. Firing rates were averaged over the best image and all neurons. Maximum baseline firing rates also increased for both novel and familiar stimuli. (C) Average baseline firing rate calculated for each trial before stimulus onset across ∼500 trials within a session. 500 trials roughly corresponds to 20 repetitions of all images. Trial-wise baseline firing rates also increased within a session.

Second, we observed familiarity-independent modulation of firing rates over the course of a trial. The average time course of firing rates during image presentation featured a prominent dip or suppression in firing rates around image onset (Figure 3.6A). This dip in firing rates reached a minimum at 75 ms before rising to reflect visually-evoked responses to the image currently being viewed. Previous IT recordings have reported adaptation-like

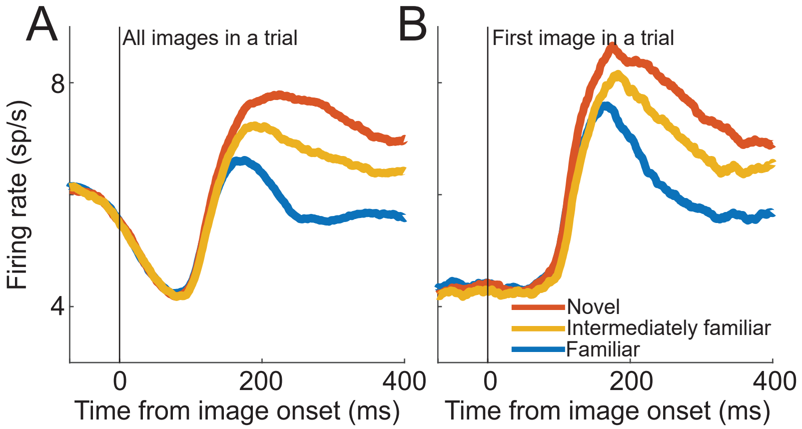

Figure 3.6: Familiarity-independent effects on firing rate dynamics within a trial. (A) Average PSTH of neurons to novel, intermediately familiar, and familiar stimuli. Activity is averaged over all images presented in a trial and over all visually selective neurons. The PSTH reveals suppression around image onset, before the visually-evoked response. Firing rates were smoothed using a 20-ms causal boxcar kernel. (B) Average PSTH of neurons to the first image in each trial, separately for novel, familiar, and intermediately familiar stimuli, averaged over all visually selective neurons. The PSTH reveals no suppression around image onset, suggesting that rapid image exposure leads to adaptation-like response around image onset. Firing rates were smoothed using a 20-ms causal boxcar kernel. Results remained the same even when no smoothing was used.

reduction in response that occurs with stimulus repetition known as repetition suppression. However, repetition suppression has been associated with a reduction in the magnitude of the visually-evoked response without affecting firing rates at image onset (McMahon and Olson, 2007). Further, repetition suppression has mostly been observed and studied during passive viewing of images. In comparison, the dimming detection task differs from previous studies in that multiple and variable (i.e. unpredictable) numbers of images are presented on every trial. We tested whether the rapid presentation of up to 5 images per trial resulted in suppression over the time course of a single trial. In all our analyses so far, we averaged over all stimuli regardless of whether they were presented in the beginning, middle or the end of the trial. On plotting the time course for only the first visual stimulus on each trial (averaged across neurons and all stimuli) and separating them by familiarity status, we observed no suppression of firing rates before or after stimulus onset (Figure 3.6B).

3.4 Discussion

We examined how neuronal responses change as initially novel objects become familiar over both long- and short-term learning, by recording from small populations of neurons while monkeys were presented with novel, intermediately familiar, and highly familiar stimuli. Intermediately familiar stimuli were presented multiple times within a session and over multiple days or recording. Unique novel stimuli, on the other hand, were presented multiple times within a session. We observed that neural responses to intermediately familiar stimuli decreased over multiple days with little change in the magnitude of firing rates within a single recording session. The trajectory of familiarity suppression could be estimated by an exponential function with a rate constant of 4 days ($\sim$ 135 views). On the other hand, within-session changes in neuronal encoding of familiarity were relatively small, since most neurons did not change their firing rates significantly. These results suggest that different learning mechanisms operate at different time scales during familiarity learning in IT cortex.

Previous work has described neurons responding to repeated images by either an average decrease in activity (Li et al., 1993; Xiang and Brown, 1998), or a sharpening of tuning (Baker et al., 2002; Erickson et al., 2000; Freedman et al., 2006; Sigala and Logothetis, 2002) over long-term learning. The sharpened tuning is believed to be achieved by an increase in activity to the preferred stimulus and suppression of activity to all other stimuli. This hypothesis takes the form of a winner-take-all network in which populations of neurons that are activated by stimuli show increased firing rates, whereas inactive or less active pools of neurons show decreased firing rates. However, in our data, although mean firing rates decrease, maximum firing rates do not increase over learning across multiple days or within a day. A previous study observed increases in maximum firing rate only in putative excitatory neurons (Woloszyn and Sheinberg, 2012). However, the proportions of putative excitatory and inhibitory neurons are low, especially when we group our neurons by day. Future studies with a larger set of simultaneously recorded neurons, perhaps chronic recordings, would be able to test whether maximum firing rates increase during familiarity learning and result in increased selectivity for familiar stimuli.

We observed a marginal impact of familiarity learning within a single recording session. Consistent with this, a previous study of familiarity learning in PFC also showed limited changes in firing rates of novel stimuli during associative learning. While the activity of single neurons differentiated between novel and familiar stimuli, novel stimuli did not show a dramatic change in firing rates between early vs. late trials even after 1500 trials or so. Trivially, these results might suggest that familiarity learning itself might not occur in IT or PFC, instead they inherit learned signals from another brain region. The most likely candidates for familiarity learning might be perirhinal cortex or the hippocampus. Both areas are influenced by familiarity (Xiang and Brown, 1998). A study in perirhinal cortex found that after one day of experience with novel stimuli, response preference of nearby neurons responded to similar stimuli (Erickson et al., 2000). This suggests that, in the initial stages, effects of learning might be present at the population level (correlations between neurons or higher order structure). It is worth noting that response changes in perirhinal cortex also occurred after one day of experience, and not within a single session. Further, sleep might play an important role in the consolidation of learned memories before effects of learning are evident in single neuron responses. Finally, novelty is a salient visual property that captures our attention and eye movements rapidly. It is possible that novel/familiar learning and recognition is also mediated by dorsal stream structures that are known to be involved in eye movement planning, like LIP and the frontal eye fields in PFC. Most likely, novel/familiar learning and recognition would involve coordination between the ventral and dorsal streams to guide behavior.

3.5 Methods

3.5.1 Behavioral task and stimulus display

Two male monkeys (Macaca mulatta) were trained to perform the RSVP dimming detection task. In this task, monkeys were rewarded for releasing a touch bar when the brightness/luminance of a stimulus decreased. On each trial, a 500 ms fixation period was followed by a variable sequence of images (1-6 images). Each image was presented for 400 ms with an inter-stimulus interval of 150 ms. The number of images on each trial ranged from 1-6. On 50% of the trials, image dimming occurred on any of the first 5 images and the monkeys were rewarded for releasing a touch bar when dimming occurred. On the remaining 50% of the trials, 6 images were presented and no dimming occurred and the monkeys were rewarded for continuing to hold the lever. The variable number of images per trial ensured that monkeys could not use task timings to predict the time at which touch bar release was required.

The stimuli used in this study were 150 x 150-pixel full color photographs downloaded from Flickr using the program Bulkr. The images consisted of natural and man-made objects and scenes and excluded face stimuli. For each recording session, 40 novel stimuli, 40 intermediately familiar stimuli and 40 highly familiar stimuli were selected at random. Highly familiar stimuli were sampled from a large library of images which were familiarized through the dimming detection task for 2-3 months before recording. Highly familiar images were presented 1000-2000 times before the start of recording. Intermediately familiar images were presented 30-800 times during recording after which they were replaced by a new set of intermediately familiar images.

Gaze positions were measured using an Eyelink 1000 optical eye tracker (SR Research) at a sampling rate of 1 KHz and stored for offline analysis. Task events, stimulus display, timings and reward delivery were controlled via a MATLAB-based toolbox, MonkeyLogic. Stimuli were displayed on a 21-inch color CRT monitor (1280x1024 resolution, 75 Hz refresh rate, 57 cm viewing distance).

3.5.2 Electrophysiological recoding

Two male monkeys were implanted with a headpost and a recording chamber. The recording chamber was implanted over the inferotemporal cortex centered 16 mm posterior to the interaural line. Stereotaxic coordinates were determined from anatomical MRI scans obtained prior to headpost and chamber implantation. All surgical procedures were in accordance with the University of Chicagos Animal Care and Use Committee and US National Institutes of Health guidelines.

About 1/3rd of IT recordings were conducted using 120-m tungsten microelectrodes (FHC), a dura-piercing 25 Ga guide tube and an electronic micromanipulator (NAN Instruments). The rest of the recordings were conducted using V-Probes (24-channel multielectrode linear array, Plexon). Neurophysiological signals were amplified, digitized, and stored for offline spike sorting (Plexon) to verify the quality and stability of neuronal isolation within a recording session.

3.5.3 Data Analysis

All our analyses were conducted on correct trials, excluding trials with incorrect responses, fixation breaks, and early responses. Unless otherwise stated, all our results were qualitatively and quantitatively similar in both monkeys. Thus, we combined data sets from both monkeys for all analyses.

CHAPTER 4

DISCUSSION AND CONCLUSIONS

4.1 Results Summary

Humans and animals excel in their ability to flexibly adapt and respond to a multitude of visual stimuli in their environment. This flexibility in behavior relies on learning-dependent changes in neural circuits. Some visual behaviors are learned through feedback over long periods of training, some others are learned rapidly without any supervision. We investigated the neural activity and dynamics underlying both supervised and unsupervised learning.

In the first experiment, we examined the task-dependence of neuronal representations during supervised visual categorization tasks. We compared neuronal activity and dynamics in LIP during categorization task switching between OIC and DMC tasks. The two tasks involved the same visual stimuli and same category rules, but they differed both in the specific behavioral demands and the effectors used to report the decisions. This experiment revealed that category representations are similar in both tasks with higher category selectivity in the delayed matching task. Further, neuronal populations in the DMC task used a stable and persistent encoding scheme for categories. In contrast, neuronal populations in the OIC task, during the same time period, used a dynamic and transient encoding scheme. This suggests that LIP mediates context-dependent categorization by dynamically transforming encoding of category information in a task-dependent manner.

In the second experiment, we examined unsupervised visual familiarity learning in the inferotemporal cortex at multiple time scales. We compared neuronal activity and dynamics over long-term and short-term familiarity as monkeys performed tasks that involved repeated presentations of initially novel stimuli. Over long time scales of days to weeks, we found that mean population activity decreased as visual stimuli became familiar. Long-term familiarity had a much stronger impact on neuronal encoding in ITC than short-term familiarity. Over short time scales of minutes to hours, we found a modest impact of familiarity on neuronal encoding. Our results suggest that familiarity learning operates at multiple time scales in IT, with different learning mechanisms over long- and short-time scales.

4.2 Questions for future research

4.2.1 Dorsal-ventral integration

The experiments reported here investigate the dorsal and ventral streams in isolation. However, almost all everyday behaviors such as the ability to recognize an object's identity (”what”) along with its position (”where”) likely depends on integrating neuronal encoding of form and spatial features across the two streams. To understand this, we examined how visual object recognition (ventral stream function) drives spatial attention and eye movements (dorsal stream function).

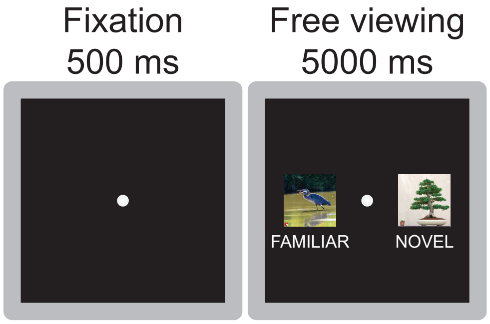

Figure 4.1: Preferential looking task.
Monkeys are shown two stimuli (one novel and one familiar), and are free to make eye movements to either or both stimuli while their eye movements are tracked.

As previously described, novel visual stimuli automatically attract our attention and eye movements. Humans and monkeys have an innate looking bias towards novel stimuli that requires no training and we can take advantage of this to investigate the coordination between dorsal and ventral visual streams. To this end, we designed an ethological preferential looking task as a naturalistic way to probe novelty. In this task, monkeys are presented with two stimuli - one novel and one familiar - and are allowed to freely scan both images while their eye movements are recorded (Figure 4.1) . As expected, monkeys preferentially viewed novel images when it was paired with a familiar image. Specifically, 80% of monkeys' first saccades were made to the novel image and they spent more time looking at the novel than the familiar image. Interestingly, the preferential looking task revealed that monkeys' viewing patterns towards initially novel stimuli saturated after 50 views within a single session. Thus, monkeys treated images viewed ∼50 times on par with images viewed 1000s of times. In order for visual familiarity information to drive visually guided eye movements, we hypothesize that ventral areas such as IT must interact with dorsal areas such as PPC and FEF (Figure 4.2). Future studies of the preferential looking task in both visual streams will illuminate the mechanisms underlying coordination and integration between dorsal and ventral visual streams.

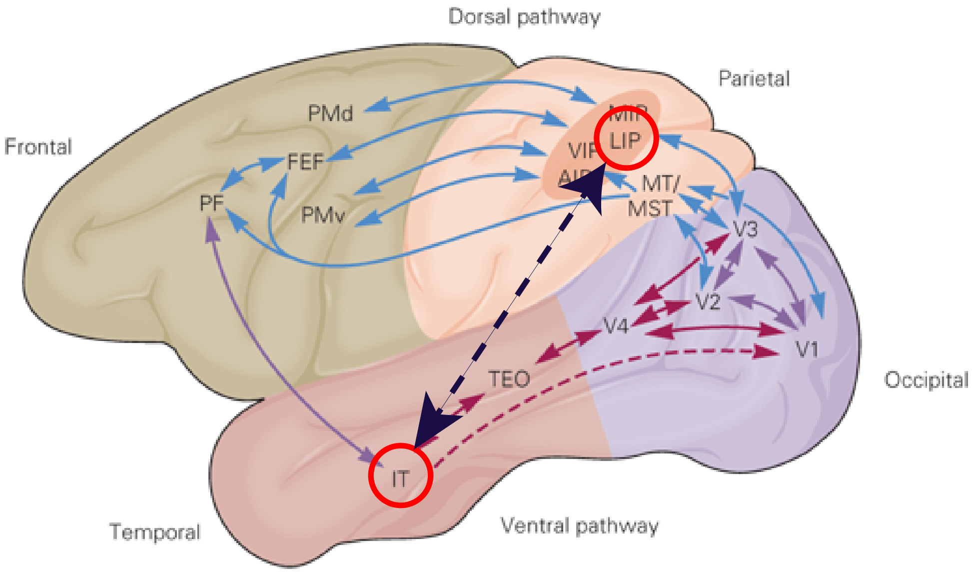

Figure 4.2: Two-streams hypothesis.
Lateral view of a macaque brain highlighting the dorsal and ventral visual pathways. Red circles represent the cortical areas relevant for the studies presented here. The dashed arrow represents potential interactions between dorsal LIP and ventral IT to mediate visually-guided behavior. Adapted from Kandel ER, et al., Principles of Neural Science. 5th ed

REFERENCES

B. Anderson, R. E. Mruczek, K. Kawasaki, and D. Sheinberg. Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral Cortex*, 18(11):2540–2552, 2008. ISSN 10473211. doi: 10.1093/cercor/bhn015.

E. G. Antzoulatos and E. K. Miller. Differences between Neural Activity in Prefrontal Cortex and Striatum during Learning of Novel Abstract Categories. *Neuron*, 71(2):243–249, 2011. ISSN 08966273. doi: 10.1016/j.neuron.2011.05.040.

W. F. Asaad, G. Rainer, and E. K. Miller. Neural activity in the primate prefrontal cortex during associative learning. Technical Report 6, 1998.

C. I. Baker, M. Behrmann, and C. R. Olson. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, 5(11):1210–1216, nov 2002. ISSN 10976256. doi: 10.1038/nn960.

S. Belkacem, H. Belbache, C. Boubekri, P. Mosset, O. Rached-Mosbah, E. Marchioni, S. Benayache, and F. Benayache. Chemical Constituents from Centaurea parviflora Desf. Technical Report 3, 2014.

S. Bennur and J. I. Gold. Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. *Journal of Neuroscience*, 31(3): 913–921, 2011. ISSN 02706474. doi: 10.1523/JNEUROSCI.4417-10.2011.

J. W. Bisley and M. E. Goldberg. Attention, Intention, and Priority in the Parietal Lobe. *Annual Review of Neuroscience*, (1):1–21. ISSN 0147-006X. doi: 10.1146/annurev-neuro-060909-152823.

S. L. Brincat and C. E. Connor. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, 7(8):880–886, 2004. ISSN 10976256. doi: 10.1038/nn1278.

M. V. Chafee and P. S. Goldman-Rakic. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. Technical Report 6, 1998.

W. Chaisangmongkon, S. K. Swaminathan, D. J. Freedman, and X. J. Wang. Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. *Neuron*, (6):1504–1517.e4, mar . ISSN 10974199. doi: 10.1016/j.neuron.2017.03.002.

C. L. Colby, J. R. Duhamel, and M. E. Goldberg. Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. Technical Report 5.

V. de Lafuente, M. Jazayeri, and M. N. Shadlen. Representation of accumulating evidence for a decision in two parietal areas. *Journal of Neuroscience*, 35(10):4306–4318, 2015. ISSN 15292401. doi: 10.1523/JNEUROSCI.2451-14.2015.

J. J. DiCarlo and J. H. Maunsell. Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature Neuroscience*, 3(8):814–821, 2000. ISSN 10976256. doi: 10.1038/77722.

J. J. DiCarlo and J. H. Maunsell. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, 89(6):3264–3278, 2003. ISSN 00223077. doi: 10.1152/jn.00358.2002.

C. A. Erickson, B. Jagadeesh, and R. Desimone. Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. Technical Report 11, 2000.

F. L. Fahy, I. P. Riches, and M. W. Brown. Neuronal activity related to visual recognition memory: long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. Technical Report 3, 1993.

V. P. Ferrera, M. Yanike, and C. Cassanello. Frontal eye field neurons signal changes in decision criteria. *Nature Neuroscience*, 12(11):1458–1462, 2009. ISSN 10976256. doi: 10.1038/nn.2434.

J. K. Fitzgerald, D. J. Freedman, and J. A. Assad. Generalized associative representations in parietal cortex. *Nature Neuroscience*, 14(8):1075–1079, 2011. ISSN 10976256. doi: 10.1038/nn.2878.

D. J. Freedman and J. A. Assad. Experience-dependent representation of visual categories in parietal cortex. *Nature*, 443(7107):85–88, 2006. ISSN 14764687. doi: 10.1038/nature05078.

D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291(5502):312–316, jan 2001. ISSN 00368075. doi: 10.1126/science.291.5502.312.

D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. Rol Biarticular Power. Technical Report 12, 2003.

D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*, 16(11): 1631–1644, 2006. ISSN 10473211. doi: 10.1093/cercor/bhj100.

S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. Technical Report 2, 1989.

A. Ghazizadeh, W. Griggs, and O. Hikosaka. Ecological origins of object salience: Reward, uncertainty, aversiveness, and novelty, 2016. ISSN 1662453X.

J. W. Gnadt and R. A. Andersen. Memory related motor planning activity in posterior parietal cortex of macaque. Technical Report 1, 1988.

J. I. Gold and M. N. Shadlen. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. Technical Report 2, 2003.

M. Golub and D. Sussillo. FixedPointFinder: A Tensorflow toolbox for identifying and characterizing fixed points in recurrent neural networks. *Journal of Open Source Software*, 3(31):1003, 2018. ISSN 2475-9066. doi: 10.21105/joss.01003.

C. G. Gross, C. E. Rocha-Miranda, and D. B. Bender. Visual properties of neurons in inferotemporal cortex of the Macaque. Technical Report 1, 1972.

T. M. Herrington and J. A. Assad. Temporal sequence of attentional modulation in the lateral intraparietal area and middle temporal area during rapid covert shifts of attention. *Journal of Neuroscience*, (9):3287–3296. ISSN 02706474. doi: 10.1523/JNEUROSCI.6025-09.2010.

H. Hong, D. L. Yamins, N. J. Majaj, and J. J. Dicarlo. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, 19 (4):613–622, feb 2016. ISSN 15461726. doi: 10.1038/nn.4247.

G. Huang, S. Ramachandran, T. S. Lee, and C. R. Olson. Neural correlate of visual familiarity in macaque area V2. *Journal of Neuroscience*, 38(42):8967–8975, 2018. ISSN 15292401. doi: 10.1523/JNEUROSCI.0664-18.2018.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Technical Report 1, 1962.

D. H. HUBEL and T. N. WIESEL. Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. Technical report, 1965.

D. Ingle, G. Schneider, C. Trevarthen, and R. Held. Locating and identifying: Two modes of visual processing. *Psychologische Forschung*, 31(1):42–43, mar 1967. ISSN 1430-2772. doi: 10.1007/BF00422384.

M. J. Jutras and E. A. Buffalo. Recognition memory signals in the macaque hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 107(1): 401–406, 2010. ISSN 00278424. doi: 10.1073/pnas.0908378107.

L. N. Katz, J. L. Yates, J. W. Pillow, and A. C. Huk. Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature*, (7611):285–288. ISSN 14764687. doi: 10.1038/nature18617.

H. Kumano, Y. Suda, and T. Uka. Context-dependent accumulation of sensory evidence in the parietal cortex underlies flexible task switching. *Journal of Neuroscience*, 36(48): 12192–12202, 2016. ISSN 15292401. doi: 10.1523/JNEUROSCI.1693-16.2016.

L. Li, E. K. Miller, and R. Desimone. The representation of stimulus familiarity in anterior inferior temporal cortex. Technical Report 6, 1993.

N. Li and J. J. DiCarlo. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, 321(5895):1502–1507, sep 2008. ISSN 00368075. doi: 10.1126/science.1160028.

N. Li and J. J. DiCarlo. Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron*, 67(6):1062–1075, sep 2010. ISSN 08966273. doi: 10.1016/j.neuron.2010.08.029.

N. K. Logothetis, J. Pauls, and T. Poggio. Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5):552–563, may 1995. ISSN 09609822. doi: 10.1016/S0960-9822(95)00108-4.

V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, nov 2013. ISSN 00280836. doi: 10.1038/nature12742.

N. Y. Masse, J. M. Hodnefield, and D. J. Freedman. Mnemonic encoding and cortical organization in parietal and prefrontal cortices. *Journal of Neuroscience*, (25):6098–6112. ISSN 15292401. doi: 10.1523/JNEUROSCI.3903-16.2017.

N. Y. Masse, G. R. Yang, H. F. Song, X. J. Wang, and D. J. Freedman. Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nature Neuroscience*, 22(7):1159–1167, 2019. ISSN 15461726. doi: 10.1038/s41593-019-0414-3.

J. L. McKee, M. Riesenhuber, E. K. Miller, and D. J. Freedman. Task dependence of visual and category representations in prefrontal and inferior temporal cortices. *Journal of Neuroscience*, 34(48):16065–16075, 2014. ISSN 15292401. doi: 10.1523/JNEUROSCI.1660-14.2014.

D. B. T. McMahon and C. R. Olson. Repetition suppression in monkey inferotemporal cortex: Relation to behavioral priming. *Journal of Neurophysiology*, 97(5):3532–3543, 2007. ISSN 00223077. doi: 10.1152/jn.01042.2006.

H. Merchant, D. A. Crowe, M. S. Robertson, A. F. Fortes, and A. P. Georgopoulos. Top-down spatial categorization signal from prefrontal to posterior parietal cortex in the primate. *Frontiers in Systems Neuroscience*, (AUGUST 2011). ISSN 16625137. doi: 10.3389/fnsys.2011.00069.

T. Meyer and N. Rust. Single-exposure visual memory judgments are reflected in inferotemporal cortex. *eLife*, 7, mar 2018. ISSN 2050084X. doi: 10.7554/eLife.32259.

T. Meyer, C. Walker, R. Y. Cho, and C. R. Olson. Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nature Neuroscience*, 17(10):1388–1394, 2014. ISSN 15461726. doi: 10.1038/nn.3794.

E. M. Meyers, D. J. Freedman, G. Kreiman, E. K. Miller, and T. Poggio. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, 100(3):1407–1419, 2008. ISSN 00223077. doi: 10.1152/jn.90248.2008.

E. K. Miller and J. D. Cohen. An Integrative Theory of Prefrontal Cortex Function. Technical Report 1, 2001.

M. Mishkin and L. G. Ungerleider. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Technical Report 1, 1982.

A. Pasupathy and C. E. Connor. Population coding of shape in area V4. *Nature Neuroscience*, 5(12):1332–1338, 2002. ISSN 10976256. doi: 10.1038/972.

G. Rainer and E. K. Miller. Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, 27(1):179–189, 2000. ISSN 08966273. doi: 10.1016/S0896-6273(00)00019-2.

G. Rainer, H. Lee, and N. K. Logothetis. The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biology*, 2(2):e44, feb 2004. ISSN 15449173. doi: 10.1371/journal.pbio.0020044.

C. A. Rishel, G. Huang, and D. J. Freedman. Independent category and spatial encoding in parietal cortex. *Neuron*, 77(5):969–979, mar 2013. ISSN 10974199. doi: 10.1016/j.neuron.2013.01.007.

J. E. Roy, M. Riesenhuber, T. Poggio, and E. K. Miller. Prefrontal cortex activity during flexible categorization. *Journal of Neuroscience*, (25):8519–8528. ISSN 02706474. doi: 10.1523/JNEUROSCI.4837-09.2010.

K. Sakai and Y. Miyashita. Neural organization for the long-term memory of paired associates. *Nature*, 354(6349):152–155, 1991. ISSN 00280836. doi: 10.1038/354152a0.

M. Sarafyazd and M. Jazayeri. Hierarchical reasoning by neural circuits in the frontal cortex. *Science*, 364(6441), 2019. ISSN 10959203. doi: 10.1126/science.aav8911.

A. Sarma, N. Y. Masse, X. J. Wang, and D. J. Freedman. Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nature Neuroscience*, 19(1): 143–149, jan 2015. ISSN 15461726. doi: 10.1038/nn.4168.

A. B. Sereno and J. H. R. Maunsell. Shape selectivity in primate lateral intraparietal cortex. *Nature*, 395(6701):500–503, 1998. ISSN 00280836. doi: 10.1038/26752.

M. Siegel, T. J. Buschman, and E. K. Miller. Cortical information flow during flexible sensorimotor decisions. Technical Report 6241, 2015.

N. Sigala and N. K. Logothetis. Visual categorization shapes feature selectivity in the primate temporal cortex. Technical Report 6869, 2002.

L. H. Snyder, A. P. Batista, and R. A. Andersen. Intention-related activity in the posterior parietal cortex: A review. *Vision Research*, (10-12):1433–1441, jun . ISSN 00426989. doi: 10.1016/S0042-6989(00)00052-3.

L. H. Snyder, A. P. Batista, and R. A. Andersen. Coding of intention in the posterior parietal cortex. *Nature*, 386(6621):167–170, 1997. ISSN 00280836. doi: 10.1038/386167a0.

G. Stoet and L. H. Snyder. Single neurons in posterior parietal cortex of monkeys encode cognitive set. Technical Report 6, 2004.

S. K. Swaminathan and D. J. Freedman. Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nature Neuroscience*, 15(2):315–320, 2012. ISSN 10976256. doi: 10.1038/nn.3016.

S. K. Swaminathan, N. Y. Masse, and D. J. Freedman. A comparison of lateral and medial intraparietal areas during a visual categorization task. *Journal of Neuroscience*, 33(32): 13157–13170, 2013. ISSN 02706474. doi: 10.1523/JNEUROSCI.5723-12.2013.

K. Tanaka. Inferotemporal Cortex and Object Vision. Technical Report 1, 1996.

J. D. Wallis, K. C. Anderson, and E. K. Miller. Single neurons in prefrontal cortex encode abstract roles. *Nature*, 411(6840):953–956, 2001. ISSN 00280836. doi: 10.1038/35082081.

L. Woloszyn and D. L. Sheinberg. Effects of Long-Term Visual Experience on Responses of Distinct Classes of Single Units in Inferior Temporal Cortex. *Neuron*, 74(1):193–205, 2012. ISSN 08966273. doi: 10.1016/j.neuron.2012.01.032.

J. Z. Xiang and M. W. Brown. Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Technical Report 4-5, 1998.

Y. Zhou and D. J. Freedman. Posterior parietal cortex plays a causal role in perceptual and categorical decisions. Technical Report 6449, 2019.