BEHAVIOURAL AND NEUROPHYSIOLOGICAL EXPRESSION OF VISUAL CATEGORIZATION IN THE MOUSE MEDIAL PREFRONTAL CORTEX

RUI CARLOS SALCEDAS PAIS
Master of Science, **University of Salamanca, 2014**

A thesis submitted
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

NEUROSCIENCE

Department of Neuroscience
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

BEHAVIOURAL AND NEUROPHYSIOLOGICAL EXPRESSION OF VISUAL CATEGORIZATION PROCESSES IN THE MOUSE MEDIAL PREFRONTAL CORTEX

RUI CARLOS SALCEDAS PAIS

Date of Defense: June 16, 2021

Dr. B. McNaughton Supervisor	Professor	Ph.D.
Dr. R. Sutherland Thesis Examination Committee Member	Professor	Ph.D.
Dr. M. Mohajerani Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. I. Whishaw Internal-External Examiner	Professor	Ph.D.
Dr. B. Winters External Examiner University of Guelph Guelph, Ontario	Associate Professor	Ph.D.
Dr. A. Iwaniuk Chair, Thesis Examination Committee	Associate Professor	Ph.D.

DEDICATION

For you Mom.

ABSTRACT

Categorization is a process that allows organisms to classify and respond to individual elements of any given experience based on their shared similarities, and to generalize their behavioural output to novel stimuli that share the same features.

In this set of experiments we explored the underlying neural dynamics of visual categorization in mice by developing a virtual reality task using an automated touchscreen operant conditioning box. By gradually incrementing the number of exemplars available in a pairwise object recognition task, mice learned to discriminate between virtual objects that belong to 2 main categories. We also tried to determine if the neural activity in the mouse prefrontal cortex, a region which has been associated with category representations in primates, reflected the acquisition of this new information. In order to do this we used in-vivo 2-photon calcium imaging to record the neural activity of a cohort of mice at different time points.

ACKNOWLEDGEMENTS

First and most importantly, I would like to thank my parents and in particular my mother, to whom I dedicate this thesis. I wish she would've been able to witness the end of my journey here, and I'm sure she would've been proud, just like she was the day I started it. To my father, who despite all the hardships he went through was always a source of unwavering support and wisdom, and always encouraged me to keep pushing forward no matter how difficult things were. To my friends, who although scattered throughout the world have always made this journey worthy, especially Vasco - the embodiment of friendship - who was always there during the toughest times.

I also want to show my gratitude to my supervisor Dr. Bruce McNaughton, for giving me the opportunity to work and grow as a scientist under his supervision and to take on such an ambitious project.

None of this would have been possible without my past and present lab and office mates, which have been an inexorable source of encouragement, and helped me to grow as a researcher and as person. Among these I want to give special thanks to Scott Deibel, Sean Lacoursier, and my good friend Darryl Gidyk who recently passed away. They were the best office mates one could ask for, and with whom I had so many insightful conversations.

I also want to thank the amazing McNaughton lab calcium imaging team: Aubrey Demchuk, Adam Neumann, Ingrid Esteves, and of course - HaoRan Chang, to whom I owe a great debt for all the help with the data analysis and who became one of my closest friends. These people were the best source of procrastination and scientific advice I had throughout this journey.

I also want to thank JianJun Sun for conducting the surgeries, and Leonardo Molina, who developed the virtual reality setup and the operant conditioning boxes used in the experiments described in this thesis, and without whom this project would simply not be possible.

Lastly, I want to thank the current and former members of my thesis committee, Dr. Robert Sutherland, Dr. Majid Mohajerani, Dr. Artur Luczak and Dr. Boyer Winters from the University of Guelph, for all the insightful discussions, guidance and scientific advice. Additionally, I would also like to thank Dr. Ian Whishaw and Dr. Andrew Iwaniuk for accepting to be the internal-external examiner and the examination chair respectively.

After five and half long years, it seems that this journey has finally come to an end. None of this would've been possible without all the support I've received, all of which allowed me to keep pushing this boulder uphill, even if only to see it roll downhill again. But in the end, one must (always) imagine Sisyphus happy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Anterior Cingulate Cortex
AG	Amygdala
ALCOVE	Attention Learning Covering Map
ANOVA	Analysis of variance
ATRIUM	Attention to Rules and Instances in a Unified Model
BG	Basal Ganglia
CA1	Cornu Amonis Area 1
CDF	Cumulative Distribution Function
CLST	Complementary Learning Systems Theory
Ctrl	Control Category
DIVA	Divergent Autoencoder Model
DLPFC	Dorsolateral Prefrontal Cortex
DMS	Delayed Matched to Sample
DMC	Delayed Match to Category
fMRi	Functional Magnetic Resonance Imaging
FOV	Field of View
GCM	Generalized Context Model
HPC	Hippocampus
ISI	Inter-Stimulus Interval
ITC	Inferior Temporal Cortex
LEC	Lateral Entorhinal Cortex
LOO-CV	Leave-One-Out Cross Validation
mPFC	medial Prefrontal Cortex
MTL	Medial Temporal Lobe Region
OCR	Object Category Recognition
OFC	Orbitofrontal Cortex
OR	Object Recognition
PFC	Prefrontal Cortex
PRh	Perirhinal Cortex
PrL	Prelimbic
PSTH	Peri-Stimulus Time Histogram
ROI	Region of interest
RULEX	Rule-Plus-Exception Model
STR	Striatum
SUSTAIN	Supervised and Unsupervised Stratified Adaptive Incremental Network
SWS	Slow Wave Sleep
S+	Positively Reinforced Conditioned Stimulus
S-	Negatively Reinforced Conditioned Stimulus
TE	Anterior Temporal Cortex
TEO	Posterior Temporal Cortex
UV	Ultraviolet light
V1	Visual Area V1
V2	Visual Area V2
V3	Visual area V3

V4	Visual area V4
VLPFC	Ventrolateral Prefrontal Cortex
VR	Virtual Reality
VTA	Ventral Tegmental Area

Chapter 1

ABSTRACT

In order to create a rich internal model of the world, and make predictions about the properties of objects or the outcomes of specific events, the brain needs to be able to encode the statistical regularities in the environment. The information can then be stored in long term memory in order to create a generalized body of knowledge that captures the categorical structure of the external world. In this chapter I will elaborate on how the brain might organize the information stored in long term memory through the process of categorization, and focus on the theoretical and physiological accounts that have provided valuable insights on this topic, particularly in the case of visual categorization. Lastly, I will outline the two main hypothesis that motivated the set of experiments described in the following chapters.

General Introduction

1. Semantic Memory and Generalization

Biological and artificial systems rely on the ability to store information on a relatively permanent basis in order to learn. This information then needs to be structured in a way which allows for the extraction of patterns that a biological or artificial agent can correlate with past experiences, in order to orient its behaviour or output, and create predictions about the world. In other words, what initially starts as a detailed record of specific events and their spatial and temporal contexts (often referred to as episodic memory), needs to be restructured, in order to create an adaptive internal model of generalized knowledge, also known (at least in the human literature) as semantic memory (Marr, 1970; McNaughton, 2010; Tulving, 1972).

Early endeavours in semantic memory modeling in the late 60's and early 70's, as well as the experimental work that followed, conceptualized the structure of semantic memory as a collection of interconnected and hierarchically organized mental structures, akin to the concept of "schema"[1] (Collins & Quillian, 1969, 1970; Quillian, 1966, 1967; Rumelhart & Norman, 1973; Rumelhart & Ortony, 1977). For example, in order to determine the truth of a sentence such as "a

[1] The term "schema" stems from the Greek word *schēmat* or *schēma*, which means "figure", and was initially used by Immanuel Kant in his "Critique of Pure Reason", before it became widespread in the field of cognitive psychology. In psychology, the term was initially introduced in the 1920's by French psychologist Jean Piaget and later popularized by Frederic Bartlett (Bartlett, 1932; Nevid, 2007; Piaget, 1923). The concept has mutated slightly over the years; while traditional psychology approaches define schemas as abstract knowledge structures that organize and categorize information in the human mind, more recent neurobiological approaches might refer to schemas as networks of interconnected neocortical representations, comprised of prior knowledge (Gilboa & Marlatte, 2017; Van Kesteren, Ruiter, Fernández, & Henson, 2012).

robin can fly", humans use their long term memory which contains representations about such concepts. According to Quillian (Quillian, 1966, 1967), these memories could be organized in two different ways. The first asserts that each bird that flies (e.g. robin, canary, eagle, etc.) is stored in long term memory along with the fact that it can fly; the second one implies that the inference about a specific type of bird that flies (e.g. a robin), is based on the generalization that birds can fly, and since a canary is a bird, the condition for it to be capable of flying is satisfied (Collins & Quillian, 1969).

This led Quillian to propose a hierarchical model for storing semantic information in a computer. In this model, a given word had stored with it an array of pointers with a specific configuration to other words in memory which, in turn, represented the words meaning. This model conceptualized knowledge as being structured in a hierarchy based on ordinary experience, whereby major concepts such as animals and plants could be divided into smaller subdivisions, such as birds and flowers, since these correspond to more specific concepts within the superordinate category. This type of semantic memory model was known by different names such as "semantic network model" or "connectionist model" and became widely used as a tool for investigating the structure of generalized knowledge. This was not only the case in human studies, which relied on reaction times when subjects were presented with a series of words and their attributes, but also in Artificial Intelligence (A.I) research (Collins & Quillian, 1969, 1970; Meyer, 1970; Quillian, 1967; Rips, Shoben, & Smith, 1973; Rumelhart & Norman, 1973; Rumelhart & Ortony, 1977; Schaeffer & Wallace, 1969; E. E. Smith, 1967).

A semantic memory model developed by Rumelhart and Todd (1993), was later used in another theoretical account made by McClelland, McNaughton and O'Reilly (1995), called "Complementary Learning Systems Theory" (CLST) (McClelland, McNaughton, & O'Reilly,

1995; Rumelhart & Todd, 1993). In this conceptual learning model, the network was given a set of inputs in the form of concept-relation pair propositions, such as “robin can”, in order to activate the nodes corresponding to the capabilities of a robin as an output. This network could learn these relationships by first propagating the activation forward to produce an output, then comparing these results with the desired ones, and finally adjusting each connection weight in the network in a gradual manner, in order to minimize error over several iterations. Initially (by construction) the concepts had distributed and almost indistinguishable representations, but after 200 epochs the difference between animals and plants was apparent, and after 500 iterations the network was capable of distinguishing among subsets of those categories such as birds and fish in the case of animals, or flowers and trees in the case of plants. The concept representation in the network assumed a sparse and hierarchical structure, as similar concepts formed clusters which were clearly identifiable. The key point here, is the gradual adjustment of the weights after each presentation of a specific concept, which allows the network to learn the structure of the domain incrementally (McClelland, 2013; McClelland & Goddard, 1996; McClelland et al., 1995).

It’s important to note, however, that, regardless of the similarities between the results obtained with this type of neural network architecture and the model initially proposed by Quillian, the fundamental processes that determine how information is represented are quite different. While Quillian’s model relies on explicit hierarchical links which denote the relationships between concepts and attributes by simply traversing between them, in the Rumelhart network used by McClelland et al., (1995) the hierarchical structure emerges in an implicit manner, through similarities among concepts, via pattern completion (McClelland, 2010). In other words, the general properties and abilities of each concept are derived from the associations between different patterns which represent each of those concepts. This creates not only an efficient way to retrieve

information, but also to learn new concepts, because it relies on category-general properties instead of concept-specific properties, which are based on minor aspects that distinguish them. And this, is one of the most important insights that stems from this work - the network was able to extract the broad structure of the inputs it received and therefore generalize to novel ones.

Without the ability to generalize over sets of inputs, the world would seem fragmented, and the variability in any given event would not only be overwhelming but also impossible to process. However, by grouping and associating different experiences and the individual elements that comprise them into functional categories, the brain is able to recognize and respond in an appropriate manner to both familiar and novel stimuli (Conaway & Kurtz, 2017; Seger & Miller, 2010; Seger & Peterson, 2013; Tversky & Itamar, 1978).

In the words of David Marr "*the world tends to be redundant in a particular way*" and "*the brain runs on this redundancy*." (Marr, 1970, p. 163).

2. Categorization

The ability to categorize is an invaluable survival skill, which allows animals to learn the commonalities between different elements in the environment (Ashby & Maddox, 2005; Richler & Palmeri, 2014).

For example, a specific type of berry or mushroom could be categorized as edible or poisonous based on its features, but also on the basis the consequences of ingesting them, and generalizing the outcomes of such action to other similar berries or mushrooms can be a highly adaptive skill. Without the ability to categorize, decision making itself would be nothing more than gambling on the outcomes, and could have fatal consequences (Cohen & Lefebvre, 2015; Iordan, Greene, Beck, & Fei-Fei, 2016; Nosofsky, 1988; Seger & Miller, 2010; Seger & Peterson, 2013). This means that

categorization requires the brain to adapt to stimulus variability in different dimensions, while disregarding other irrelevant features and sources of random noise in the perceptual system (Gauthier & Tarr, 2016; Seger & Peterson, 2013; Townsend & Ashby, 1986). This is what ultimately allows for the clustering of similar stimuli in what is often referred as a "perceptual space" or "psychological similarity space" (Conaway & Kurtz, 2017; Op de Beeck, Wagemans, & Vogels, 2003; Tversky, 1977; Tversky & Itamar, 1978). In other words, categorization is a process that relies on the perceived features of a given stimulus and the subjective weight that those same features acquire when processed by different regions in the brain.

Over the years, the field of visual category learning not only has been leading the discussion on this topic, but it has also been the most prolific in terms of theoretical and experimental output. For the most part, researchers have focused primarily on the nature of stimulus representations (i.e. object features and feature variability), the theoretical models which try to explain the mechanisms behind categorization (i.e. different types of categorization and criteria for category membership), and the neurophysiological basis of this process.

I will address each of these points in detail in the following sections and conclude with the basic hypotheses being tested in this set of experiments.

2.1 Stimulus representations

In terms of sheer computation, the main problem the visual system has to deal with, is the massive level of variability of any visual scene (DiCarlo, Zoccolan, & Rust, 2012; Nikos K. Logothetis & Sheinberg, 1996; Poggio & Riesenhuber, 2000; Rolls & Milward, 2000). At every moment, the distinct visual elements are subject to changes in light conditions, morphology and different types of affine transformations such as rotations or scaling (Gross, 2008). Furthermore,

the visual elements that comprise a given scene can differ from each other along multiple dimensions. Objects can vary in terms of more global properties such as shape (Folstein, Gauthier, & Palmeri, 2012; Folstein, Palmeri, & Gauthier, 2013; Freedman, Riesenhuber, Poggio, & Miller, 2003; Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008; Jiang et al., 2007), simple dimensions such as the size or brightness (Goldstone & Steyvers, 2001), or more complex dimensions which involve specific object components (Biederman, 1987; Erez, Cusack, Kendall, & Barense, 2016; Nosofsky, 1986; Richler, Wilmer, & Gauthier, 2017; Sigala & Logothetis, 2002).

Depending on the type of behavioural task used, object dimensions are often referred as being separable if the variability observed on irrelevant dimensions (i.e. dimensions which are thought to play a lesser role in identifying a given object) doesn't hinder performance during categorization tasks (Folstein et al., 2012; Op de Beeck et al., 2003; Richler et al., 2017). However, this separability between dimensions is not always possible. A good example is the extreme difficulty in differentiating between brightness and saturation, since variations in one of these dimensions can lead to perceived differences in the other one.

The context and purpose of categorization is also a fundamental part of the process, and unfortunately one which is often overlooked due to the nature of the experimental paradigms. In order to exclude any bias related to previous knowledge about a set of visual stimuli, researchers often use visual stimuli which bear little to no resemblance to real world objects (Hauffen, Bart, Brady, Kersten, & Hegdé, 2012; Kromrey, Maestri, Hauffen, Bart, & Hegdé, 2010; Okamura, Yamaguchi, Honda, Wang, & Tanaka, 2014; Tafazoli, Di Filippo, & Zoccolan, 2012; G. Wang, Obama, Yamashita, Sugihara, & Tanaka, 2005; Zoccolan, Oertelt, DiCarlo, & Cox, 2009). On the one hand, this is indeed a suitable way of assessing the ability of the visual system to group objects

based on dimensions which were specifically manipulated by the experimenter. However, this also raises questions about the real world applicability of such paradigms, since one cannot exclude the behavioural significance of real world objects as one of the (if not the) most important factor(s) in categorization (Peelen & Downing, 2017). Having said this, it is easy to find studies which use a multitude of visual stimuli in order to study this phenomenon, ranging from 2D to 3D objects, and encompassing anything from clouds of dots, computer generated objects such as "greebles", faces of human and non-human primates, among many others (Curby, Hayward, & Gauthier, 2004; Erez et al., 2016; Folstein et al., 2012; Freedman, Riesenhuber, Poggio, & Miller, 2002; Gauthier & Tarr, 1997; Goldstone & Steyvers, 2001; Kriegeskorte, Mur, Ruff, et al., 2008a; Palmeri & Nosofsky, 2001; Richler et al., 2017; Shepard, Hovland, & Jenkins, 1961; Todd Maddox, Gregory Ashby, & Bohil, 2003).

Another way of differentiating among objects, is by selectively attending to their components, a strategy that might be particularly useful when distinguishing between similar exemplars or when information from other dimensions is not available[2] (Erez et al., 2016; Sripati & Olson, 2010; Ullman, Vidal-Naquet, & Sali, 2002). In other words, the need to use any available visual information to quickly identify a given object becomes an essential survival tool for animals, which need to recognize whether a specific visual cue can be related to a predator or prey. The terms "diagnostic regions" or "diagnostic features", are often used by researchers interested in the problem of object recognition (OR), and refer to the regions or specific features that are used in order to identify a given image/object (Gosselin & Schyns, 2001; Nielsen, Logothetis, & Rainer, 2006). In one study, Nielsen et al., (2006) compared the strategies that humans and Rhesus monkeys used in order to discriminate between sets of natural images. In this task both monkeys

[2] For example low lighting conditions can render information related to colour virtually irrelevant, and the same can be said when the more global object dimensions such as their shape or size are partially or totally occluded.

and humans had to show that they could associate an image that was presented for hundreds of milliseconds with one of 3 dots that were presented in a computer screen afterwards (Nielsen et al., 2006). After reaching peak performance the images were partially occluded with bubbles, a technique first used by Goselin and Shynes (2001) in which an opaque mask punctured by randomly located circular windows is overlapped with the image (Gosselin & Schyns, 2001; Karimi-Rouzbahani, Bagheri, & Ebrahimpour, 2017; Nielsen et al., 2006; Royer, Blais, Gosselin, Duncan, & Fiset, 2015). The results showed robust differences between in the diagnostic feature size, with monkeys relying on features that covered around 7% of the images while humans used diagnostic regions covering 51% on average. Similarly, in a 2015 study, Rosselli et al., showed that when rats had to discriminate between easily distinguishable objects they relied on more stable and view-invariant features, but when the discrimination between objects was harder they tended to rely on a wide variety of specific view-dependent diagnostic features, which differed between animals (Rosselli, Alemi, Ansuini, & Zoccolan, 2015).

In essence, these findings seem to indicate that, in order to discriminate between different dimensions, the visual system needs to privilege certain features more than others; that is to say, the visual system needs to ascribe a heavier weight to specific features. As such, visual stimuli which share invariant features in any given dimension can be perceived as being similar.

2.2 Theoretical Models of Categorization

Over the years, several models of categorization have been proposed, but by far the ones which have received more attention, and have succeeded in explaining the results of several experiments are known as “reference point models” (Conaway & Kurtz, 2017). These models have their roots in the associative learning tradition, as they firmly place the process of category learning

as a form of stimulus generalization. According to the reference point framework, subjects make category judgements based on the subjective evaluation of similarity between a target object and the existing information in the subject's knowledge database (Ashby & Maddox, 2005; Homa, Sterling, & Trepel, 1981; J. D. Smith & Minda, 1998a). However, the main difference between the two main models within this framework, is in how the comparison process unfolds, and what the target stimulus is being compared against.

In 1968, Posner and Keele proposed a model called "Prototype theory", which assumes that categorization is a processes whereby the target stimulus is compared against a category prototype, which can be defined as the central tendency or average, of multiple observations of a given category (Posner & Keele, 1968, 1970). This means that the process of categorization is one of pure abstraction of the features that comprise a given set of stimuli, which ultimately creates a mental representation that serves as a template. In 1970, Posner and Keele also argued that the prototype pattern, unlike any other pattern that is derived from it, was resistant to decay; something which was later corroborated in several other studies (Homa & et al, 1973; Homa & Vosburgh, 1976; Posner & Keele, 1970; Rosch, 1973, 1975; Rosch & Mervis, 1975; Strange, Keeney, Kessel, & Jenkins, 1970)

About a decade later, a new model called Exemplar Theory was proposed (Estes, 1986; Medin & Schaffer, 1978; Medin & Schwanenflugel, 1981; Nosofsky, 1986, 1988). According to this model, the generalization which underlies the categorization process is based on the specific exemplars which have been stored during learning, and retrieved depending on the perceived similarity to the object or pattern being categorized. In this way, there is an assumption that each category is represented by its individual exemplars, which means that it does not need to rely on any kind of abstraction. Exemplar models have in fact enjoyed great success in explaining the

results obtained in different tasks. Their ability to fit human performance has been particularly effective, especially when expanded in order to include concepts such as selective attention and error driven learning such as in the GCM or ALCOVE models (Kruschke, 1992; Medin & Schwanenflugel, 1981; Nosofsky, 2011).

Later, in a series of experiments Smith and Minda (1998) have also hypothesised that both strategies - one which relies on the abstraction of a prototype and one which relies on specific exemplars - could be adopted depending on the type of task, the type of pattern being categorized (ill-defined categories vs well defined), and the number of patterns or items used in the categorization tasks (J. D. Smith & Minda, 1998).

It's also noteworthy, that the performance of both prototype and exemplar based models, seems to be highly correlated with the perceived typicality[3], as shown by studies involving virtually constructed face stimuli (Davis & Poldrack, 2014; Iordan, Greene, Beck, & Fei-Fei, 2016b; Nosofsky, 1988; Rosch & Mervis, 1975; Rosch, Simpson, & Miller, 1976; Sigala & Logothetis, 2002)

Other categorization models include Decision Bound Model, which conceptualizes categories as being represented in terms of a boundary that separates them in a continuous perceptual/psychological space, or Rule Based Models, which posits that category learning can be formalized based on specific (sometimes verbalized) rules (Busemeyer & Myung, 1992; Casale, Roeder, & Ashby, 2012; Snyder & Munakata, 2010).

Since their inception, these models have been hybridized in order to accommodate different behavioural tasks and specific parameters which might influence performance, and we now have

[3] Typicality refers to how typical a given exemplar is, of its superordinate category. Humans tend to respond more quickly to an exemplar that is more representative of a specific category (e.g. a robin is a bird) than to a more atypical one (e.g. a penguin is a bird) (Iordan et al., 2016; Rosch et al., 1976).

a plethora of models which can combine reference point and rule based approaches such as RULEX or ATRIUM (Erickson & Kruschke, 1998b; Nosofsky, Palmeri, & McKinley, 1994); SUSTAIN (Love, Medin, & Gureckis, 2004) which is based on clustered representations where some clusters can represent the central tendency or prototype, while others represent single exemplars; or even DIVA (Kurtz, 2007), which relies on principal component analysis (PCA) and an autoencoder artificial neural network.

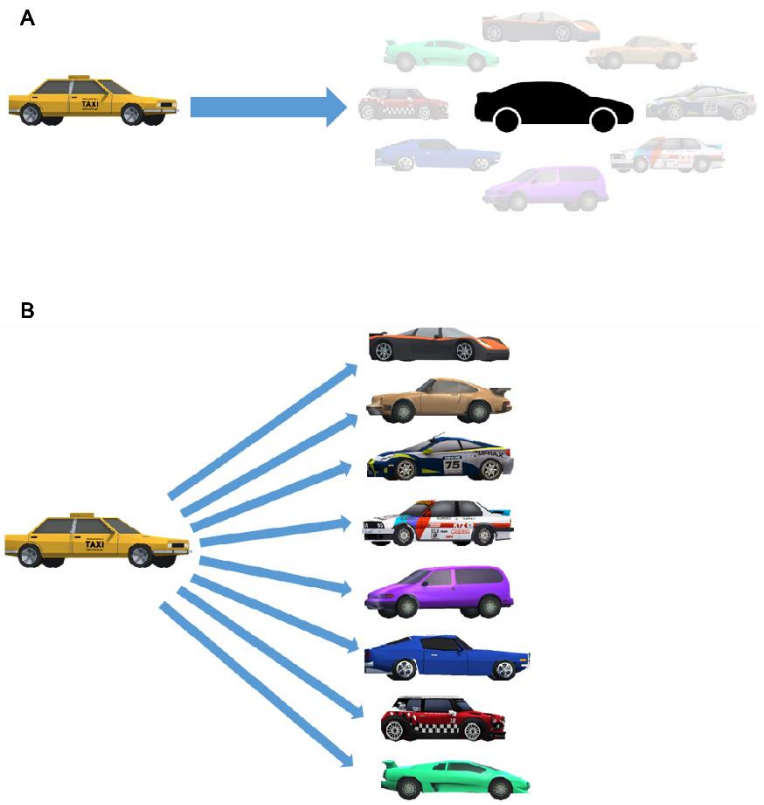

Figure 1.1. The two major theoretical models of visual categorization. A) Prototype Theory suggests that a new exemplar is compared against the category prototype, which is defined as the central tendency of multiple observations. **B)** Exemplar Theory posits that visual categorization relies on the direct comparison between the new object and the individually stored exemplars.

Overall, these models have permeated the study of human category learning for the past 50 years, and have provided valuable insights on a very complex process. However, the inner workings of categorization have received considerably less attention. This in part due to the fact that some authors have approached this problem from a uniquely human point of view, and on the other hand to the fact that the it's not always easy for different research fields to converge on empirical grounds. But over the last 20 years we've witnessed a significant progress in our understanding of how the brain learns to categorize and the specific neuronal blueprint associated with this complex process.

2.3 Neurophysiological Basis of Visual Categorization

The rapid extraction of regularities and the subsequent discrimination of its constituent elements is one of the most important tools for the survival of any animal species. In most mammals this remarkable ability is due to the way their visual system is structured, and how its functional organization enables an internal representation of a given visual scene to emerge. This representation can then be compared with previous stored ones and ultimately used to make decisions about the visual elements that have been perceived. All of this happens in a fraction of a second, and not only does it have to be fast, but it also has to be tolerant to changes in the environment that can potentially hinder the generalization between experiences.

Any attempt to decipher the neuronal dynamics of any cognitive process is invariably tied to the neuroanatomy of the brain, and the process of visual categorization is no exception. The idea of two separate processing streams of visual information can be traced back to early papers on the golden hamster visual system (Schneider, 1969). This apparent dissociation between a visual pathway that is mainly responsible for processing the location or the identification of a stimulus

have been reported in different species and although inter-regional communication between both pathways exists, converging lines of evidence seemed to indicate that they are responsible for very different aspects of processing visual scenes (Ettlinger, 1990; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013). In a seminal paper published in 1969, Gerald Schneider hypothesized that the identification of a given stimulus would take place along the genicolustriate pathway, a circuit that seems to be phylogenetically more recent than the retinotectal pathway which seems to be involved in visuo-spatial processing and oculomotor tasks (Lehky, Kiani, Esteky, & Tanaka, 2014; Schneider, 1969).

This basic idea was then further elaborated in a seminal paper by Mishkin and Unglerleider (1982), where the authors proposed a "where" versus "what" distinction for has become known as the dorsal and ventral visual streams (Mishkin and Ungerleider, 1982). This model has been updated over the years, and in a more recent paper it has been suggested that the dorsal visual stream can be further divided into three different pathways: (1) the parieto-preforntal, (2) the parieto-premotor and (3) the parieto-medial temporal pathways. Each of these pathways is believed to support different functions that play an important role in visual guided actions, spatial working memory and visuo-spatial processing (Kravitz, Saleem, Baker & Mishkin, 2011).

Further evidence from non-human primate research pointed in the same direction. In a study conducted by Weizkratz and Saunders (1984), lesions in the inferior temporal cortex (ITC) of monkeys critically impaired the animal's ability to generalize across different viewpoints in a 3-D shape recognition task, while lesions in posterior parietal cortex didn't seem to have any effect in this task (Goodale & Milner, 1992; Weiskrantz & Saunders, 1984).

The ventral visual stream comprises what is known as the occipito-temporal network, which is a set of bidirectional connections between regions in charge of processing visual

information along the rostro-caudal axis. These connections range from early visual areas in the occipital cortex such as V1, V2, V3 and V4, up to different temporal lobe regions such as the posterior temporal cortex (TEO) and the anterior temporal cortex (TE) in the ITC, which are considered to be at the apex of the cortical visual hierarchy in the primate brain. This set of regions seem to process visual information in a hierarchical fashion, with cells in the ITC responding to increasingly more complex stimuli compared to the relatively basic features that can elicit a response in early visual areas (Desimone, Thomas, Gross, & Bruce, 1984; D. H. Hubel & Wiesel, 1959; D. Hubel & Wiesel, 1964). Additionally, the receptive field properties of ITC neurons seem to rely on the integration of a broader information spectrum, with clusters of cells that can respond to full objects instead of simple elements and with an increased invariance to changes in light conditions, translations, rotations, color or size (Kravitz et al., 2013). And since neurons in the ITC cortex can respond to complex visual stimuli, and are sensitive to diagnostic features that allow for a more view-invariant identification of objects, they seem to have all of the appropriate features for a role in visual categorization (Booth & Rolls, 1998; Gross, 2008; Kobatake, Wang, & Tanaka, 1998; Nikos K. Logothetis & Sheinberg, 1996; Perrett, Rolls, & Caan, 1982; Sigala & Logothetis, 2002; Tanaka, 1996; Yasushi Miyashita, 1988).

One of the most remarkable studies to highlight the role of the ITC in the visual categorization process was done by Kiani and collaborators in 2007 (Kiani, Esteky, Mirpour, & Tanaka, 2007). In this study, the group decided to train three rhesus monkeys (*macaca mulatta*) in a simple fixation task, while recording the activity of single neurons (one cell at a time) in the ITC, as the monkey's visualized >1000 images of real world objects over multiple sessions. A neuron was regarded as category selective if the responses to the different category exemplars were significantly larger than to any other category. Concurrently, the population response was

determined by arranging the average response of every neuron to the image set in vectors which were then normalized.

Their results showed that the categorical structure of the images was represented by distributed patterns of activity over the recorded cell population (674 cells), with the population code showing clearly defined clusters that corresponded to animate and inanimate objects. Within the animate objects category, smaller clusters for primate and non-primate faces, which then could be divided into human and non-human primate faces, were also observed. The presence of cells that were selective for bodies of humans, non-human primates, birds and four-limb animals was also noteworthy, as well as the presence of cells that were more selective towards lower animals such as reptiles, insects and fish, resulting in a tree of interdependent population activity clusters.

However, it should be noted that the animals were not naïve and had been exposed to many animals and inanimate objects before, as they were raised in both human houses and zoos before the experiments took place. Furthermore, the fact that the data was collected over multiple sessions can also imply a learning component due to repeated exposures to the same set of stimuli, even if they were shown in a pseudorandom fashion each session.

The aforementioned results were corroborated in a different study conducted by the same group, where the authors compared the ITC response patterns in both monkeys and humans when they were presented the same set of images (Kriegeskorte, Mur, Ruff, et al., 2008). While the human subjects were presented 92 pictures in an event related fMRI experiment, the monkeys had their neuronal activity recorded using tungsten electrodes, as described in Kiani et al., (2007). To this effect, the authors decided to focus instead in comparing the response patterns elicited by the images for each subject within each of the 2 experimental groups and generating a representational

dissimilarity matrix for each species (for a review on their category identification methods see Kriegeskorte, Mur, & Bandettini, 2008).

The results seemed to align with the findings reported by Kiani et al., (2008), with defined category sub-clusters that corresponded to animate and inanimate objects identified in both groups as well. Whereas the categorical information pertaining to inanimate objects (particularly man-made objects) was less defined, especially in the monkey ITC, the categorical structure related to animate objects, and in particular faces and body parts appeared to be remarkably well preserved across species.

This being said, the results reported by Kriegeskorte et al., (2008) should be interpreted with some caution since the recording techniques and data collection used in each experimental group were quite different. Even though a lot of progress has been made in terms of comparing monkey and human brain activity with fMRI, the differences in terms of both spatial and temporal resolution of each recording technique is by no means negligible.

Nevertheless, previous studies had already shown that neurons along the IT seem to be highly selective to both human and non-human primate faces, with some cells firing only when a particular viewpoint is presented or when certain elements or features are present in a given configuration (Desimone et al., 1984; N. K. Logothetis, Pauls, Bülthoff, & Poggio, 1994; Wallis & Rolls, 1997). We now know that there is in fact a network of 6 interconnected regions along the primate ITC that are responsible for processing faces, with neurons that can encode for either left of right profiles (view-specific) and others that present a more view-invariant preference (Freiwald & Tsao, 2010). The aforementioned clusters of ITC neurons that encode for faces, body parts or scenes, seem to have very defined functions in both humans and monkeys, with some even suggesting that the consistency of their location can be attributed to an evolutionary based

selectivity for important visual elements, which relies on an efficient way of encoding information and an inter-regional communication mechanism that resemble small world, or hierarchical modular networks[4] (Hilgetag & Goulas, 2016; Telesford, Joyce, Hayasaka, Burdette, & Laurienti, 2011; Uhlhaas et al., 2009; Watts & Strogatz, 1998).

The ITC is also interconnected with medial temporal lobe regions (MTL) such as the perirhinal cortex (PRh); an association region that combines projections from multiple sensory areas in order to form a multimodal representation of objects as well as the specific features of a given environment, and then projects to the hippocampus (HPC) both directly through a network of relatively sparse connections to CA1, and via lateral entorhinal cortex (LEC) as well (Agster & Burwell, 2013; Brown & Aggleton, 2001; Burwell, 2001; Burwell & Amaral, 1998; Cloke, Jacklin, & Winters, 2015; Furtak, Wei Shau-Ming, Agster, & Burwell, 2007; Winters & Reid, 2010; Winters, Saksida, & Bussey, 2008). This information related to category membership is then sent to another major hub in the process of visual categorization – the prefrontal cortex (PFC) – through direct and indirect pathways which ultimately converge in this association region (De Curtis & Paré, 2004; Maurer, Burke, Diba, & Barnes, 2017; Webster, Bachevalier, & Ungerleider, 1994).

3. The Prefrontal Cortex

3.1 Functional Neuroanatomy

The term prefrontal (or *pre-frontal*) was first used in 1884, in a publication by David Ferrier and Gerald Yeo, and it was initially defined as the anterior two-thirds of the frontal convolutions in the primate brain (Ferrier, 1886; Ferrier & Yeo, 1884). After Brodmann's seminal studies

[4] Small world networks were first described by Watts and Strogatz in 1998, and refer to highly clustered networks with small path lengths which maximize the overall connectivity while minimizing the number of connections (Watts and Strogatz, 1998).

(Brodmann, 1909), the prefrontal cortex became almost uniquely associated with primate species and was often referred as "frontal granular cortex" or by the less common Latin term *regio frontalis* (Uylings, Groenewegen, & Kolb, 2003; Uylings & Van Eden, 1991). A definition predicated on the fact, that in the primate brain, the PFC is located rostral to the agranular pre-motor cortex, whereas the rodent frontal cortex is completely agranular and therefore considered phylogenetically more primitive (Seamans, Lapish, & Durstewitz, 2008; Uylings et al., 2003; Zaitsev et al., 2009).

From an evolutionary perspective, the prefrontal cortex has been the most controversial cortical region to define and to compare across species, since it has always been considered one of the hallmarks of human cognition due to its role in some of the most characteristic aspects of human behaviour (Fuster, 2001; Goldman-Rakic, 1984; Preuss, 1995). In the vast body of literature concerning this region, one can usually find references to its specific functions under the umbrella of cognitive control, top-down modulation, executive functions or inhibitory control; terms that often refer to very different aspects of brain activity, which are contingent on specific goals and relevant behavioural drives.

In primates the PFC can be broadly divided into 3 different parts. In the ventral-dorsal axis we find: (1) the orbitofrontal cortex, which corresponds to Brodmann's areas 13, 47 and the inferior part of areas 10 and 11; (2) the medial and cingulate prefrontal cortex, which encompasses areas 12, 24, 32, and the more medial parts of areas 8, 9 10 and 11; and (3) the dorsal and lateral (often referred as *dorsolateral*) prefrontal cortex, which corresponds to area 46 and the lateral portion of areas 8,9,10 and 11. These areas have subsequently been the subject of more detailed histochemical and immunohistochemical analysis which revealed a mosaic of 22 different subregions (Carmichael & Price, 1994; Seamans et al., 2008).

Conversely, the rodent PFC (here limited to mice and rats) is often divided into the medial prefrontal cortex (mPFC) and the orbitofrontal cortex (OFC), two topologically distinct regions that then can be divided further into several subregions. Among the mPFC subregions we can identify 3 main areas with noticeable differences in terms of laminar organization and involvement in specific cognitive functions. These 3 subregions evolved from both archicortical and paleocortical moiety and they are usually referred as: (1) the infralimbic cortex; (2) the prelimbic cortex; and (3) the anterior cingulate cortex (Pandya & Yeterian, 1990).

This being said, given the connectivity patterns, cytoarchitecture and the involvement in higher cognitive functions often ascribed exclusively to anthropoid primates, the existence if a homologous region in rodents became a polarizing topic among neuroanatomists, with early accounts (Rose & Woosley, 1948) identifying the prefrontal cortex solely on the basis of the afferents it receives from the medial dorsal nucleus of the thalamus (MD) (Preuss, 1995; Rose & Woosley, 1948; Seamans et al., 2008; Uylings et al., 2003; H. J.J.M. Van De Werd, Rajkowska, Evers, & Uylings, 2010; Henri J.J.M. Van De Werd & Uylings, 2014). But despite the somewhat controversial comparisons between species, the current view in terms of anatomical homologies suggest that: (1) the infralimbic region in rodents roughly corresponds to Brodmann area 25; (2) the prelimbic region can be seen as a more primitive version of the primate dorsolateral cortex (area 46) in terms of its overall functions, but in anatomical terms it seems to be more closely related to the ventromedial PFC (area 32); and finally, the rodent anterior cingulate cortex would correspond to area 24 in the primate brain (Seamans et al., 2008; Uylings & Van Eden, 1991; Henri J.J.M. Van De Werd & Uylings, 2014).

The connectivity within the dorsal and ventral subdivisions of the rodent mPFC seem to be quite robust, contrasting with the more sparse connections that exist between them. But the

differences between these mPFC sub-divisions, can also be observed in their connections with other cortical as well as subcortical areas (Condé, Maire-lepoivre, Audinat, & Crépel, 1995; Datiche & Cattarelli, 1996; Heidbreder & Groenewegen, 2003). The PFC possesses a remarkably vast array of cortical as well as subcortical connections, which makes it one of the final destinations for information arriving from different streams. Its connections can be seen as far as the spinal cord (Van Eden & Buijs, 2000), and several brainstem nuclei. Equally important are the direct connections between the mPFC and neuromodulatory systems, such as: (1) the dopaminergic innervation which predominantly stems from the ventral tegmental area (VTA) and to a lesser extent from the substantia nigra pars compacta (David B. Carr & Sesack, 2000; Thierry, Blanc, Sobel, Stinus, & Glowinski, 1973); (2) the serotoninergic projections from the raphe nuclei that reach the infralimbic and ventral prelimbic cortices (Heidbreder & Groenewegen, 2003; Harry B.M. Uylings et al., 2003); (3) the efferents of cholinergic neurons arriving from the nucleus basalis magnocellularis and the mesopontine laterodosral tegmental nucleus responsible for heightening arousal in the mPFC (Lamour, Dutar, & Jobert, 1984; Ragozzino & Kesner, 1998); and finally (4) the noradrenergic projections that reach the mPFC arriving from the locus coeruleus, which modulate the levels of other neurotransmitters such as dopamine in both the prelimbic and infralimbic cortex (Morrison, Molli Ver, Grzanna, & Coyle, 1979; Öngür & Price, 2000; Tronel, Feenstra, & Sara, 2004). Besides the aforementioned regions, the mPFC also has connections with other important subcortical regions that play a vital role in different brain functions such as the hypothalamus, thalamus, the amygdala and the basal ganglia. Through the extensive efferent connections to the latter, the mPFC plays an important role in decision making, anticipating the outcomes and valence of one's actions, and even in participating in motor output functions (Heidbreder & Groenewegen, 2003). The array of connections with these regions

supports important aspects of prefrontal functions such as the homeostatic regulation of processes related to basic drives, attention, level of motivation, emotional appraisal and even social behaviour (Fuster, 2000, 2001; Riga et al., 2014).

There is also considerable evidence that the mPFC plays a major role in memory stabilization on a time scale that can range from seconds to weeks, as well as in the process of memory retrieval. It's no surprise then, that the mPFC receives strong projections from the hippocampus, particularly from ventral hippocampus and subiculum. The received inputs might be reciprocated via indirect connections arriving at the HPC through the nucleus reunions of the thalamus (NR) and the PRh – lateral entorhinal cortex (LEC) pathway (Agster & Burwell, 2009; M. C. Anderson, Bunce, & Barbas, 2016; Brod, Lindenberger, Werkle-Bergner, & Shing, 2015; Euston, Gruber, & McNaughton, 2012; Euston & McNaughton, 2006; Godsil, Kiss, Spedding, & Jay, 2013; Hallock, Wang, & Griffin, 2016; Hernandez et al., 2017; Jarovi, Volle, Yu, Guan, & Takehara-Nishiuchi, 2018; Peters, David, Marcus, & Smith, 2013; Richards et al., 2014; Tripathi, Schenker, Spedding, & Jay, 2016; Xia et al., 2017).

Even though the PFC, and in particular the mPFC, is also involved in the early stages of information encoding (Bero et al., 2014; Kitamura et al., 2017; Lesburguères et al., 2011), this region is mostly known for its pivotal role in the processes of memory consolidation, and retrieval of remote memories (Euston, Gruber, & McNaughton, 2012b; Euston, Tatsuno, & McNaughton, 2007; Hebscher & Gilboa, 2016; Milivojevic, Vicente-Grabovetsky, & Doeller, 2015). The level of engagement of the PFC appears to be correlated with a slower component of memory consolidation that occurs concomitantly with a progressive disengagement (for some types of memory) from temporal lobe regions such as the hippocampus, an effect which seems to be particularly dependent on post encoding sleep (Gais et al., 2007; Takashima et al., 2006; Tse et al.,

2011; Wolbers & Buchel, 2005 but see J. Q. Lee, Zelinski, McDonald, & Sutherland, 2016; Sutherland, Sparks, & Lehmann, 2010; Sutherland & Lehmann, 2011). This was made evident in a now seminal study by Euston and collaborators (2007), where it was shown that during sleep, the rat mPFC replayed task-related spatiotemporal patterns of neural activity that were compressed by a factor of 6 to 7 (Euston et al., 2007). In a follow up paper, the same group also demonstrated that reactivation of those same neural patterns was correlated with the density of down-to-up state transitions, and were mostly associated with K-complexes and low voltage spindles; two distinctive electrophysiological features of the complex interplay between hippocampus and cortex during slow wave sleep (SWS) that are presumed to support memory consolidation (Johnson, Euston, Tatsuno, & McNaughton, 2010).

3.2 The Role of the Prefrontal Cortex in Visual Categorization

Due to the vast network of cortical fibers that converge in this region, the PFC cortex and its subdivisions are in a position to compare multimodal information arriving from both dorsal and ventral cortical streams (Condé et al., 1995; Heidbreder & Groenewegen, 2003; Room, Russchen, Groenewegen, & Lohman, 1985; Sakagami & Pan, 2007; Sakagami, Pan, & Uttl, 2006; Sakagami & Tsutsui, 1999).

In primates, the ventrolateral prefrontal cortex (VLPFC) receives information primarily from the ventral visual pathway, which mediates object recognition. On the other hand, the dorsolateral prefrontal cortex (DLPFC) receives projections from the dorsal stream regions in order to determine the spatial configuration of objects in the environment. This information can then be classified based on the emotional valence and other motivational aspects and used in order

to initiate the motor planning of goal directed actions (Pan & Sakagami, 2012; Sakagami & Pan, 2007; Sakagami et al., 2006).

The specific role of the PFC in visual categorization was addressed in a series of experiments by Freedman and collaborators (Freedman, Riesenhuber, Poggio, & Miller, 2001; Freedman et al., 2002, 2003). In these studies, the authors developed a variation of the Delayed Match to Sample task (DMS) called Delayed Match to Category (DMC), and used a three-dimensional morphing system to produce a set of images that belonged to 2 different categories: cats and dogs. These 3D images were generated through linear combinations of every possible arrangement between them, which allowed the researchers to define the category boundary based on the amount of "cat" or "dog" features that any given exemplar displayed. The goal of these experiments was to record the activity of dlPFC neurons in rhesus monkeys, while the animals decided whether a sample and a testing image belonged to the same category, and how the neural activity would reflect those same choices.

The results showed that the monkeys could accurately classify between objects (about 90% of the time), even when their physical appearance was close to the category boundary (60:40 cat-dog and vice versa), furthermore the recorded dlPFC neurons also showed a remarkable category selectivity as well. In the first paper, Freedman and colleagues observed that out of a total of 395 dlPFC neurons, 253 of those (64%) were active during the sample and/or delay interval, with roughly one-third of those (82/253) exhibiting category selective responses regardless of the degree of dog or cat features within the category boundaries (Freedman et al., 2001).

In a more recent paper by the same group, Roy et al., (2010) generated the morphing images by varying the percentage of two dog and two cat prototypes. This resulted in both within and between category morphing spectrums, with the images now being blended along six morph lines

instead of one. With this new set of images, the authors defined two categorization schemes with orthogonal boundaries that were then used to train the monkeys at different time points, and subsequently evaluated their performance during the recording sessions when the animals had to switch between the two schemes (Roy, Riesenhuber, Poggio, & Miller, 2010). Interestingly, they found that the same images, depending on the category scheme, were represented by largely different neuron population in the PFC. Most neurons (29.3% or 157 out of 536) showed category sensitivity for only one category scheme but not both, while a relatively small number of neurons were category sensitive to both category schemes (7.1% or 38 out of 536).

The degree of specialization and generalization was addressed in a following study by Cromer et al., (2010), where the authors expanded the original morphing task to include another set of images, in this case, besides the cat versus dog category (one single morphing spectrum), the monkeys had to learn to distinguish between two types of cars - sedans versus coupes (Cromer, Roy, & Miller, 2010). The purpose behind this 2x2 category task, was to see if neurons would show a more general type of encoding where individual neurons could respond to more than one stimulus set, or, if their activity would be very specific to one set, similar to the findings reported by Roy and collaborators. Their results showed that many PFC neurons (44% or 104 out of 236) were "multitasking", and showed a significant difference in their average firing rates for both category distinctions. But just like in previous studies, Cromer and colleagues also found neurons that were mostly category specialists (i.e. responding either to Cars or Animals, but not both). One possible explanation for the increased number of multitasking neurons lies in the structure of the task itself. Even though the results reported by Roy et al., (2010) seemed to contrast with previous reports on the multitasking abilities of PFC neurons (Duncan & Miller, 2002), the fact that 2 different category schemes, with the same fixed number of images were in direct competition

might explain the level of independence in neural representations. According to the authors, it is possible that neuronal specialization is driven by high cognitive demands such as when the same set of images is categorized in 2 different ways. On the other hand, when the categories are independent from each other, or not in direct conflict, the same neurons can be recruited to encode category information by either displaying an increase or a decrease in their average firing rates.

Different oscillatory patterns within PFC subregions such as the dlPFC and the vlPFC, during an abstract categorization task have also been reported. It appears that the level of activity in these two regions was contingent on the level of stimulus abstraction; with gamma oscillations in the vlPFC more engaged in lower levels of stimulus abstraction and dlPFC beta oscillations becoming more prominent with higher levels of abstraction (Wutz, Loonis, Roy, Donoghue, & Miller, 2018). Similarly, category learning seems to be accompanied by an increase in beta synchrony between the PFC and the striatum during correct trials (Antzoulatos & Miller, 2014), with the striatum exerting a stronger influence on the PFC. The anatomical loops between the basal ganglia (BG) and the PFC can facilitate the establishment of a functional circuitry that would enable the selection of the appropriate motor programs in the BG, based on the category information encoded in the PFC (Antzoulatos & Miller, 2011, 2014; Miller & Buschman, 2007; Uhlhaas et al., 2009). In addition, the striatum also seems to be able to predict the behavioural response before the PFC in the initial stages of a dot-based categorization task, when learning is more reliant on stimulus-response associations (S-R). The explanation for this phenomenon might lie in faster plasticity mechanisms within the striatum circuitry that could then facilitate the slower learning rate of the PFC (Antzoulatos & Miller, 2011, 2014; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008). Concurrently, this dynamic between the STR and the PFC gradually shifted as

the number of category exemplars increased, possibly reflecting the important role the PFC plays in category abstraction (Antzoulatos & Miller, 2014).

4. Purpose of this project

The main focus of this project was to determine how the dynamics of neuron ensembles change after the acquisition of substantial semantic knowledge, specifically, how the generation of categorical representations was reflected within specific regions in the neocortex. This project addresses two of the main postulates of Complementary Learning Systems Theory by McClelland et al. (1995).

Hypothesis 1: *Neural representations of related concepts share a high degree of similarity between them and become hierarchically clustered.*

Not much is known in terms of how the accumulation of semantic knowledge might affect the patterns of connectivity and the underlying neural interactions in different cortical regions, particularly in higher order association regions such as the mPFC.

One of the main assumptions of CLST states that as the cortex gradually adapts its weight matrix to accommodate new knowledge (McClelland, 2013; McClelland et al., 1995), the representations category exemplars are expected to cluster, similar to what has been reported by Kriegeskorte et al., (2008). This being said, I hypothesise that in the mPFC, the neural activity patterns elicited by exemplars from the same category will be highly correlated, as opposed to when the category exemplars belong to different categories, which are expect to be more orthogonal.

Hypothesis 2: *As experiences of certain stimuli become integrated, mice will form sparse conjunctive representations of specific stimuli in higher modules of the cortical hierarchy.*

One of the main underlying questions on how the brain represents similar as well as dissimilar experiences is related to the type of coding scheme. In order to encode a multitude of stimuli and their respective associations, the neural representations must minimize interference and redundancy without compromising the idiosyncratic aspects of each experience, in addition, the brain must maximize the amount of information that can be stored and ensure that the storage capacity in any given region is not exceeded, otherwise information would be lost or irretrievable (Földiak, 2002; McNaughton, 2010; Olshausen & Field, 2004; Rolls, 2016).

In order to efficiently encode the acquired categorical knowledge, higher modules of the cortical hierarchy (i.e association regions) seem to develop the ability to encode higher order conjunctive features (Kriegeskorte, Mur, Ruff, et al., 2008; Nikos K. Logothetis & Sheinberg, 1996). By taking advantage of the common or redundant properties across category exemplars, high order regions can get away with using a small set neurons and fewer spikes, without compromising the amount of information. Since there will be many category exemplars to store, and each might have a unique representation, a sparse code can be particularly useful in terms of preventing categories from interfering with each other.

Sparse coding has been hypothesized in several theoretical and experimental papers to optimize the number of different activity patterns in associative networks, making it an efficient way of encoding information (Marr, 1970, 1971; Rolls & Treves, 1990; Treves & Rolls, 1991). A

sparse representation or sparse coding can refer to 2 different but often related concepts – lifetime sparseness (or lifetime kurtosis) and population sparseness (Willmore & Tolhurst, 2001).

In lifetime sparseness a given neuron is silent most of the time, but displays high firing rates only when specific stimuli are presented or at certain time-points. Lifetime sparseness is usually calculated using kurtosis, the fourth moment of a distribution which measures its "peakedness". It has been observed at different stages of the cortical hierarchy in several animal species and it's commonly used computational models as well (Barlow, 1972; Field, 1987; Graham & Field, 2007; Olshausen & Field, 2004; Perez-Orive, 2002; Vinje & Gallant, 2000).

On the other hand, in population sparseness there's a large set of neurons available with only a small subset of them active at any particular time, therefore minimizing the number of units that are involved in the representation of a given event or in response to an external stimulus. This has been proposed as the most efficient way of storing information, somewhere in between the overly sparse or localist coding scheme which hinders generalization, and a fully distributed one, which requires a higher number of neurons and although robust, it is also more prone to interference between patterns as well as energetically expensive This means that a highly efficient code is one that is both highly selective and that increases the number of independent associations while minimizing the number of modifiable synapses used during encoding (Földiak, 2002; Marr, 1970, 1971; Ohiorhenuan et al., 2010; Rolls, 2016; Treves & Rolls, 1991; Wixted et al., 2014) .

In the first set of experiments we exposed a cohort of mice to several exemplars of object categories using a virtual reality setup, and addressed their ability to distinguish between them. In the second part, along with the exposure to the virtual object categories, the neural activity was also recorded at different time points using in-vivo 2-photon calcium imaging.

Chapter 2

ABSTRACT

Categorization is a process whereby individual experiences and single instantiations are unified by their commonalities into functional groups. In this paper we describe a novel visual categorization task for mice using an automated touchscreen operant conditioning chamber. By gradually incrementing the number of exemplars available in a pairwise object recognition task, mice learned to discriminate between virtual objects that belong to 2 different categories, one rewarded and one non-rewarded. This further allowed the animals to maintain the same level of performance when two new sets of objects from the initial categories were introduced all at once. Similar results were also observed even when the non-rewarded category was switched to a completely new one. Taken together our results suggest that through this incremental goal-directed task, mice can easily incorporate information into distinct visual categories associated with specific outcomes in a relatively short amount of time and to generalize the behavioural response to new exemplars.

A novel visual categorization task for mice using a touchscreen operant conditioning chamber.

Introduction

The ability to detect invariance between individual elements across different environments is crucial for survival. Grouping different types of stimuli reduces the vast complexity of any given environment by decreasing the number of elements for which a similar behavioural response can be selected (Cohen & Lefebvre, 2005; Gauthier & Tarr, 2016; Hélie, Turner, & Cousineau, 2018). This parcellation of the external world into distinct categories allows the brain not just to process information faster, by decreasing information load, but also to generalize the same behavioural output when encountering inputs that share similarities between them (Iordan et al., 2016a; Rosch & Mervis, 1975; Seger & Miller, 2010). But categorization is not merely a passive process that creates abstractions from individual perceptual elements; it is an active inference mechanism, indissociable from the functional aspects and valence of the elements being categorized, as well as with the outcomes associated with them (Peelen & Downing, 2017; Richler & Palmeri, 2014; Seger & Peterson, 2013).

The study of categorization has been at the core of different academic disciplines for many years and has spawned a plethora of theories, models and experimental works that approached the problem from different angles. This resulted in a heterogeneous body of knowledge that has allowed our understanding of this process to become increasingly complex (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008; Lindh, Sligte, Assecondi, Shapiro, & Charest, 2019; Seger & Miller, 2010). Although categorization can be studied in relation to

different sensory modalities, most of the scientific literature in the field of experimental psychology and behavioural neuroscience have focused on the concept of visual categorization in order to better understand the basic processes behind category learning.

Even though there is a considerable variability between tasks, types of stimuli and theoretical models that have been proposed, most studies have relied on either human or non-human primates (Freedman et al., 2001; Homa et al., 1981; Kriegeskorte, Mur, Ruff, et al., 2008; Nosofsky, 1988; Rosch, 1973; J. D. Smith, Redford, & Haas, 2008; Strange et al., 1970); with some notable exceptions such as pigeons (Cook & Smith, 2006; Güntürkün, Koenen, Iovine, Garland, & Pusch, 2018; Herrnstein & Loveland, 1964; Troje, Huber, Loidolt, Aust, & Fieder, 1999; Wasserman, Kiedinger, & Bhatt, 1988), dogs (Range, Aust, Steurer, & Huber, 2008), or even honeybees (Benard, Stach, & Giurfa, 2006).

Surprisingly, the most widely used class of mammals in scientific research – rodents – has been ignored for the most part when it comes to study the visual system. Due to their poor visual acuity, and the less sophisticated neural architecture of their visual system (when compared to primates), rodents have been regarded as unsuitable when it comes to study complex visual processes such as object recognition or categorization (Artal, De Tejada, Tedó, & Green, 1998; Balkema & Pinto, 1982; Lashley, 1930). However, over the last decade, rodents, and in particular mice, have become increasingly popular when it comes to study the general properties of the visual system.

The rodent and primate visual system share many similarities despite the former lacking of important visual features such as a foveal pit, having fewer cone cells, lacking ocular dominance columns and having a small number of visual areas. Notably, the rodent brain retains key aspects of the primate brain's functional architecture, with: (1) functional modules that correspond to a

dorsal and ventral visual streams (Glickfeld, Andermann, Bonin, & Reid, 2013; Q. Wang, Gao, & Burkhalter, 2011; Q. Wang, Sporns, & Burkhalter, 2012), (2) a hierarchically organized cortical scaffold (Coogan & Burkhalter, 1993; Felleman & Van Essen, 1991; Laramée & Boire, 2015; Rockland & Pandya, 1979) and (3) a network architecture which resembles both small-world networks (typically found in primates), and scale-free networks (Oh et al., 2014; Sporns & Bullmore, 2014).

On a more practical level, mice are also becoming increasingly popular as experimental subjects due to the widespread availability of transgenic lines as well as the molecular tools, which allow for in-vivo recordings and circuit labeling. In addition, the overall size of the mouse brain also allows for larger scale recordings, and they are also less expensive and relatively low maintenance when compared to other larger mammals (Huberman & Niell, 2011). Consequently, several studies have now shown that rodents can also be informative subjects with which to study high level visual processing in the brain.

In a series of experiments conducted in rats, Zoccolan et al., (2009), Tafazoli et al., (2012) and Rosselli et al., (2015), demonstrated not only that rodents are suitable for object recognition experiments using touchscreens, but also that they possess a remarkable flexibility terms of switching between different strategies that would allow them discriminate between the presented objects based on their morphological features (Rosselli et al., 2015; Tafazoli et al., 2012; Zoccolan et al., 2009). For example, in their paper Rosselli et al., showed that when rats had to discriminate between easily distinguishable objects, they relied on a more stable and invariant strategy, but when the discrimination between objects was harder, they tended to rely on a wide variety of specific features from those same objects.

More recently, rodents have also been used in classification or categorization-like tasks. Brooks et al., (2013) and Vinken et al., (2014) have trained rats to discriminate over a series of pictures with different aspect ratios or movie sequences respectively. Creighton et al., (2019) used mice in a one-trial object category recognition (OCR) task. In this adaptation of the object recognition task (Ennaceur & Delacour, 1988), the authors exposed mice to real exemplars of two different categories during a sampling phase, and then assessed their ability to discriminate between an object of one of the categories previously presented, and one object from a novel category, in a "Y maze" task (Brooks et al., 2013; Creighton et al., 2019; Vinken, Vermaercke, & Op de Beeck, 2014). Mice preferred the object belonging to the novel category, which indicated a generalized recognition of the initial categories to which the animals had been previously exposed.

In this study, I designed a behavioural task to test mice's ability to categorize based on a pairwise discrimination protocol, using a touchscreen operant conditioning box (Bussey, Muir, Everitt, & Robbins, 1997; Kim, Kwak, Yu, & Kaang, 2016; Markham, Butt, & Dougher, 1996; Mitchnick et al., 2018; Talpos, Winters, Dias, Saksida, & Bussey, 2009). However, instead of training mice with a fixed set of virtual objects, I decided to create a task that would allow the gradual incorporation of new objects into the existing datasets, as the animals learn to discriminate between them, so that the brain could slowly adapt to the variability between exemplars (McClelland et al., 1995; Seger & Peterson, 2013)

Materials and Methods

Subjects

A total of 12 adult C57/BL6 mice (Jackson Laboratories, 23 – 35 g, 3 – 8 months old; 4 males, 8 females) were used in this study. The animals were single housed in standard mouse

cages, with a room temperature of 24 °C under a 12 h light/dark cycle with the lights on at 7:30 AM and free access to food and water before the beginning of the behavioural training. The procedures were in accordance with the guidelines established by the Canadian Council on Animal care and with the protocols approved by the Animal Welfare Committee of the University of Lethbridge.

Mice were water deprived throughout the duration of the behavioural training. During this period mice were given a daily *ad libitum* access to water for 30 minutes in their home cages 30 minutes after the last training session, and their weight was maintained to at least 85% of the baseline value (average weight during the 3 days prior to the beginning of the training sessions).

Touchscreen Operant Conditioning Box

Mice were trained in a custom built automated operant chamber (230 x 230 x 230 mm) with a computer tablet (Samsung Galaxy Tab A: SM-T350; 208.28 x 137.16 x 8.2 mm; Android 5.0) for the virtual object presentation. The reward consisted of a drop of sucrose water (10% concentration) that was delivered through a silicone tube, which was connected to a metal tube positioned below the computer tablet's screen. The reward delivery was controlled by a pinch valve that would open every time a correct response was made, delivering approximately 2.5 µl each time. The synchronization between the touchscreen and the valve was achieved through an Arduino Mega 2560. The wall where the Arduino, the pinch valve, the reward tube and the computer tablet were inserted is removable (front window: 150 x 164 mm, with a divider measuring 150 x 5 x 5 mm), which means that it can be used outside the operant conditioning box and adapted to different behavioural paradigms.

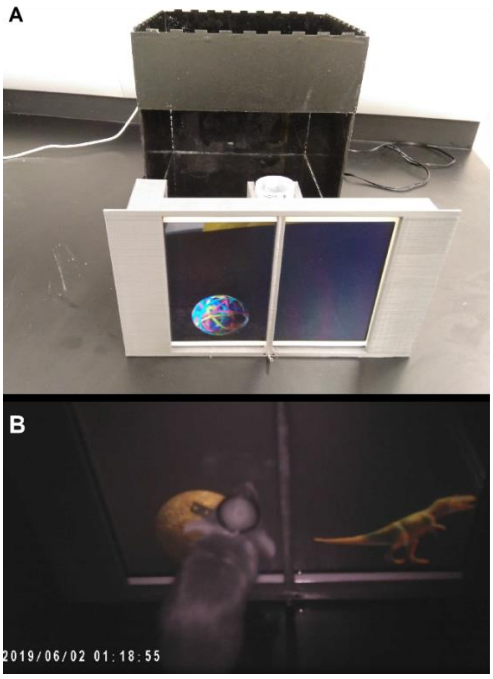

Figure 2.1. Custom-built touchscreen operant condition box. A) Operant conditioning box with the removable wall where the computer tablet is inserted. **B)** Picture taken during one of the sessions.

Virtual Reality objects

Most of the virtual reality (VR) objects were purchased at Unity Asset Store (Unity technologies), while others were virtually rendered from real objects using Autodesk's (Autodesk, inc.) single-camera photogrammetry software 123D Catch (discontinued). The VR objects were then modified using 3DS Max (Autodesk, inc), compiled and then finally rendered using Unity (Unity technologies).

The VR object categories were selected based on three criteria. First, the objects had to possess a degree of visual similarity within each category in terms of their overall shape. Second, although visual similarity is a requirement in order to belong to a given category, we ensured that the VR objects pertaining to a given category possessed distinctive features between them, such as the perceived texture, object components or colour. Third, the chosen categories should have a

high degree of visual dissimilarity between them. In short, we tried to maximize the intra-category visual similarity and the inter-category visual dissimilarity (fig. 1). I decided to use 3 main categories with 22 objects in total: ball, car and prism (columns, bottles, buildings), and a fourth one – dinosaur (bipedal) - comprised of only 5 objects, which we used as a control category. I also tried to keep the aspect ratio between objects belonging to a given category relatively stable (Ball: $\bar{x}$ = 45 mm x 45 mm; Car: $\bar{x}$ = 67.5 mm x 25 mm; Prism: $\bar{x}$ = 19.5 mm x 98 mm; Dinosaur: $\bar{x}$ = 68 mm x 40 mm). Finally, the virtual objects were then positioned on each side of the screen against a black background during the pairwise discrimination task.

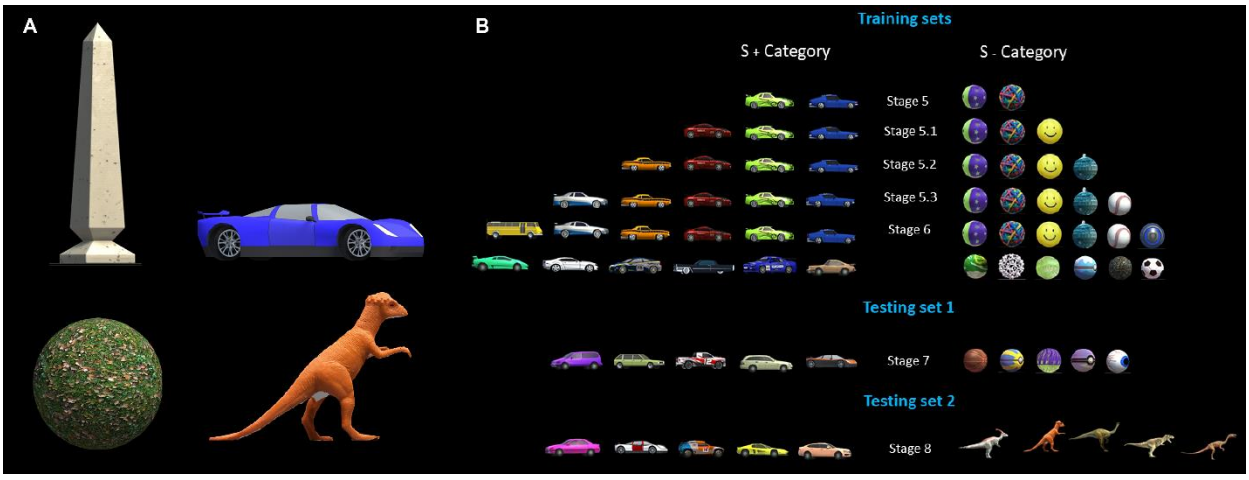

Fig 2.2. Virtual reality objects used in the behavioural task. **A)** Exemplars of the different categories as they appear in the computer tablets. Before the beginning of the pre-training phase, one of the three categories (prism, car and ball) is selected as the S+ category and one as the S- for each mouse. All of the categories are permutated between animals with the exception of the control category (Ctrl) which is comprised of a fixed group of objects (dinosaurs) that will subsequently be used as the last testing set in stage 8 and will only serve as a substitute for the initial S- category for every mouse. **B)** The touchscreen categorization task design. After the pre-training sessions where mice learn the basic rules of the pairwise discrimination task, they are gradually introduced new objects belonging to two given categories during the training sessions in order to facilitate generalization. Their ability to discriminate is then evaluated with 2 different testing sets, one comprised of completely new sets of objects from the previously defined categories, and another one where a completely new S- category (Ctrl category) was introduced.

Experimental design

The touchscreen categorization task is based on the touchscreen pairwise discrimination task described in previous studies (Horner et al., 2013; Mar et al., 2013; Nithianantharajah et al., 2015). In this task mice are required to make a choice between two images/virtual objects appearing on each side of the screen divider, by touching the surface where the objects are displayed.

Before pairwise discrimination sessions, the mice need to be shaped, which means that they must undergo some form of pre-training. The pre-training sessions were divided into four stages that would allow the animals to habituate to the specifics of the required task.

Habituation: The mouse was placed in the chamber for 10 and 20 min in the first and second day, respectively (stage 1 and 2). The screen was turned off and there was no tone nor reward. In the third day the mouse was introduced to the reward. After the animal was placed into the chamber, a 3 kHz pure tone was played every 10 sec to signal reward availability in the tube, and it continued irrespective of the reward collection. This phase took 2 sessions in 2 different days with the duration of the first session set to 20 min and the second one to 40 min.

Object presentation: In stage 3 mice were introduced to two VR objects (e.g. car 1 and 2) that belong to one of the defined categories (S+). These were presented one at a time in a pseudorandom fashion, in either the left or right side of the screen, and they were paired with a tone and a reward. The reward was always delivered at this stage, regardless of the animal's input. The screen was cleared after 30 sec, followed by a 10 sec. inter-stimulus interval (ISI) before a new trial started, with the total duration per session set for 60 min or 30 trials.

The rewarded and non-rewarded categories (S+ and S-) were permutated between mice to ensure that the accuracy of the discrimination was not exclusive to a specific pair of categories (e.g. mouse 1: car versus ball; mouse 2: ball versus car, etc.)

Touchscreen interaction: This is the stage where the animal must learn the association between touching the screen where the object is presented, and collecting the reward. No reward is delivered if other parts of the screen were touched and as in the previous stage, the session ended after either 60 min or 30 correct trials, but from this stage onwards there was no time limit for the object display. In the next stage, the animals were introduced to a small time-out on commission of an error. If the screen is touched anywhere besides where the object belonging to the S+ category is displayed, a 1.5 kHz pure tone and a white screen were presented for 5 seconds, and no reward was delivered; this condition was followed by the normal 10 sec ISI. The total duration of this stage depended on how fast the animals could reach the passing criterion, which was defined as 80% of correct responses (24/30) for two consecutive sessions.

Training Sets (stages 5, 5.1, 5.2, 5.3 and 6): After successfully completing pre-training, mice begin the training phase, which is divided into 5 different stages (fig. 2) that were designed to allow the animals to become familiarized with the categories and their respective objects in a gradual manner. From this point onward, training takes place twice a day, 2 sessions in the morning and afternoon respectively.

Stage 5 starts with the same 2 exemplars from the S+ category that were presented during pre-training plus 2 new objects that belong to an S- category. These stimuli are presented in a pseudorandom fashion on either the right or left side of the screen. The animals must make a choice by touching the screen where either the S+ or the S- object is displayed (e.g. S+ a or b versus S- a or b). If mice touched the object from the S- category a 1.5 KHz pure tone was presented with no

reward delivery, followed by a 5 sec time out and normal 10 sec ISI. As in the previous pre-training stage the passing criteria was defined as 80% correct trials per session for 2 consecutive sessions. I then added one more object to each category for the next three stages of the task (stages 5.1, 5.2 and 5.3), allowing the animal to associate the new objects that are gradually added to the previous ones. In the last stage of the training phase however (stage 6), unlike the previous ones, I accessed the ability of mice to generalize the same response to seven new objects that were added to the previous five, all at once. In this stage mice had a total of 12 S+ and S- objects that they need to correctly discriminate before moving to the testing sets.

Testing set 1 (Stage 7): Here, the previous 24 virtual objects (12 from each category) were removed, and in their place 5 new objects from both S+ and S- were presented. The animals had never seen any of these objects before, which means that they had to rely on the morphological similarities between them and the previous ones in order to reach the passing criteria. If mice had developed an S+ and S- category representation during the training sessions, they would have been able to maintain a similar level of performance with the new sets.

Testing Set 2 (Stage 8): The final testing set consists of 5 entirely new S+ objects and the introduction of a new S- category, to which I called "control category" (Ctrl). This category is comprised of 5 bipedal dinosaurs and was the same for all mice regardless of their starting S+ and S- categories. This is clearly the most difficult stage of the task, not only due to the fact that the animals are not familiar with the morphology of these new objects, but also because rodents also tend to be drawn towards novelty (Winters et al., 2008).

After each session the data was automatically stored in a .CSV file and analysed using MATLAB R2018b and GraphPad Prism version 9.0.0.

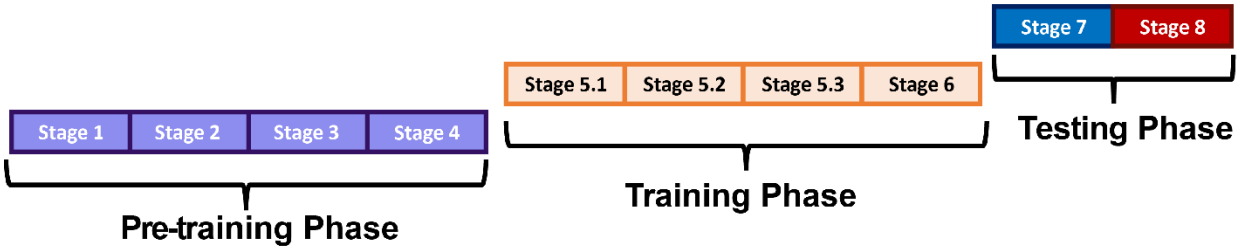

Fig 2.3. Experimental timeline 1. The touchscreen categorization task can be divided in 3 main phases: pre-training (stages 1 and 2: habituation; stage 3: object presentation; stage 4: touchscreen interaction), training (stages 5.1, 5.2, 5.3 and 6), and testing (stages 7 and 8).

Results

All of the animals used in this study were able to learn the categorization task. Even though there were individual differences in the amount of time that was necessary for them to learn the task, the individual learning curves converged into a specific pattern for the most part (Fig. 4). The initial exposure to the pairwise discrimination task, where any of the two objects from the S+ and S- category were presented seemed to present a bigger challenge than the subsequent ones where new objects were being added to the sets from previous stages. However, as soon as the mice extracted the basic rule of the task during that initial stage, they had no problems in generalizing the behavioural output to newer exemplars. On the other hand, when the S- category was replaced by the Ctrl category, most mice had a significant decrease in their performance, with some animals reaching chance level responses (15 out of 30 correct trials), which resulted in more sessions spent in stages 5 and 8 in order to reach the passing criteria (stage 5: $\bar{x}$ = 10.17; stage 5.1: $\bar{x}$ =3.25; stage 5.2: $\bar{x}$ =3.08; stage 5.3: $\bar{x}$ = 3.08; stage 6: $\bar{x}$ =2.08; stage 7: $\bar{x}$ = 2.16; stage 8: $\bar{x}$ = 7.75).

Nevertheless, after a few sessions, all mice reached the passing criterion of 80% correct trials and were able to accurately discriminate between the control category and the rewarded one.

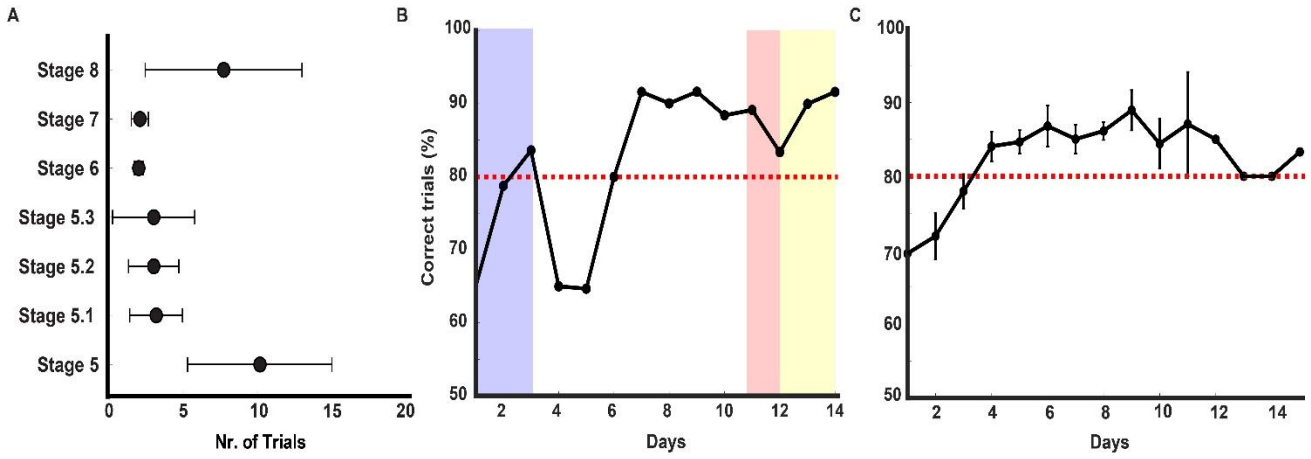

Fig 2.4. Performance in the touchscreen categorization task. **A**) Average number of trials to reach the passing criterion for all mice. **B**) Learning curve of one of the mice tested in the categorization task. The learning curve starts from the last stage of the pre-training phase (blue shaded area). The red and yellow shaded areas represent the time the animal spent in the two testing sets, stage 7 and 8 respectively. The dashed red line indicates the passing criterion of at least 80% correct trials for two consecutive sessions. **C**) Average learning curve of all animals during the training phase (Stages 5, 5.1, 5.2, 5.3 and 6). Error bars represent the SE across days. Since mice finished the task at different time points, fewer animals are still undergoing behavioural training/testing as time progresses, which explains why the error bars disappear after day 10, since only 1 animal went past that point.

To assess the disparity between the average number of mistakes in each of the training stages, an analysis of variance was conducted. The analysis yielded a significant variation among the errors in each stage, $W(6, 131.4) = 27.58$, $p < .001$, with a significant difference between the average number of mistakes between stage 5 and all of the subsequent stages ($p < 0.001$, Welch ANOVA with Dunnett T3 *post hoc* test). However, there were no differences in the average number of mistakes between stages 5.1, 5.2, 5.3 and 6, indicating that after the initial learning stage of the pairwise discrimination, the number of mistakes remained relatively low, even when several new objects were introduced at once.

Next, the number of errors in the two testing sets was compared in order to evaluate the changes in the task performance between them. The number of mistakes incremented substantially

when the unrewarded category was switched for a new one. A Mann-Whitney test indicated that the distributions in stage 7 (Mdn = 4) and stage 8 (Mdn = 7) differed significantly (U = 655.5, n_1 = 30, n_2 = 117, p < .0001 two-tailed). The sharp decrease in response accuracy during stage 8 was actually comparable to the number of mistakes in stage 5 (fig 3 (B)), the initial stage of the training set (Mann-Whitney U = 6480, n_1 = 122, Mdn_1 = 8; n_2 = 117, Mdn_2 = 7, p = .217, two-tailed). These results indicate that such manipulation brings the animals to a level of performance almost as low as in the period where the task is completely new. The starting S+ category didn't seem to have any effect on the average number of errors in stage 8 (Welch ANOVA W(2, 62.02) = 3.02, p = 0.071), and out of the 12 mice tested, only two didn't show any decrease in the performance in that same stage, and kept the correct response rate above 80%.

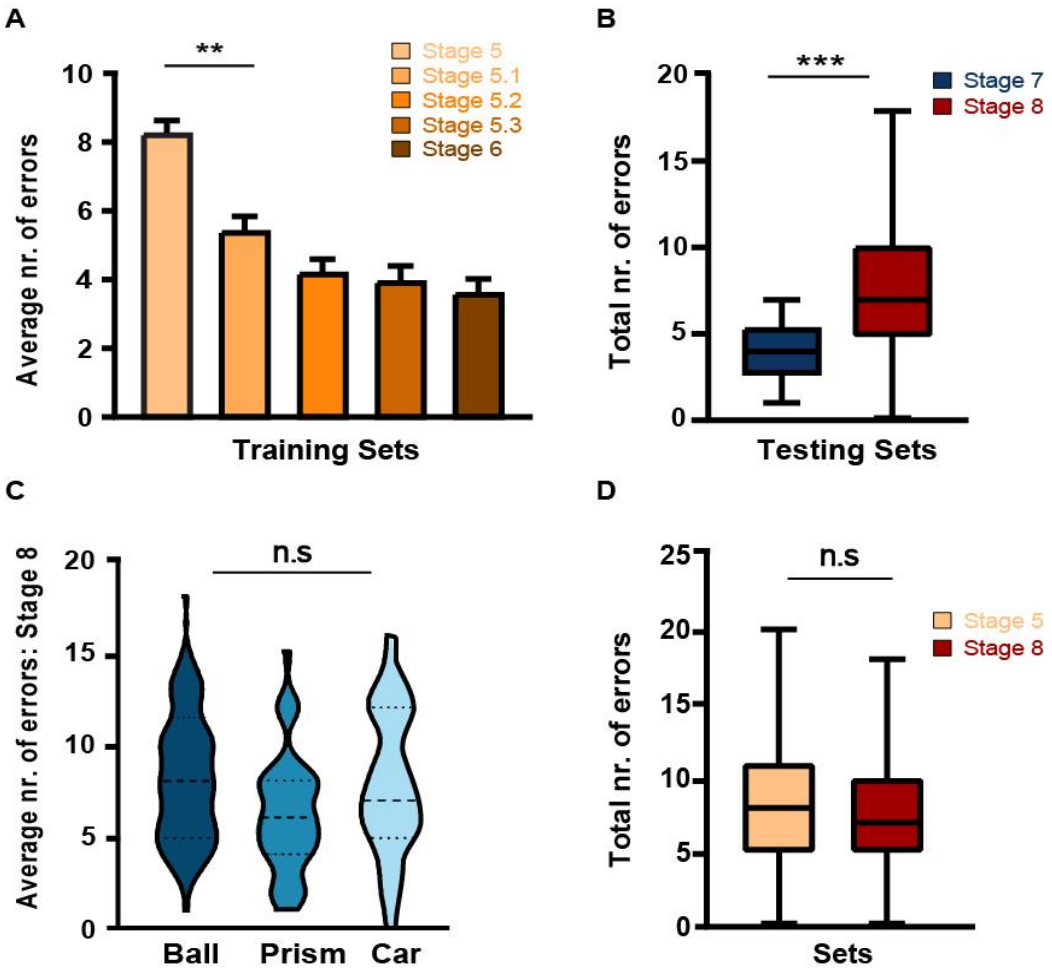

Fig 2.5. **Number of errors steadily decrease with more exemplars of the same initial categories but increase sharply as a new S- (Ctrl) category is introduced**. **A**) Difference in the average number of errors during the training stages (chance level = 15 errors: mean ± SEM in each stage, *** $p < .001$). **B**) Box plot of the total number of errors in the two testing sets (line: median; box: 25th and 75th percentile; whiskers: minimum and maximum values, ***$p < .0001$). **C**) Violin plot showing the average number of errors during the last testing set when the Ctrl category (the new S-) is introduced, based on the initial S+ category, which depending on the mouse can be "ball", "prism" or "car". **D**) Total number of errors in the initial stage of the training set and the last stage of the testing set (line: median; box: 25th and 75thpercentile; whiskers: minimum and maximum values). No significant difference was found between the distributions of the two groups.

Discussion

The overall goal of these experiments, was to develop a task that allowed for the gradual incorporation of categorical knowledge about different sets of objects, and in particular, to test how robust the generalization within the S+ category is, despite the introduction of different S- categories and their respective exemplars.

Based on the results I obtained, it's important to emphasize that the number of objects presented throughout the task didn't diminish the animals' ability to discriminate between categories, as long as they were added incrementally in the initial stages. And although the exposure to more exemplars of each category contributed to the improvement in the overall performance, suggesting a generalization of the objects morphologic features, it's also possible that this improvement was mostly due to the continuous training on the contingency. Therefore, a wiser assessment of the reported results should mostly focus on the robustness of the responses to the S+ category exemplars, regardless of what was presented as the S- category.

It's also noteworthy that mice were able to extend the behavioral response to completely new sets of objects such as the ones presented in stages 7, but most of them showed some initial difficulties in adapting when a new unrewarded category was introduced in stage 8. One possible explanation for decrease in performance is the fact that rodents tend to be naturally driven towards novelty, and even though there was no change in either the rewarded category or the rules of the task, this tendency could explain the sharp initial contrast with the previous stages.

One could outline the basic rule of the pairwise discrimination in this task as: *"When any of the elements that comprise A and B are present,* (e.g. A_1, B_1; A_2, B_2…A_n, B_n) *choose the ones that predicts reward* (e.g. A subset). This means that when B is replaced by a new category C, the mouse needs to adapt the rule so that now, any other object besides the ones that comprise A (not

just B), will not predict a reward and therefore should not be touched. Due to the nature of the task used here, where objects are presented in pairs, the categorization process implies a rule whereby the alternative choice to A leads to an outcome (no reward + time out), and therefore is part of the categorization process itself, as the set of exemplars against which A and the respective associated outcomes are compared.

These aforementioned assumptions are deeply related to the nature of categorization *per se*, and over the years several models have been proposed. From reference point approaches such as Prototype and Exemplar models (Estes, 1986b; Homa et al., 1981; Lamberts, 2000; Medin & Schaffer, 1978; Nosofsky, 1986; Reed, 1972; Rosch, 1973), to Decision Boundary or Rule-based models (Ashby & Maddox, 1994; Lockhead, 1966; Salatas & Bourne, 1974; Townsend & Ashby, 1986; Wright & Katz, 2007). Mixed or hybrid models such as RULEX (Nosofsky et al., 1994), ATRIUM (Erickson & Kruschke, 1998a), SUSTAIN (Love et al., 2004) and DIVA (Kurtz, 2007) have also gained considerable attention, either by mixing aspects of different models or by conceptualizing them in a spectrum.

Nonetheless, it is important to mention that these models have been tested using different stimuli, with different properties in different tasks. Anything ranging from dot patterns, Gabor patches, simple shapes, and abstract or real world objects has been used and tested in either human or non-human subjects (Folstein et al., 2012; Freedman et al., 2002; Goldstone & Steyvers, 2001; Palmeri & Nosofsky, 2001; Shepard et al., 1961; Todd Maddox et al., 2003). This extreme variability in experimental setups has resulted in some rather inconclusive results in terms of fitting an overarching theoretical framework for category learning, which further showcases the difficulty in defining these processes in both theoretical and empirical grounds (Folstein et al., 2012).

One could argue that in our task, the animal's ability to categorize might be explained by more than one strategy, by relying on both object features and a specific set of rules determined by the pairwise discrimination task. However, I cannot infer about the specific type of strategy or categorization mechanism used by these mice, since I didn't set up this particular task to test any of the foregoing models, nor their accuracy in describing the way these animals learned to categorize.

I also wanted to develop a virtual reality categorization task that could potentially be used in conjunction with neural recording techniques such as in-vivo 2-photon calcium imaging. And although the task, in its current form, was designed for using an automated operant conditioning box, it can easily be modified and adapted to other experimental paradigms where a choice between 2 items and their associated outcomes needs to be made; similar to what has been reported in previous studies (Andermann, Kerlin, & Reid, 2010; Komiyama et al., 2010; Mayrhofer et al., 2019).

To summarize, I think that the behavioural task presented here is well suited for accessing the process of visual categorization in rodents by gradually incorporating new exemplars of each category. In addition, this task can also be adapted for use in combination with any *in vivo* recording technique in order to further inquire the neural substrate of these mechanisms.

Chapter 3

ABSTRACT

The ability to categorize different stimuli based on shared features a fundamental principle of cognition. It has been suggested that such process recruits different brain regions which might play a unique role depending on the task rules, object features, category boundaries or even the overall context. One of such areas is the primate dorsolateral prefrontal cortex (dlPFC), a region which has been shown to encode category relevant information. However, not much is known when it comes the mouse brain and the specific role its different regions play in visual categorization. In this chapter, I describe a set of experiments that were conducted in order to assess the neural correlates of visual categorization in the mouse medial prefrontal cortex, the homologous region to the primate dlPFC, using 2-photon calcium imaging. Our goal was to determine if there were any changes the network dynamics related to the gradual acquisition of categorical information, both at the single neuron level and at the population level.

Visual categorization in the mouse Medial Prefrontal Cortex

Introduction

Animals form internal representations of the world in which they navigate, and recalibrate those same internal representations as they learn with each new experience. Our internal model of the world is therefore shaped through a type of information encoding that allows for the extraction of basic properties and regularities in the environment that can be generalized across different experiences. This accumulation of generalized knowledge and its statistical and categorical structure is what allows the brain to make predictions about the possible outcomes of specific events and thus, orient behavior.

A theoretical account on how information is encoded and organized in the brain was made by McClelland, McNaughton and O'Reilly in 1995, in a conceptual learning model called "Complementary Learning Systems Theory" (McClelland et al., 1995). In the original paper the authors tried to investigate some of the main mechanisms behind memory consolidation and knowledge acquisition using an Artificial Neural Network model first introduced by Rumelhart and Todd (Rumelhart & Todd, 1993). This network was trained using a set of specific propositions regarding several concepts or categories, in order to learn relationships between them over a specified number of iterations. Initially (by construction) the internal representations were highly distributed and didn't seem to follow any pattern of similarity; however as learning progressed, those representations started to display more structured patterns in terms of the concept relationships. The population activity became sparser and the selectivity of the responses increased, resulting in category representations that were further apart.

But how does the brain learn to create categories, and what regions are involved in this complex process of abstraction? Over the past few decades several studies have shown that the

ability to categorize depends on different regions which are responsible for different aspects of the overall process (Freedman et al., 2001; Kriegeskorte, Mur, Ruff, et al., 2008; Pan & Sakagami, 2012; Seger & Miller, 2010). One such region is the PFC, a major hub in the cortical hierarchy that receives inputs from several regions involved the process of memory consolidation, object recognition and categorization, such as the HPC and the PRh and the ITC; it also shares reciprocal connections with an array of other cortical and subcortical regions responsible for a variety of sensory, motor and cognitive functions (Brod et al., 2015; ~~D B~~ Carr & Sesack, 2000; Cowen & McNaughton, 2007; Euston, Gruber, & McNaughton, 2012; Szabo et al., 2006).

The role of the primate PFC (in particular the lPFC or dlPFC) in the process of categorization has been the subject of several studies that showcased its ability to distinguish between sets of visual stimuli and group them based on shared sensory features, or according to more abstract rules (Wutz et al., 2018).

In a series of seminal experiments, Freeman and collaborators trained monkeys in a variation of the classical delayed match to sample (DMS) paradigm where they used a three-dimensional morphing system to create a set of stimuli that would fall into 2 main categories – dogs and cats (Freedman et al., 2001, 2002, 2003). These morphed images were linear combinations between the exemplars of these 2 categories, and by blending them the authors created a continuum where the most extreme exemplars of each category would sit on opposite sides of the spectrum. The results of these experiments showed that monkeys could effectively distinguish between the 2 categories, and a substantial portion of lPFC (approximately 1/3 of the randomly selected neurons), was category responsive. Interestingly, not only were the monkeys able to distinguish between exemplars that were close to the category boundary (e.g. 60% dog and 40% cat and vice versa), but there was a sharp difference in the selective responses of lPFC neurons

to each category, regardless of how close the stimulus was to the boundary. The results in these and other experiments by the same group in more recent years (Cromer, Roy, & Miller, 2010; Meyers et al., 2008; Roy et al., 2010), point towards an ability to create abstract representations and generalize information carried out by the PFC, which can then be used to select the appropriate response according to the task demands.

The present set of experiments was designed to address two of the main CLST predictions: (1) as experiences of certain stimuli become integrated, mice will form sparse neuronal representations of specific stimuli in higher modules of the hierarchy such as the PFC; and (2) neural representations of related experiences share a high degree of similarity and are less orthogonal when compared with unrelated or novel experiences.

In order to address these questions, I trained three mice to discriminate between different object categories in an automated touchscreen conditioning box, mentioned in the previous chapter. I then recorded the neural activity at different time points during learning, with in vivo 2-photon Ca2+ imaging using a microprism implanted in the mPFC (Low, Gu, & Tank, 2014), a region that is generally viewed as the homologous to the primate dlPFC.

Materials and methods

All animal procedures were conducted in accordance with the guidelines established by the Canadian Council for Animal Care and were approved by the Animal welfare Committee of the University of Lethbridge.

Three female Ras-CRE Ai162D (TIT2L-GC6s-ICL-tTA2, Jackson laboratories) transgenic mice (19 – 23 g, 2 – 4 months old at the time of surgery) were used in this study. This transgenic mouse model was specifically chosen due to its strong GCaMP6s expression along midline cortical regions, a feature that seems to be lacking the widely used Thy-1 GCaMP6s. However, the overall

GCaMP expression throughout the brain is also quite heterogeneous across animals and in some mice I observed a very small number of neurons (< 20). The entire process ended up being quite challenging due to the variability of the GCaMP6 expression, especially since this could only be observed after the necessary post-surgical recovery time.

Surgical procedure

Mice were administered dexamethasone (0.2 mg/kg, intramuscular) followed by 0.5 ml of a mixture of 5% atropine and dextrose (3 μg ml−1 , subcutaneous) before being anesthetized with isoflurane (1%–1.5%, O2: 0.5-1 L/min). A subcutaneous injection of Lidocaine (7 mg/kg) was administered under the incision site with the animal's body temperature maintained at 37°C by an infrared heating pad.

The surgical procedure for the placement of the microprism was based on the one reported by Low et al., (2014), who were the first group to successfully implant a microprism in the mPFC. In order to access the mPFC, I used a right angle microprism (1.5 mm side length, BK7 glass; Tower Optical Corporation) with an aluminum coating and a dielectric overcoat protection on the prism hypotenuse to enable internal reflection. The prism was then bonded to the center of a circular glass window (3.0 mm diameter coverslip, BK7 glass; Warner Instruments) using UV curing optical adhesive (Norland #81).

The craniotomy coordinates varied between 2.0 – 2.5 mm in the anterior to bregma, and 1 – 1.5 mm posterior to bregma. The target coordinates for the microprism placement were somewhere between 0.7 and 1 mm anterior to bregma, with the prism placed with its front face along the midline, pressed against the wall of the contralateral hemisphere. Brain vasculature, and in particular the most anterior branches of the superior sagittal sinus were also used as a reference

in order to place the microprism in the appropriate region. But similar to what was reported by Low et al., (2013) these coordinates were subject to slight adjustments as the brain vasculature can differ significantly among animals.

The microprism and coverslip compound was implanted in the most anterior part of the craniotomy and attached to the skull using tissue adhesive (Vetbond, 3M). A custom-built titanium headplate was then fixed to the skull (Metabond, Parkell) with a rubber ring attached along its perimeter, which allows for the immersion medium (dH_2O) used in the 2-photon recordings to be retained and also to act as an additional light shielding mechanism. Mice were then allowed to recover for 2 weeks before the first imaging session.

Two-photon Imaging

I used a Thorlabs Bergamo II multiphoton microscope for data acquisition in all of the experiments. A Ti: Sapphire pulsed laser (Coherent) was set to an excitatory wavelength of 920 nm (~ 80 – 120 mW laser power measured at the sample) in order to penetrate the brain tissue and excite the fluorophores. This was achieved through a laser scanning controlled by Galvo-Resonant X-Y mirrors through a 16x water immersion objective (NA = 0.8, Nikon).

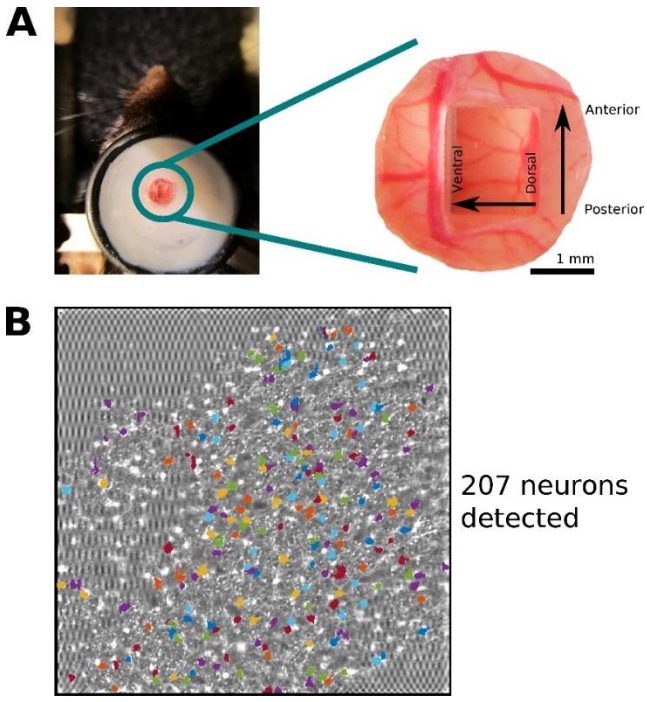

Figure 3.1. Cranial window and neurons detected using 2-photon calcium imaging. A) Top view of the cranial window and microprism implant. **B**) Neurons detected in one of the sessions from mouse nr. 1 after the preprocessing stage.

The emitted fluorescent signals were detected and amplified by a GaAsP photomultiplier tube (Hamamatsu) and digitized at a sampling rate of 19 Hz to a resolution of 800px x 800px.The imaging samples were collected from an 835 µm × 835 µm field of view (FOV) over layers II and III of the left prelimbic (PrL) cortex and the rostral portion of the anterior cingulate cortex (ACC) at depths between 100 – 200 µm. Before the beginning of each experiment a strip of Velcro was wrapped around the objective. The Velcro was then lowered slightly bellow the rubber ring in order to shield the sample from the light emitted by any external source, and in particular, the computer tablet (Samsung Galaxy Tab A: SM-T350; Android 5.0) used in these experiments.

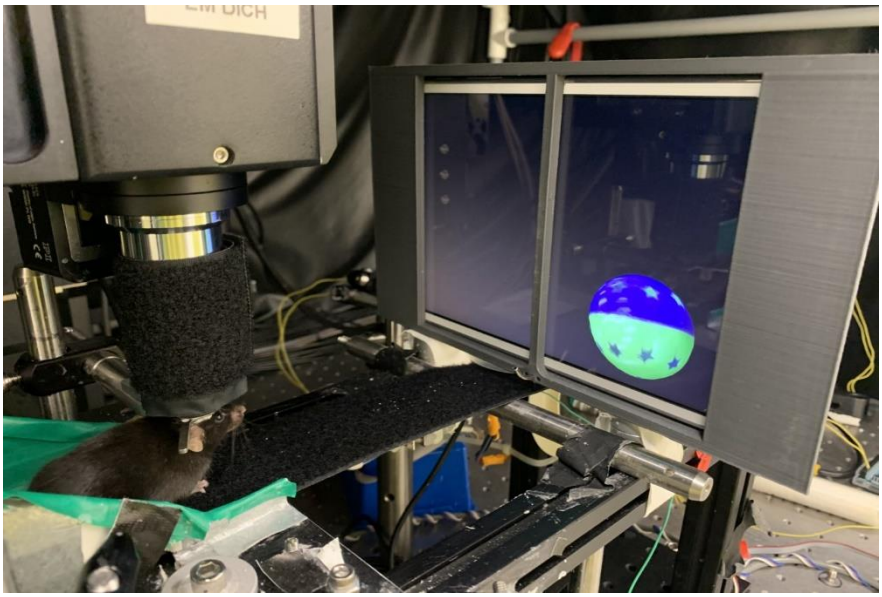

Figure 3.2. Experimental setup for the imaging sessions. The removable wall of the computer tablet was positioned in front of the mice with one object appearing at a time, on the right side of the screen, contralateral to the hemisphere where the imaging occurred.

Experimental design

Figure 2.3 shows the experimental timeline for the 3 mice. In this set of experiments, I used the same behavioural task described in detail in the previous chapter (mouse nr.1: Ball v.s Prism; mouse nr.2: Prism v.s Car; mouse nr.3: Car v.s Prism). The only difference in the experimental design was the assessment of the neural activity as the mice learned to distinguish between the predefined categories. The recording sessions took place at specific time points during learning. The purpose behind this design was to observe the changes in the mPFC network as the animals learned to distinguish between an incrementing number of object exemplars and how the activity at both single neuron level and at the neural population level would reflect that same learning process.

A preliminary imaging session took place 2 weeks after surgery. The goal of this session was to assess the fluorescence expression and the overall number of neurons. If the number of available cells was too low (< 50) the animal was excluded for the study. This step is particularly

important since small tears in the brain's microvasculature when implanting the microprism can occlude areas within the region of interest (ROI) for several weeks.

The baseline session took place one or two days after the initial assessment of the recording region and it was the first time the animals were exposed to the virtual reality objects. During each imaging session the animal was head-restrained on a fixed platform while passively viewing a set of 60 virtual objects that belonged to the categories defined in chapter 1, plus 4 objects that didn't belong to any category (spheres = 16; cars = 16; prisms = 16; dinosaurs = 7; misc. = 5). The duration of the whole recording session was set to 15 min, where the stimulus would be presented for 5 seconds, followed by a 10 second ISI. The synchronization between the photosensor and the computer tablet where the objects were displayed was achieved through and Arduino Mega 2560 using a custom built code that would generate a .CSV file at the end of each session. This file contained the timestamps of the microscope pulses and the appearance of different objects on the screen.

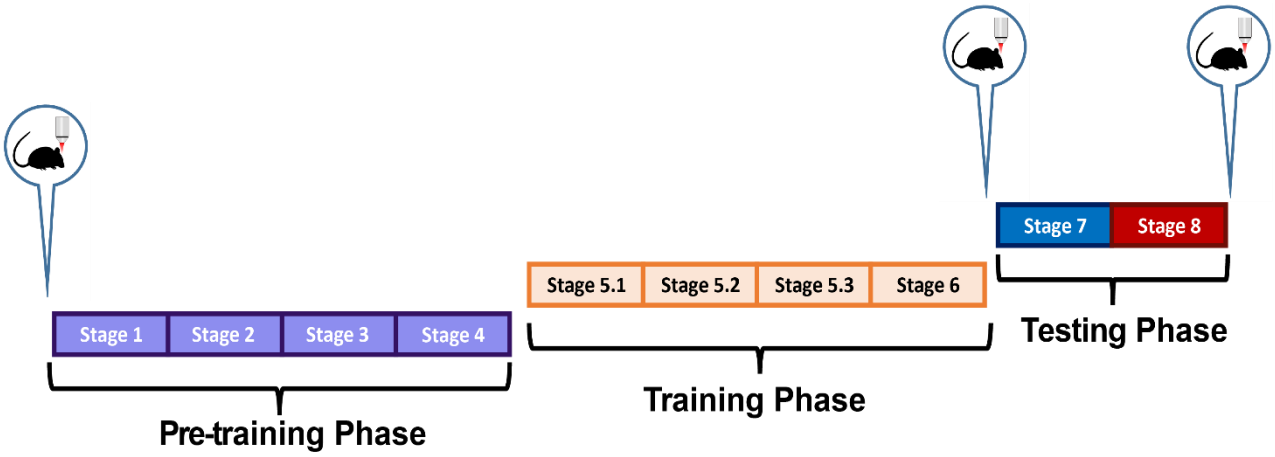

Figure 3.3. Experimental timeline 2. A baseline imaging session took place before the beginning of the pre-training phase when the animals were still naive. A second imaging session took place after the completion of the training phase, when the animals had already been exposed to several category exemplars. The final imaging session took place after the completion of stage 8, where the control category was introduced.

Data analysis

Image preprocessing

The image registration and the estimation of ROIs (regions of interest) was conducted automatically using the Suite-2P software (Pachitariu et al., 2016), as described in previous studies from our lab (Chang et al., 2020; Esteves et al., 2021; Mao, Kandler, McNaughton, & Bonin, 2017). The ROIs were then inspected through graphical user interface that allowed for a manual curation of the results and the raw fluorescence traces were determined for each ROI. Neuropil contamination was estimated and subtracted from the surround of the ROI's, and the baseline fluorescence was estimated by the Δ*F/F* ratio (Bonin, Histed, Yurgenson, & Clay Reid, 2011). In order to infer the neuron firing rates for each ROI the Δ*F/F* time courses were then deconvolved using the constrained non-negative matrix factorization method (Pnevmatikakis et al., 2016). The data were subsequently analysed using the deconvolved time courses in MATLAB versions R2018a and R2019a (MathWorks).

Peri-Stimulus Time Histogram (PSTH)

The mean firing rate of neuron i as a function of time from stimulus onset $f_i(t)$ was calculated as:

$$f_i(t) = \frac{1}{Nt\Delta} \sum_{\tau=1}^{N_t} \sum_{\iota=1}^{\Delta} n_i(\tau, \iota)$$

Where N_t is the number of trials, Δ is the width of the time bin t and n_i is the time-course vector of the deconvolved firing rates for neuron i. The deconvolved time-courses was circularly shuffled 1,000 times to obtain a null distribution of PSTH. Neurons whose response curve either exceeded the 95^{th} percentile or fell short of the 5^{th} percentile of the shuffled responses over

continuous time segments longer than two seconds were classified as stimulus-receptive neurons (SRN). I conducted the same analysis over the trials of individual categories to identify neurons that expressed significant response specificity for particular categories.

To measure the consistency of an individual neuron's response across trials, a reliability coefficient was computed. Reliability is taken as the Pearson correlation coefficients obtained from the neuronal response time vectors between pairs of trials of the same category (a total of $\frac{(T-1)T}{2}$ comparisons for T trials). Subsequently, this sample distribution was tested against the null hypothesis of $r = 0$ using a one-sample Wilcoxon Signed Rank test. Neurons with median Pearson coefficients higher than the null hypothetical distribution at $\alpha = 0.05$ were considered as category specific. Previous studies that evaluated the stability of the representations across trials by Pearson correlation did so by comparing the even vs odd trials (Salz et al., 2016). Here, to account for biases introduced by small trial numbers, I evaluated all pairs of trials.

Characterization of Population and Lifetime Sparseness

Treves and Rolls (1991) defined *population sparseness* as:

$$s = \frac{E[R]^2}{E[R^2]} = (\frac{1}{N}\sum_{i=1}^{N} r_i)^2 / \frac{1}{N}\sum_{i=1}^{N} r_i^2$$

Where r_i is the firing rate of neuron i in response to a stimulus for N neurons. Using this equation, I computed the population sparsity with r_i as the mean population vector over the 5 seconds of stimulus presentation for distinct object categories.

I used the same equation to measure lifetime sparseness by simply replacing the mean population vector (r_i) for the averaged activity vector of every individual neuron during the 5 seconds of stimulus presentation.

The population activity was decoded using an independent Bayesian decoder (Esteves et al., 2021; Mao et al., 2018). In brief, for every time bin, we estimated the probability of the animal viewing an exemplar of a specific category given the population response of all imaged neurons.

I used *stratified* K-fold cross validation ($K = 5$) to estimate the accuracy of decoding. Trials were partitioned into five equal-sized subsamples so that, at each iteration, a fifth of the trials was used for testing, while the remainder of the trials were used for training. While trials were randomly drawn, the proportions of each category were approximatively the same in each partition. Accuracy was reported as the mean over the five iterations. Alternatively, leave-one-out cross validation (LOO-CV) was used, and in this case, the confusion matrix was obtained by cumulating the results over all trials.

$$\Pr(c|n) = \frac{\Pr(n|c)\Pr(c)}{\Pr(n)}$$

Where Pr(c|n) is the probability of category "c" given the population firing rate vector "n". Assuming that the deconvolved firing rates of neurons obey a Poisson distribution we have:

$$\Pr(n|c) = \prod_{i=1}^{N} \Pr(n_i|c) = \prod_{i=1}^{N} \frac{(\tau f_i(c))^{n_i}}{n_i!} \exp(-\tau f_i(c))$$

$$= \prod_{i=1}^{N} \left(\frac{(\tau f_i(c))^{n_i}}{n_i!}\right) \exp\left(-\tau \sum_{i=1}^{N} f_i(c)\right)$$

$$Pr\ (\mathrm{c}|\mathrm{n}) = C\ (\tau, n)\ Pr\ (c) \left(\prod_{i=1}^{N} (fi(c)^{ni})\right) \exp\left(-\tau \sum_{i=1}^{N} fi(c)\right)$$

Where $f_i(c)$ is the mean deconvolved fluorescence of neuron i as a function of category c, and n_i is the time-course vector of mean activity within time bins of length τ. $C(\tau,n)$ is the

normalization factor that depends on τ and the population firing rate vector n, and that sets the sum of $Pr(c|n)$ to 1.

Results

Most of the neurons which were classified as stimulus-receptive fired at different time points during the stimulus presentation and not necessarily at the onset of the stimulus as one might have expected. Others had their activity increase prior to the stimulus presentation, which indicates that its activity was not category-dependent, and even if there is any "category preference" embedded in their firing rates, such information is not evident, since I didn't find a single neuron that reliably fired at each presentation of a given category, one of the hallmarks used to determine if a neuron was category selective in previous studies (Meyers et al., 2008; Pan & Sakagami, 2012).

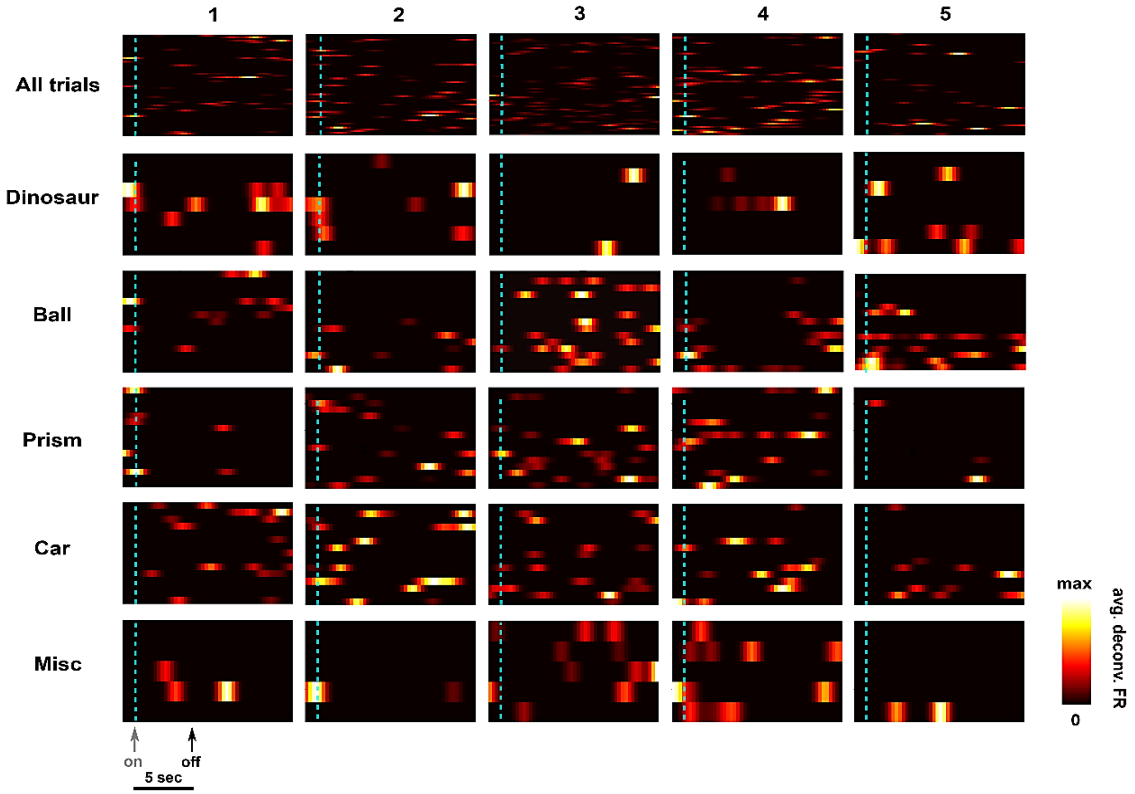

Figure 3.4. Single neuron PSTH. Normalized activity of 5 representative neurons from mouse nr. 1, for all trials (top row), and for individual trials of the same category of objects (rows 2 through 6). The vertical axis represents the number of trials and the horizontal axis represents time (-1 sec. to 14sec) with the dashed green line indicating the onset of the stimulus.

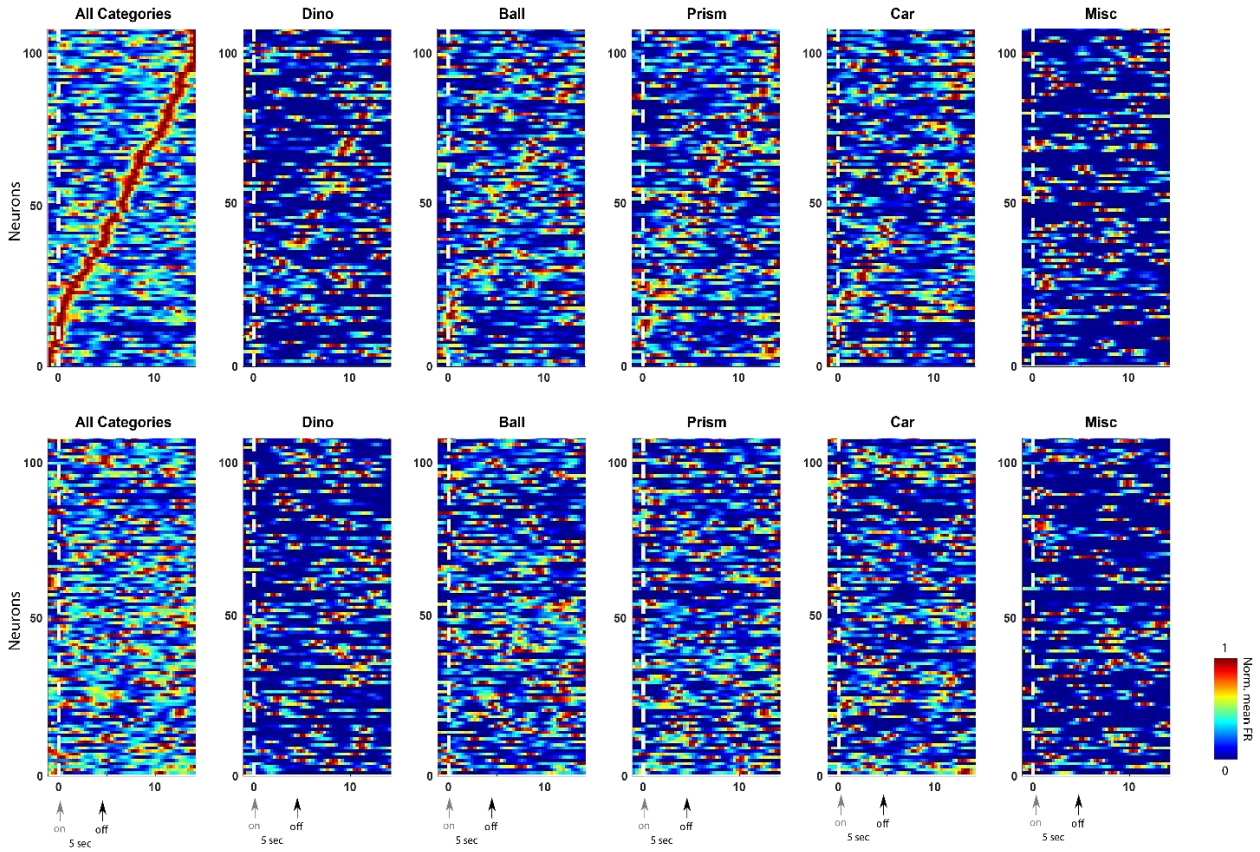

Figure 3.5. Neuron Population PSTH. Neurons' average normalized activity for all categories (first column) and individual categories of objects (columns 2 through 6). Neurons were classified based on the selectivity of their response towards individual categories (see Methods). The vertical axis represents the number of neurons and the horizontal axis represents time (-1 sec. to 14sec) with the dashed white line indicating the onset of the stimulus. On the top row the neurons are organized according to their peak activity, which means the organization changes in each panel, whereas the bottom row shows the neurons always in the same order. Only neurons that expressed significant response tuning towards the respective categories are plotted. Data showing the same animal and recording session as in figure 3.1.

Bayesian Classifier

As mentioned in the previous section, I decided to build a Bayesian classifier in order to ascertain whether the neural activity reflected the viewed objects category membership and if there was any distinctive pattern emerging during the stimulus presentation.

The overall accuracy of the Bayesian decoder was quite low, and I was unable to estimate which object was being presented based on the neural activity alone, even when I leave all but one

trial for our testing set (LOO-CV) (figure 2.6). Given that the stimuli were not uniformly sampled across the categories, but instead biased in favour of "Ball", "Prism" and "Car", chance level was not 20%. Rather, the sampling distribution was approximately 11.6%, 26.6%, 26.6%, 26.6% and 8.3% for the categories "Dinosaur", "Ball", "Prism", "Car" and "Miscellaneous" respectively.

In the case of mouse nr. 1, the mean ranks corresponding to the dinosaur category and the miscellaneous category are significantly different from the mean rank of the prism category. However, this might be the result of the overrepresentation of the 3 main categories, since there was no difference between the mean ranks of the ball and prism and car category (Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (4)= 17.32, p = 0.0017). Similar results were found in the data acquired from mouse nr. 2, where we found a statistically significant difference in the mean ranks of the ball category and the miscellaneous category for mouse nr. 2 (Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (4)= 11.55, p = 0.021), and for mouse nr. 3 as well, where the mean ranks corresponding to the dinosaur and miscellaneous category were significantly different from the mean rank of the prism category (Kruskal-Wallis one-way ANOVA on ranks: $\chi 2(4)$= 19.29, p = 0.0007).

When I analyzed the decoder accuracy just for the three main categories each animal was exposed to (fig 2.6 B and D), I observed that, with the exception of mouse nr. 2, the difference between the accuracy in regards to the S+ and S- categories when compared to the Ctrl category was statistically significant (mouse nr. 1: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 8.82, p = 0.0121; mouse nr.2: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 1.85, p = 0.3956; mouse nr.3: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 10.2, p = 0.0061), even if the overall accuracy was still quite low.

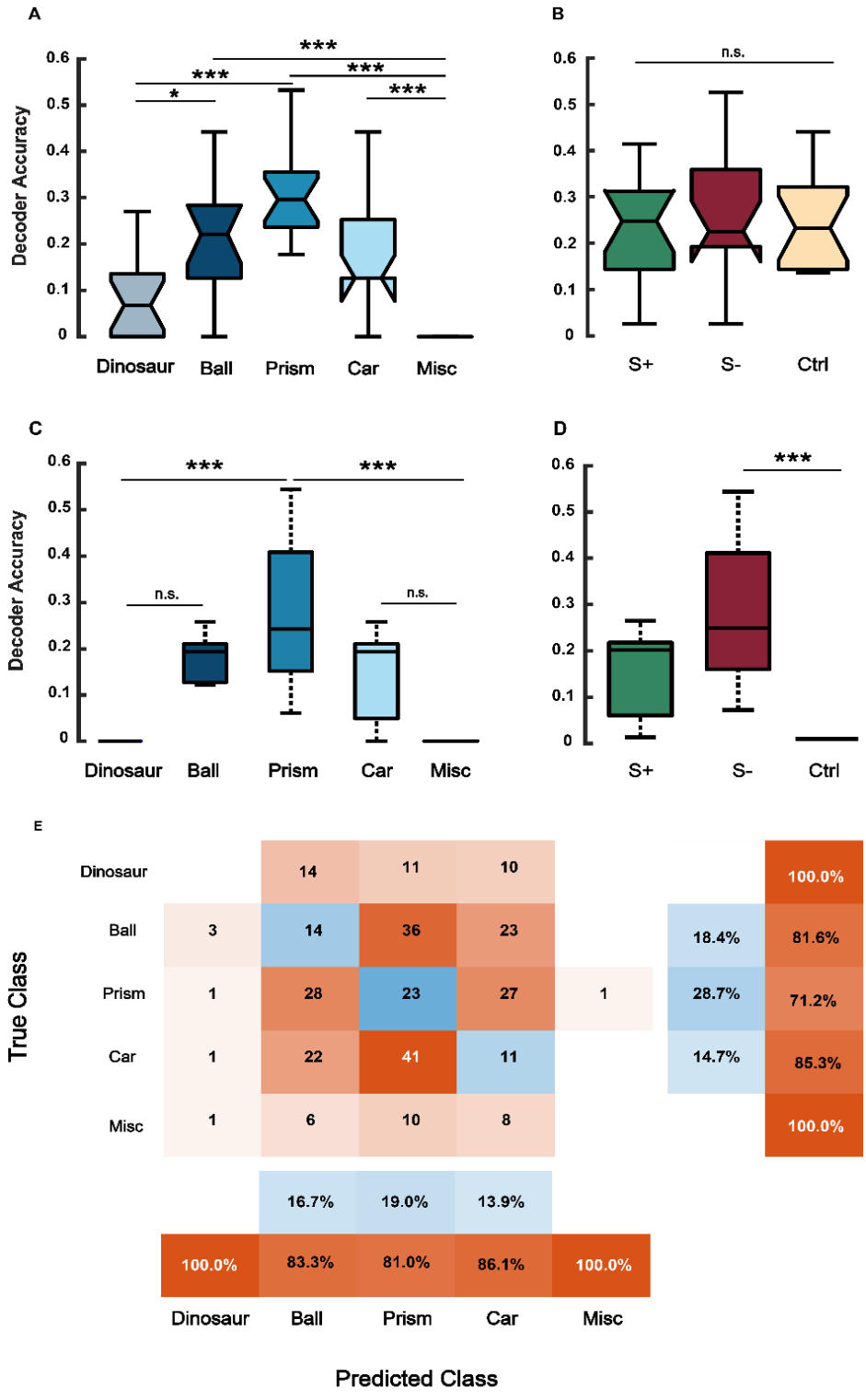

Figure 3.6. Accuracy of Bayesian decoding as obtained through Leave-One-Out cross-validation for individual object categories. On the left (**A** and **C**), scores for individual categories are reported irrespective of valence for all animals (n = 3 mice; Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (4) = 18.21; *p<0.05 **p<0.01 ***p<0.001) and for mouse nr. 1

respectively. On the right (**B** and **D**) object categories are grouped by the S+, S- and control categories for all animals (n = 3 mice; Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 1.32) and for mouse nr.1 respectively. Box plots centre line: median, box limits: first and last quartiles, whiskers: data range excluding outliers. **E**) Confusion matrix chart for the Bayesian decoder accuracy for mouse nr. 1. The rows of the confusion matrix correspond to the true class, whereas the columns correspond to the predicted class. The diagonal and off-diagonal values show the correctly and incorrectly classified objects respectively. The row-normalized summary shows the percentages of correctly and incorrectly classified objects for each true class, whereas the column-normalized summary displays the percentages of correctly and incorrectly classified objects for each predicted class.

Similarity

The correlation matrices corresponding to the baseline and endpoint sessions for the three mice also indicated that there were no discernible effects of the category learning process over time (Figure 2.7). In addition, the correlation between the neural representations of the 5 categories was non-significant for each animal, regardless of the particular session (mouse nr.1 baseline: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 3.5, p = 0.4784; mouse nr.1 endpoint: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 3.63, p = 0.4581; mouse nr.2 baseline: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 3.29, p = 0.5111; mouse nr.2 endpoint: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 3.39, p = 0.4951; mouse nr.3 baseline: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 3.34, p = 0.5029; mouse nr.3 endpoint: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 3.19, p = 0.5273)

I tried to see if there was any correlation in terms of neural activity between the S+, S- and Ctrl category, and if there was any difference between those (figure 2.7). Only for mouse nr. 2 I was able to observe a difference between the S+ and S- correlation and the remaining groups (S+ vs S- median = 0.3823, S+ vs Ctrl median = 0.2579, S- vs Ctrl median = 0.2880; Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 7.91, p = 0.0192). The other two mice didn't show any difference between the correlations being tested (mouse nr.1: S+ vs S- median = 0.3616, S+ vs

Ctrl median = 0.2122, S- vs Ctrl median = 0.1328; Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 4.22, p = 0.1212; mouse nr. 3: S+ vs S- median = 0.4142, S+ vs Ctrl median = 0.2429, S- vs Ctrl median = 0.2661; Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 1.34, p = 0.5117). The correlations observed for each animal, although moderately positive, don't allow for any meaningful conclusion about the representations of the different categories the animals were exposed to, even when in terms of their specific valence.

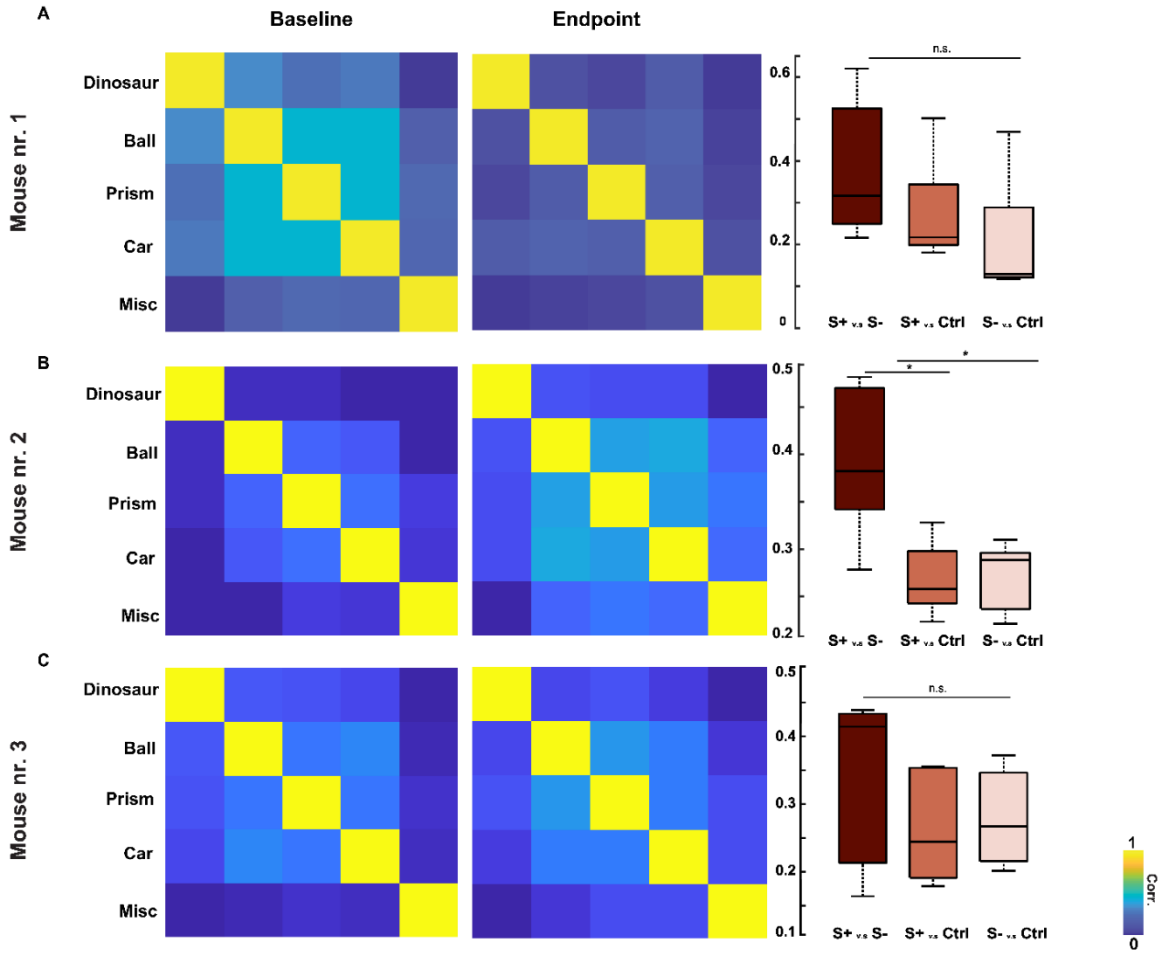

Figure 3.7. Similarity between response vectors during stimulus presentation for each category. Pearson's correlation matrix for each animal (**A**) corresponding to the baseline and endpoint sessions respectively. **B)** Between category similarity of response vectors for each animal.

Sparseness

The measure of population sparseness used in this analysis is based on the one proposed by Treves and Rolls (1991), where the sparseness of the code increases as the values approach 0.

As we can see in figure 3.8 (A), it seems that there is no difference in the degree of population sparseness between the different categories, not even when grouped based on their valence, regardless of how low the overall values are (mice = 3: Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (4)= 1.03, p = 0.9051; Kruskal-Wallis one-way ANOVA on ranks: $\chi 2$ (2)= 0.85, p = 0.6525).

However, it should be noted that we observed the same trend in the previous recordings, including the baseline session, where the animals had no previous exposure to the virtual objects or the behavioural task.

I also decided to analyze the lifetime sparseness over time, by comparing the cumulative distribution function (CDF) of the first and the last imaging sessions during the 5 sec. of stimulus presentation for all mice. Even though when I take the results of every mouse, the CDF of the last session is larger than the one corresponding to the first session, which indicates sparser neuronal activity at a later time point (One-sided two-sample KS test: n = 378 neurons baseline; n = 458 neurons endpoint; n = 3 mice; K = 0.1325, p < 0.001), this is most likely biased due to the results obtained by comparing the baseline and endpoint distributions of mouse nr. 3 (One-sided two-sample KS test: n = 162 neurons baseline; n = 121 neurons endpoint; K = 0.202, p = 0.002) since there was no difference between the distributions of the other two mice (One-sided two-sample KS test; mouse nr. 1: n = 104 neurons baseline; n = 108 neurons endpoint; K = 0.1527, p = 0.076; mouse nr. 2: n = 112 neurons baseline; n = 256 neurons endpoint; K = 0.084, p = 0.315)

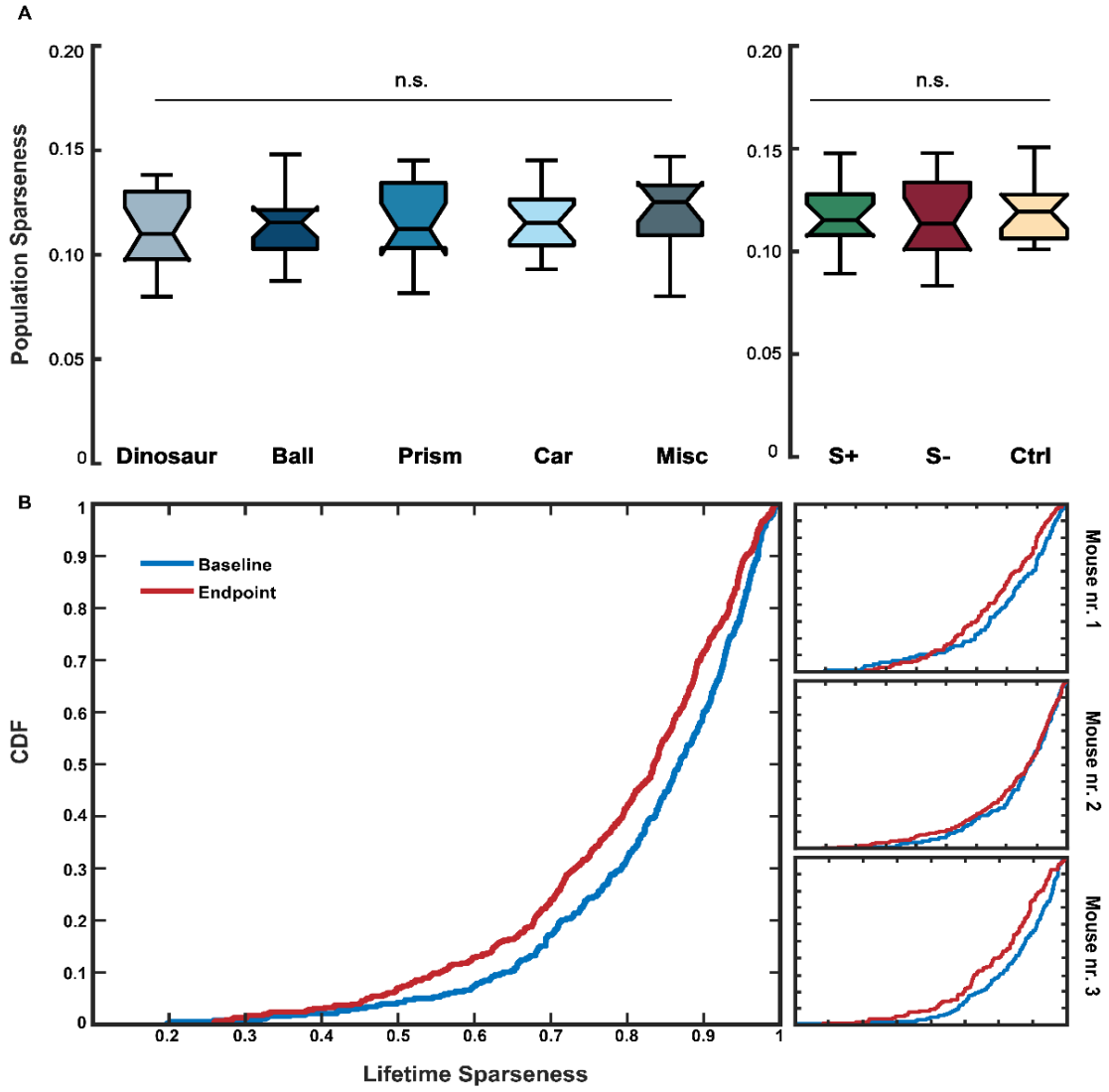

Figure 3.8. Population and lifetime sparseness doesn't increase with the acquisition of categorical knowledge in the mPFC. **A**) Population sparseness during stimulus presentation for the individual object categories in the last recording session (n = 3 mice). **B**) Population sparseness during stimulus presentation in the last recording session grouped by the S+, S- and Ctrl categories (n = 3 mice). Data reported in the same manner as in 3.3. **C**) Cumulative distribution function for lifetime sparseness corresponding to baseline (blue) and endpoint (red) sessions for all mice on the left, and for each individual animal on the right.

Discussion

The results of this study are insufficient to draw any conclusion about the learning process that took place during the behavioural categorization task, as I was unable to extract any information pertaining to the different categories, nor the specific objects being displayed on the screen. Furthermore, although a small portion of neurons seemed to display some preferences in terms of object categories, in most cases this activity preceded the onset of the visual stimulus, which doesn't reveal much in terms of category specificity. It is also worth mentioning that there was no detectable degree of selectivity for object exemplars, that is, a given neuron could present some degree of selectivity for "ball 1" in the first trial, but not when the same object appeared in trial number ten for example.

The Bayesian classifier used in order to create a probabilistic inference about the stimuli being presented also performed rather poorly, even when all but one trial was used as the training dataset. It should be noted that I tried different classifiers before using the one described in this chapter, but their accuracy was even lower.

There was no tangible difference in terms of how similar or dissimilar the neural activity was when the animals were viewing different category exemplars. In fact, it's fair to say that if there was no information about any undergoing learning process (*i.e.* the categorization task), I couldn't tell the difference in terms of neural activity between the different imaging sessions.

The sparseness of the neural code was also accessed both at the population level (population sparseness) and single neuron level (lifetime sparseness), and based on the results reported here, the population sparseness was quite low, even after the animals had been exposed to several object exemplars. Given that I observed similarly low values for sparseness (*i.e* a sparser population code) in previous recording sessions, this can lead to the conclusion that the activity in

the mPFC is probably sparse to begin with, an effect which might be even more amplified due to some recording noise when imaging through the microprism, and also the fact that during the preprocessing phase, only the most active neurons are automatically detected with our current Suite2P pipeline. This essentially means that the low number of active neurons is most likely a consequence of both the recording and preprocessing procedures and not a direct consequence of any particular experimental effect, especially since there's no tangible difference between the different categories.

As for the neuron's lifetime sparseness, I did observe a statistically significant difference between the first and last imaging sessions when I combine the data from the 3 mice. However, when analyzed individually, only one mouse showed a significant difference in terms of the baseline and endpoint CDF. It's also worth mentioning that when I conducted the same analysis, using the 5sec. window that preceded the object presentation, the difference between the 2 distributions was still present, which might indicate that the stimulus presentation doesn't have any effect on the neuronal responses *per se*.

It is unlikely that different methods to analyse the data would've yielded different results. In fact before the data analysis methods described in this chapter were implemented, I tried different approaches that had been reported in previous publications and the results were arguably even more lackluster.

On my first approach, I tried to use the PCA analysis method described by Lopes-dos-Santos et al, (2011), where the authors used the theoretical bounds of the Marcenko-Pastur distribution in order to detect neuron ensembles (Lopes-dos-Santos, Conde-Ocazionez, Nicolelis, Ribeiro, & Tort, 2011). Even though this method was applied to multielectrode recordings in rats, it seemed that it could be applicable to calcium deconvolved traces by simply converting the

deconvolved matrix into a binary one. The main purpose behind this idea was the detection of neuron ensembles that could hypothetically be present when the animals were seeing objects belonging to the same category. However, the results obtained revealed a handful of small ensembles sometimes comprised of two or three neurons, many of them not related to any particular object (*i.e* when the same object appeared more than once, the neuron ensemble was not detectable) or object category. For this reason the method was discarded.

I also tried to address the clustering of different category exemplars using an agglomerative hierarchical clustering algorithm, similar to what McClelland et al., (1995), Kiani et al., (2007) and Kriegeskorte et al., (2008) used in order to better visualize the relationship between the different categories. However, not only were the clusters obtained unrelated to category membership (*i.e* the population vectors for each trial for a given category were not clustered), but they also seemed to display the characteristics of a known clustering problem called "chaining". This problem is commonly observed when using single-link (shortest distance) as the linkage method, but the case of the dataset used here it was apparent even when different methods such as complete–link (furthest distance) or average link were used. Since the distance metric used in this dataset was based on Pearson's correlation (one minus the correlation between data points) and not on Euclidean distance, other linkage methods such as Ward's method were not employed. This being said, it is unlikely that different results would've been obtained by implementing these slight modifications.

All things considered, based on these results one cannot make any assertion about the role of the rodent mPFC in visual categorization. However, I believe that there are two major factors that can account for these lackluster results.

First, we should consider the context where learning occurred. The animals were trained in an operant conditioning box with a touchscreen tablet where the objects were displayed, which means that the learning process related to the virtual objects, as well as the nature of the pairwise discrimination task itself, become associated with a set of contextual cues present at the time of learning.

It's widely recognized that animals form spontaneous associations between objects and the context in which they are found (Barker & Warburton, 2020). This behavioural mechanism is thought to be supported by a network comprising several regions, chief among which are the HPC, PRh and PFC, with the latter being associated with top-down modulation through the PRh and lateral entorhinal cortex (lEC), of the hippocampal context-appropriate object representations (Bar, 2004; Eichenbaum, 2017; Fenske, Aminoff, Gronau, & Bar, 2006). Recent studies have also shown that action-selective mPFC neurons display a high degree of context-dependent modulation, and suggest that the mPFC is responsible for creating a rich contextual representation that incorporates sensory cues as well as specific actions and even time (Hyman, Ma, Balaguer-Ballester, Durstewitz, & Seamans, 2012).

Concurrently, perturbations in the PFC's activity via muscimol injections can also negatively affect task performance in by hindering the flexibility of context-appropriate responses towards ambiguous objects (Lee & Lee, 2012). Similar results have been observed when bilateral PFC lesions in rats impaired the performance in both object-context and object-place-context tasks (Barker & Warburton, 2020). Equally relevant, given the design of our behavioural task, is the fact that neurons in the PFC encode for context-appropriate behavioural initiation during reward seeking, which is based on the response-outcome contingency of the task (Moorman & Aston-Jones, 2015).

This means that, in the experiments described in the second chapter, there is a clear problem when it comes to the overall experimental setup. The animals were trained and imaged under very different circumstances, and with different sources of sensory stimulation which are inherent to the two experimental contexts. This can be quite problematic when it comes to assessing the network changes which derive from the cumulative experience of different category exemplars, since the contextual cues present at the time of learning were absent during the imaging phase, when the animals were tested.

Second, one should consider the imaging setup as well. Whereas the behavioural task required a specific set of actions to be taken in order to learn the similarities between category exemplars, the imaging/testing phase didn't require any action. The animals were passively viewing different objects appearing on the tablet's screen while being head-fixed. This becomes quite problematic when it comes to the mPFC's engagement in the categorization task.

In one of the studies conducted by Freedman et al., (2003), the group decided to compare the response patterns in both ITC and lPFC using the already mentioned DMC task (Freedman et al., 2003). Interestingly, they observed that when the visual stimuli were presented in a task which didn't require any input from the animals, the response patterns observed in the lPFC during the DMC task completely disappeared. This stands in sharp contrast to what was found in purely visual areas such as the ITC, as demonstrated by different studies (Dehaqani et al., 2016; Freedman et al., 2003; Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Lehky et al., 2014; Meyers et al., 2008).

Ultimately, this finding seems to indicate that the category information encoded by lPFC is heavily dependent on task demands, and requires some level of engagement and goal oriented

behaviour. I will further elaborate on this topic when revisiting the role of the PFC in visual categorization in the next chapter.

Chapter 4

General Discussion

The purpose of this set of experiments was to understand the role of the mPFC in the process of visual categorization, and in particular, if two of the main findings described in a conceptual learning model proposed by McClelland et al. (1995), could be observed in biological systems (McClelland et al., 1995). In order to assess this, I decided to use mice as our animal model, and to develop a behavioural categorization task that would allow us to evaluate the ability of these animals to generate categorical representations of different virtual objects. The main purpose of this task was to allow for the gradual incorporation of different objects belonging to two distinct categories into their knowledge database.

I then decided to evaluate the animal's performance with two different testing sets, the first with completely new objects from the initial categories (stage 7) and the second with a completely new S- category, or Ctrl category (stage 8). The results showed that the mice were able to generalize between objects of the S+ category as learning progressed, and even when tested with a complete new set of S+ and S- objects (stage 7) the correct response rate was remarkably high. Lastly, in the second testing stage (stage 8), the initial S- category was replaced by a new one, with 5 new objects. Even though there was a decrease in the performance, to levels comparable to the initial stage of training (stage 5), the mice gradually adapted to this new S- category and after a few sessions they reached the passing criteria (80% correct trials for two consecutive sessions). Thus, the data indicates that mice are able to form visual category representations, presumably based on sets of features that humans are also likely also to use; however, a possible caveat is that the mice might be using low level features such as aspect ratio in order to discriminate between the objects. One can speculate that there isn't enough variability between exemplars, and therefore

an abstract representation based on pixel variability in one or more axes could account for the results obtained, especially given the limited visual acuity observed in mice. A possible solution for this problem, is to use abstract objects such as “Greebles” or “Geons” which have been used by different groups as mentioned in Chapter I.

In the next set of experiments, I took a subset of the animals used in the behavioural task and examined the neural activity in the mPFC using in-vivo 2-photon calcium imaging. Our goal was to observe the neural dynamics in the mPFC, as the animals gradually learned to discriminate between the S+ and S- categories with different exemplars throughout the different stages of the behavioural task. However, as mentioned in the previous chapter, even though the three mice used in these experiments were able to reach the end of the categorization task, the analysis of the neural activity collected at different time points didn’t show any particular difference across sessions. And despite finding a very small set of neurons in each session which seemed to display some category preference, I also couldn’t find any category information in the population code.

It would also be reasonable to assume that the results reported here, simply reflect the fact that the mouse mPFC, unlike the primate dlPFC, does not play a role in encoding category information. However, given the perceived flaws in the experimental setup used in these experiments, such conclusion cannot be drawn, at least until these issues are properly addressed.

But even though I couldn’t provide a satisfactory answer to the questions I had initially proposed, I can find in the literature a glimpse of what the neural dynamics in the primate lPFC look like, and acquire some valuable insights about our two main hypotheses.

Hypothesis 1: *Neural representations of related experiences share a high degree of similarity between them and become hierarchically clustered.*

Building on previous experiments performed in their lab, Cromer et al., (2012) decided to investigate which neurons acted as multitasking or category generalists, and how many were category specialists (Cromer et al., 2010). Using a variation of the DMC task, the group added another category distinction (coupes versus sedans) to the one used in previous studies (cats versus dogs). This study was already mentioned in the general introduction of this thesis, but in short, a multitasking/generalist neuron would respond to both animals and cars (e.g. responding to dogs and sedans) during the delay period, whereas a specialist neurons was expected to respond just to one category scheme (either animals or cars). The results of this new DMC task revealed that a significant portion of dlPFC neurons were selective for both category schemes, and that these neurons were also the most category sensitive. There is of course, a sharp contrast between these results and the findings previously reported by the same group, where by Roy et al., (2010), used a pool of cat and dog images comprised of two prototypes for each category, as opposed to just one; this allowed the experimenters to create two different category schemes using the same two categories (Roy et al., 2010). In this variation of the task, Roy and collaborators found that there was very little overlap in terms of category representations across lPFC neurons. In other words, if the competing categories have very similar exemplars, and therefore are more prone to be misclassified, the lPFC can employ a more orthogonal coding scheme in order to increase pattern separation and therefore minimize uncertainty. This essentially means that, based on the task demands, the PFC might employ a different coding scheme in order to distinguish between different categories.

On the other hand, when objects being categorized belong to two independent sets such as in Cromer et al., (2012), with a larger degree of separation terms of their overall features, the PFC can reuse the same pool of neurons in order to represent categorical information, which leads to more overlap between category representations across neurons. Thus, the difficulty of the task seems to play a bigger role in terms of determining how information is encoded in the lPFC; contrasting with a pure bottom up signal processing approach, which would lead us assume that the more similar the stimuli are, the more correlated their representations would be (Kiani et al., 2007; Kriegeskorte et al., 2008). That is to say, neuronal specialization in the PFC occurs when the cognitive demands are high (e.g. using the same set of images with two different category schemes), whereas neuronal multitasking prevails when there is a marginal independence between the different categories.

This invariably leads us to one of our main hypotheses, which was derived from findings reported by McClelland et al., (1995) in their CLST model. In their paper, the authors hypothesized that representations of similar concepts would be highly correlated as the network acquires more information and is able to extract the statistical regularities across different categories.

If we analyse the aforementioned results from Miller's lab (Cromer et al., 2010; Roy et al., 2010) under this assumption, one can conclude that for a region such as the rhesus monkeys lPFC, this might not be the case. However, it should be noted that the CLST model didn't make any predictions regarding specific cortical regions where such clustering between representations of similar concepts (or categories) would exist. Furthermore, the results reported by Kiani et al., (2007) and Kriegeskorte et al.,(2008) seem to align with the predictions made by CLST model. A possible explanation for this might be the fact that the ITC is the last purely visual area in the ventral visual stream, which encodes categories based on their shared features, whereas the PFC

receives inputs from an array of cortical as well as subcortical regions, which need to be weighed according to the task parameters

Hypothesis 2: *As experiences of certain stimuli become integrated, mice will form sparse conjunctive representations of specific stimuli in higher modules of the hierarchy.*

As we've seen in the previous chapter, based on the data collected in experiment II, one cannot conclude that the representations in the mouse mPFC, display any of the characteristics of a sparse coding scheme, and this applies to both population sparseness and lifetime sparseness.

However, similarly to what I've described in regard to our first hypothesis, experiments conducted in Millers' laboratory can provide some valuable insights into how category information is encoded in the homologous region of the rodent mPFC – the primate lPFC.

In 2008, Meyers and collaborators also examined the coding schemes in both LPFC and ITC, to see if information was represented in a sparse/compact[5] fashion or in a more distributed manner (Meyers et al., 2008). In order to do this, the authors trained a classifier with the best "k" neurons in each region (lPFC and ITC) and then tested the classifier using only these neurons. They defined the best "k" neurons as those with the smallest p-values on a t-test that was applied to all the dog trials versus all of the cat trials on their training dataset; this procedure was also conducted separately for each time bin they used. Out of the 256 neurons recorded, the authors were able to extract most of the information available for both ITC and PFC using the best 16 neurons at all time points. In addition, during the decision period, they were able to retrieve most

[5] In their paper, the authors used the term "compact" or "compactness" due to the strong association between firing rate and sparseness. This might also be related to the fact that in many publications, the concepts of population sparseness and lifetime sparseness are sometimes used interchangeably or even combined (Meyers et al., 2008; Willmore & Tolhurst, 2001).

of the information from the lPFC using only 8 neurons, with a decoding accuracy of 78.2% +/- 1.2%. Remarkably, these neurons contained almost as much information as the whole population (79.4% +/- 1.7%). On the other hand, in the ITC the code was less sparse, with 64 neurons containing nearly all of the information available. Equally relevant, is the fact that excluding these 64 neurons, didn't lead to chance level accuracy by the classifier, a finding which point towards a considerable amount of redundant information contained in the remaining 192 neurons.

Nevertheless, the same laboratory, in 2010 equated multitasking neurons with a less sparse, and more distributed coding (Cromer et al., 2010). This apparent mismatch between what both papers claim can perhaps be attributed to the common misunderstanding between population sparseness and lifetime or kurtotic sparseness (for an extensive review see Wilmore and Tolhurst, 2001). While Meyers et al., (2008) point towards a population sparseness definition, Cromer et al., (2010) might be alluding to something more akin to lifetime sparseness. However, this interpretation is merely speculative, since the authors never clarified in any publication these seemingly contradictory statements.

Taken together these findings imply an extremely sparse coding in the monkey lPFC during the decision period, and an overall sparse representation in both regions during the three periods of the task. Nevertheless, I cannot infer anything about the evolution of the network's coding scheme as learning progressed, since the recordings were conducted after the extensive training period.

1. Visual categorization: revisiting the role of the Prefrontal Cortex

Neural correlates of visual categorization have been reported in several brain regions. This complex process of partitioning the visual elements of the external world according to their visual

similarity and/or their behavioural significance, relies on several of brain areas. These areas range from primary sensory cortex, in particular the visual cortices which process increasingly complex features of a given stimuli, to the hippocampus, and of course higher order processing regions such as the parietal cortex, inferior temporal cortex and the prefrontal cortex (Pan & Sakagami, 2012; Seger & Miller, 2010). Moreover, regions associated with stimulus valence such as the striatum and the midbrain dopaminergic neurons also collaborate in the process of category learning, which is often comprised of a decision making component as well (Antzoulatos & Miller, 2011; Seger & Miller, 2010; Seger & Peterson, 2013). All things considered, both the IT cortex and the PFC, in particular the lPFC, seem to play major roles when it comes to category learning.

As mentioned in the introduction, the ITC is not only an important region for object recognition, with neurons exhibiting tolerance for several viewpoints of the same object (also known as view invariance), but its also heavily involved in the categorization of visual stimuli (Freedman et al., 2003; Kiani et al., 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Nikos K. Logothetis & Sheinberg, 1996; Vogels, 1999a, 1999b).

Vogels and collaborators (1999a, 199b) showed that monkeys were capable of discriminating tree from non-tree images in a classification task, and that the ITC neurons were actively encoding categorical information pertaining to the tree category (Vogels, 1999a, 1999b). The studies mentioned earlier in the general introduction, by Kiani et al., (2007) and Kriegeskorte et al., (2008) also revealed that while viewing a set of more than 1000 images, the population activity in the monkey's ITC showcased a robust ability to classify images based on feature similarity. But even though the neurons in the ITC were capable of encoding category information, no single neuron responded to all exemplars of a given category in a similar manner. For example, many neurons which were responsive to human faces were mostly silent when the faces of other

primates were shown, and even those which showed selectivity for to human faces, didn't respond to all human faces.

Meyers and collaborators analysed the responses of 443 ITC and 525 lPFC neurons from two rhesus monkeys, from a set of previous experiments conducted by Freedman and collaborators (Freedman et al., 2003; Meyers et al., 2008). The researchers found that, in general, even though the ITC neurons encoded abstract category information, for the most part they seemed to encode detailed visual information. These results echoed the findings reported by Freedman et al., (2003), where the data suggested a greater involvement of the ITC in encoding the properties of images currently being viewed.

On the other hand, the lPFC neurons seemed to attribute more weight to the behavioural relevance of the stimuli, given the current task demands, and in storing such information in working memory (Meyers, 2018; Meyers et al., 2008; Pan & Sakagami, 2012).

Using a linear classifier, Meyers et al., (2008) managed to get a glimpse of how much detailed visual information was preserved regardless of the spike count variability from trial to trial, and between the different phases of the behavioural task. Interestingly, they observed that in both ITC and PFC, the information about the category of a stimulus presented during the sample phase (i.e. cat or dog), seemed to increase immediately before the onset of the decision phase, where the monkeys had to make a decision about the category membership of the sampled stimulus. Its also important to note that while neurons along the ITC possessed category information in each phase of the task (sample, delay and decision phases), the effect related to the overall ratio of abstract information relative to total category information was much more pronounced in the lPFC. Furthermore, Freedman et al., (2003) had already observed that before the initial training on the contingency, the neurons in the ITC didn't show any ability to

discriminate between images close to the category boundary. Only after several sessions, did the neurons in the ITC start to display a sharper tuning for the specific features of the stimuli, and consequently, to the distance between the morphed images near the category boundary, which contrasted once again, with the lPFC neurons which were capable of representing the category boundary between the 2 categories.

Another interesting finding from Freedman's experiment was the fact that the category signal observed in the lPFC could be observed earlier in the task (sampling phase), compared to the ITC, in which category selectivity was only observed during the delay and decision phase. According to Pan and Sakagami (2010), this could mean that perhaps, the category information related to specific object features and the perceptual commonalities between them, could be sent from the ITC to the lPFC, which would then extract category information based on the current task demands and motivational state, and send that information back to the ITC (Pan & Sakagami, 2012; Sakagami et al., 2006; Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999). In spite of that, the importance of the lPFC in visual categorization was challenged by another group.

In a study published in 2010, Minamimoto et al., decided to see if the lPFC was actually necessary for the process of visual categorization (Minamimoto, Saunders, & Richmond, 2010). In humans, category specific visual agnosia has been observed in patients with ITC damage, but not lPFC damage (Gainotti, 2000; Minamimoto et al., 2010). This led Minamimoto and collaborators to test the role of the lPFC in the process of visual categorization, using a modified version of a reward-delay task, which had been previously developed by the same group (Minamimoto, La Camera, & Richmond, 2009). In this experiment, four monkeys were trained to associate a given category (dogs) with a higher amount of reward and another one (cats) with a smaller amount, and after four days of testing with these two categories, the animals were given a

bilateral lPFC lesion. The results showed that not only could the monkeys perform the same task after the bilateral lesion, using the same set of images, but could also learn to categorize novel images from the previous categories and even learn 2 new categorical distinctions (cars versus trucks). These results are diametrically opposed to the evidence accumulated over the years by other groups (Freedman et al., 2001, 2002, 2003; Jiang et al., 2007; Meyers et al., 2008), and they raise a lot of questions related to the process of stimulus generalization, and exactly what regions are essential for the process of visual categorization.

A possible explanation for the results reported by Minamimoto et al., (2010), was proposed by Pan and Sakagami in 2012. In line with some of the main points outlined in a previous review paper by Buckley and Sigala (2010), the authors raised the possibility that the task used in Mimaminoto's study could be accomplished without the involvement of the lPFC since it might be less-demanding in comparison to the type of tasks reported in previous studies (Buckley & Sigala, 2010; Pan & Sakagami, 2012). The main argument seems to revolve around the fact that the reward-delayed task used by Minamimoto and collaborators relies on a fast perceptual mechanism, which might be accomplished in the absence of the lPFC and its top-down influence, either by using different pathways or by purely relying on visual cues, which could theoretically be accomplished by the ITC alone.

In 2009, Peelen and collaborators conducted an fMRi study, where they showed that, in a rapid categorization task, the PFC appeared to be silent, whereas the ITC was actively engaged when it came to detect the presence of people or cars in natural scenes (Peelen, Fei-Fei, & Kastner, 2009). A possible explanation for these results lies in the fact that the type of fast visual processing required could be achieved via ITC and OFC connections without the involvement of the PFC. (Buckley & Sigala, 2010). Lastly, one should also consider that the lPFC lesioned monkeys might

find more complex tasks such as the ones which make use of morphed images, more difficult to learn, but for now we are still lacking the experimental evidence to make such assertion.

All thing considered, the important role played by the lPFC in categorization has been demonstrated by the unique ability of its constituent neurons to represent the boundary between different categories in a morphing visual categorization task (Freedman et al., 2001, 2003; Meyers et al., 2008). Furthermore, it seems that the primate PFC as a whole is involved in representing not just perceptual categories, but also in grouping together objects or visual stimuli with no resemblance to each other, but which share the same context-dependent behavioural relevance (Pan & Sakagami, 2012; Seger & Miller, 2010). This means that the primate PFC and more precisely the lPFC, is capable of extracting important information across stimuli, while ignoring irrelevant dimensions, which can essentially be seen as a form of abstraction.

Given the aforementioned results it is fair to assume that the ability to abstract across stimuli might also rely on the lPFC's capacity to retain and/or access a long-term record of stimulus-reward associations. In turn, this feature can facilitate the emergence predictions about a positive or negative outcomes based on specific perceptual information, as well as the inputs stemming from the vPFC, OFC, HPC, AG, BG, STR, and even the dopaminergic afferents originating in the midbrain. Lastly, it's also important to note that neurons in the lPFC can be responsive not only to the category exemplars, but also to rewards, making them category-reward specific (Pan, Sawa, Tsuda, Tsukada, & Sakagami, 2008). This is particularly relevant for tasks that rely on any kind of reward-based learning in order to probe the subjects' ability to discriminate between different categories, since the type of category specificity observed at the neuronal level, or even the strength of the category response, might be directly linked to the existence of a reward signal that can amplify their ability to discriminate.

2. Conclusion and Future Directions

Visual Categorization is a complex process that has been the subject of intense research since the second half of the 1960's. For the most part researchers have focused on uncovering the mechanisms that might be in place when a categorical judgement is made, and the parameters which determine how things around us are categorized. But over the past 20 years, the focus within the field of Neuroscience has been in the specific regions involved in this complex process that relies on many different brain regions. The PFC has been particularly important in terms of providing insights about this process. This is mostly due to its unique connectivity with many cortical as well as subcortical regions which provide an impressive array of inputs that are then combined and reweighted according to specific goals.

But although a considerable number of physiological studies have been conducted in human and non-human primates, there is much less information about the specific mechanisms behind visual categorization in other animal species. With the advancements in recording techniques developed for rodents, such as 2-photon calcium imaging, it made sense for us to explore the role of the mouse mPFC in this process, especially given the complete lack of studies that could bridge the gap between homologous regions between species and their specific role in categorization. And even though the 2-photon calcium imaging results reported here do not bring us closer to any meaningful understanding about the role of the rodent mPFC in visual categorization, I consider that this has been a valuable experience. The experiments reported in this thesis were essentially a first attempt at not only imaging the mPFC in mice (within our lab), but more broadly, at examining the neural mechanisms of category learning, and how the network dynamics change as more categorical knowledge is acquired over time. There were a lot of things that needed to be in place in terms of the overall logistics required to conduct these experiments,

and for a first attempt, I did gain some valuable knowledge, even in the absence of evidence for any of our specific hypotheses.

There is still a lot to be done within the larger field of visual categorization, particularly when it comes to rodent research. Many important areas such as the HPC, the AG, the PRh, or even the different sub-regions of the PFC are still unknown variables when it comes to determining their contribution to this complex process and what sorts of information is represented within those areas.

It's quite possible that different areas weight the inputs they receive based on parameters such as purely visual information, previous encounters with a given object or set of objects as well any emotional valence associated with them, which might ultimately dictate how category information is represented in those same areas. This will require tasks that not only evaluate some of these different parameters but also recording techniques that can access different brain regions, ideally in a simultaneous manner similar to the ITC and PFC recordings mentioned throughout this thesis. This being said, one of the major shortcomings of those studies, lies precisely on the lack of information on how the network adapts its weight matrix as learning progresses, and as the subjects are exposed to more exemplars of the same categories. This is what will ultimately allow us to uncover how this form of semantic knowledge, how schematic representations arise, and therefore, how the brain learns to create a rich model of the world.

REFERENCES

Agster, K. L., & Burwell, R. D. (2009). Cortical Efferents of the Perirhinal , Postrhinal , and Entorhinal Cortices of the Rat, *1186*(April), 1159–1186. https://doi.org/10.1002/hipo.20578

Agster, K. L., & Burwell, R. D. (2013). Hippocampal and subicular efferents and afferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. *Behavioural Brain Research*, *254*, 50–64. https://doi.org/10.1016/j.bbr.2013.07.005

Andermann, M. L., Kerlin, A. M., & Reid, R. C. (2010). Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front Cell Neurosci*, *4*(March), 3. https://doi.org/10.3389/fncel.2010.00003

Anderson, M. C., Bunce, J. G., & Barbas, H. (2016). Prefrontal–hippocampal pathways underlying inhibitory control over memory. *Neurobiology of Learning and Memory*, *134*, 145–161. https://doi.org/10.1016/j.nlm.2015.11.008

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

Antzoulatos, E. G., & Miller, E. K. (2014). Article Increases in Functional Connectivity between Prefrontal Cortex and Striatum during Category Learning. *Neuron*, *83*(1), 216–225. https://doi.org/10.1016/j.neuron.2014.05.005

Artal, P., De Tejada, P. H., Tedó, C. M., & Green, D. G. (1998). Retinal image quality in the rodent eye. *Visual Neuroscience*, *15*(4), 597–605. https://doi.org/10.1017/S0952523898154020

Ashby, F. G., & Maddox, W. T. (1994). A Response Time Theory of Separability and Integrality in Speeded Classification. *Journal of Mathematical Psychology*. https://doi.org/10.1006/jmps.1994.1032

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Balkema, G. W., & Pinto, L. H. (1982). Electrophysiology of retinal ganglion cells in the mouse: A study of a normally pigmented mouse and a congenic hypopigmentation mutant, pearl. *Journal of Neurophysiology*, *48*(4), 968–980. https://doi.org/10.1152/jn.1982.48.4.968

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Barker, G. R. I., & Warburton, E. C. (2020). Putting objects in context: A prefrontal–hippocampal–perirhinal cortex network. *Brain and Neuroscience Advances*, *4*, 239821282093762. https://doi.org/10.1177/2398212820937621

Barlow, H. B. (1972). Single units and sensations: a neuron doctrin for perceptual psychology? *Perception*, *1*, 371–394. https://doi.org/10.1068/p010371

Bartlett, F. C. (1932). REMEMBERING: A STUDY IN EXPERIMENTAL AND SOCIAL

PSYCHOLOGY. In *Remembering: A Study in Experimental and Social Psychology* (Vol. 3, pp. 197–214). Cambridge, MA: Cambridge University Press. https://doi.org/https://doi.org/10.1111/j.2044-8279.1933.tb02913.x

Benard, J., Stach, S., & Giurfa, M. (2006). Categorization of visual stimuli in the honeybee Apis mellifera. *Animal Cognition*, *9*(4), 257–270. https://doi.org/10.1007/s10071-006-0032-9

Bero, A. W., Meng, J., Cho, S., Shen, A. H., Canter, R. G., Ericsson, M., & Tsai, L.-H. (2014). Early remodeling of the neocortex upon episodic memory encoding. *Proceedings of the National Academy of Sciences*, *111*(32), 11852–11857. https://doi.org/10.1073/pnas.1408378111

Biederman, I. (1987). A Theory of Human Image Understanding. *Psychological Review*, *94*(2), 115–147. https://doi.org/10.1038/nprot.2008.36

Bonin, V., Histed, M. H., Yurgenson, S., & Clay Reid, R. (2011). Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *Journal of Neuroscience*, *31*(50), 18506–18521. https://doi.org/10.1523/JNEUROSCI.2974-11.2011

Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, *8*(6), 510–523. https://doi.org/10.1093/cercor/8.6.510

Brod, G., Lindenberger, U., Werkle-Bergner, M., & Shing, Y. L. (2015). Differences in the neural signature of remembering schema-congruent and schema-incongruent events. *NeuroImage*, *117*, 358–366. https://doi.org/10.1016/j.neuroimage.2015.05.086

Brodmann, K. (1909). *Localization in the cerebral Cortex: The Principles of Comparative Localization in the Cerebral Cortex Based on Cytoarchitectonics*. Leipzig: Verlag von Johan Ambrosius Barth.

Brooks, D. I., Ng, K. H., Buss, E. W., Marshall, A. T., Freeman, J. H., & Wasserman, E. A. (2013). Categorization of photographic images by rats using shape-Based image dimensions. *Journal of Experimental Psychology: Animal Behavior Processes*, *39*(1), 85–92. https://doi.org/10.1037/a0030404

Brown, M. W., & Aggleton, J. P. (2001). Recognition Memory: What Are the Roles of the Perirhinal Cortex and Hippocampus. *Nature Neuroscience Reviews*, *2*(January). Retrieved from papers2://publication/uuid/FEDB8C78-DD9B-401A-91EB-A26D05EEC079

Buckley, M. J., & Sigala, N. (2010). Is top-down control from prefrontal cortex necessary for visual categorization? *Neuron*, *66*(4), 471–473. https://doi.org/10.1016/j.neuron.2010.05.012

Burwell, R. D. (2001). Borders and Cytoarchitecture of the Perirhinal and Postrhinal Cortices in the Rat, *41*(May), 17–41.

Burwell, R. D., & Amaral, D. G. (1998). Cortical Afferents of the Perirhinal , Postrhinal , and Entorhinal Cortices, *205*(November 1997), 179–205.

Busemeyer, J. R., & Myung, I. J. (1992). An Adaptive Approach to Human Decision Making: Learning Theory, Decision Theory, and Human Performance. *Journal of Experimental*

Psychology: General, *121*(2), 177–194. https://doi.org/10.1037/0096-3445.121.2.177

Bussey, T. J., Muir, J. L., Everitt, B. J., & Robbins, T. W. (1997). Triple dissociation of anterior cingulate, posterior cingulate, and medial frontal cortices on visual discrimination tasks using a touchscreen testing procedure for the rat. *Behavioral Neuroscience*, *111*(5), 920–936. https://doi.org/10.1037/0735-7044.111.5.920

Carmichael, S., & Price, J. L. (1994). Architectonic Subdivision of the Orbital and Medial Prefrontal Cortex in the Macaque Monkey. *Journal of Comparative Neurology*, *402*, 346–366. https://doi.org/10.1002/cne.10609

Carr, David B., & Sesack, S. R. (2000). Projections from the rat prefrontal cortex to the ventral tegmental area: Target specificity in the synaptic associations with mesoaccumbens and mesocortical neurons. *Journal of Neuroscience*, *20*(10), 3864–3873. https://doi.org/10.1523/jneurosci.20-10-03864.2000

Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory and Cognition*, *40*(3), 434–449. https://doi.org/10.3758/s13421-011-0154-4

Chang, H. R., Esteves, I. M., Neumann, A. R., Sun, J., Mohajerani, M. H., & McNaughton, B. L. (2020). Coordinated activities of retrosplenial ensembles during resting-state encode spatial landmarks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1799). https://doi.org/10.1098/rstb.2019.0228

Cloke, J. M., Jacklin, D. L., & Winters, B. D. (2015). The neural bases of crossmodal object recognition in non-human primates and rodents: A review. *Behavioural Brain Research*, *285*, 118–130. https://doi.org/10.1016/j.bbr.2014.09.039

Cohen, H., & Lefebvre, C. (2005). *Handbook of Categorization in Cognitive Science*. (H. Cohen & C. Lefebvre, Eds.), *Handbook of Categorization in Cognitive Science* (1st ed.). Oxford: Elsevier.

Collins, A. M., & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, *247*(8), 240–247. https://doi.org/https://doi.org/10.1016/S0022-5371(69)80069-1

Collins, A. M., & Quillian, M. R. (1970). Facilitating retrieval from semantic memory: the effect of repeating part of an inference. *Acta Psychologica*, *33*, 304–314. https://doi.org/https://doi.org/10.1016/0001-6918(70)90142-3

Conaway, N., & Kurtz, K. J. (2017). Similar to the category, but not the exemplars: A study of generalization. *Psychonomic Bulletin and Review*, *24*(4), 1312–1323. https://doi.org/10.3758/s13423-016-1208-1

Condé, F., Maire-lepoivre, E., Audinat, E., & Crépel, F. (1995). Afferent connections of the medial frontal cortex of the rat. II. Cortical and subcortical afferents. *Journal of Comparative Neurology*, *352*(4), 567–593. https://doi.org/10.1002/cne.903520407

Coogan, T. A., & Burkhalter, A. (1993). Hierarchical organization of areas in rat visual cortex. *Journal of Neuroscience*, *13*(9), 3749–3772. https://doi.org/10.1523/jneurosci.13-09-03749.1993

Cook, R. G., & Smith, J. D. (2006). Stages of abstraction and exemplar memorization in pigeon category learning. *Psychological Science*, *17*(12), 1059–1067. https://doi.org/10.1111/j.1467-9280.2006.01833.x

Cowen, S. L., & McNaughton, B. L. (2007). Selective delay activity in the medial prefrontal cortex of the rat: contribution of sensorimotor information and contingency. *Journal of Neurophysiology*, *98*(May 2007), 303–316. https://doi.org/10.1152/jn.00150.2007

Creighton, S. D., Collett, H. A., Zonneveld, P. M., Pandit, R. A., Huff, A. E., Jardine, K. H., … Winters, B. D. (2019). Development of an "Object Category Recognition" Task for Mice: Involvement of Muscarinic Acetylcholine Receptors. *Behavioral Neuroscience*, *133*(5), 527–536. https://doi.org/10.1037/bne0000331

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. *Neuron*, *66*(5), 796–807. https://doi.org/10.1016/j.neuron.2010.05.005

Curby, K. M., Hayward, W. G., & Gauthier, I. (2004). Laterality effects in the recognition of depth-rotated novel objects. *Cognitive, Affective and Behavioral Neuroscience*, *4*(1), 100–111. https://doi.org/10.3758/CABN.4.1.100

Datiche, F., & Cattarelli, M. (1996). Reciprocal and topographic connections between the piriform and prefrontal cortices in the rat: A tracing study using the B subunit of the cholera toxin. *Brain Research Bulletin*, *41*(6), 391–398. https://doi.org/10.1016/S0361-9230(96)00082-2

Davis, T., & Poldrack, R. A. (2014). Quantifying the internal structure of categories using a neural typicality measure. *Cerebral Cortex*, *24*(7), 1720–1737. https://doi.org/10.1093/cercor/bht014

De Curtis, M., & Paré, D. (2004). The rhinal cortices: A wall of inhibition between the neocortex and the hippocampus. *Progress in Neurobiology*, *74*(2), 101–110. https://doi.org/10.1016/j.pneurobio.2004.08.005

Dehaqani, M. R. A., Vahabie, A. H., Kiani, R., Ahmadabadi, M. N., Araabi, B. N., & Esteky, H. (2016). Temporal dynamics of visual category representation in the macaque inferior temporal cortex. *Journal of Neurophysiology*, *116*(2), 587–601. https://doi.org/10.1152/jn.00018.2016

Desimone, R., Thomas, A., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience*, *4*(8), 2051–2062.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Duncan, J., & Miller, E. K. (2002). Cognitive Focus through Adaptive Neural Coding in the Primate Prefrontal Cortex. In D. T. Stuss & R. T. Knight (Eds.), *Principles of Frontal Lobe Function* (1st ed., Vol. 15, pp. 278–291). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195134971.001.0001

Eichenbaum, H. (2017). Prefrontal-hippocampal interactions in episodic memory. *Nature Reviews Neuroscience*, *18*(9), 547–558. https://doi.org/10.1038/nrn.2017.74

Ennaceur, A., & Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research*, *31*(1), 47–59. https://doi.org/10.1016/S0166-4328(05)80315-8

Erez, J., Cusack, R., Kendall, W., & Barense, M. D. (2016). Conjunctive Coding of Complex Object Features. *Cerebral Cortex*, *26*(5), 2271–2282. https://doi.org/10.1093/cercor/bhv081

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107–140. https://doi.org/10.1037//0096-3445.127.2.107

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, *18*(4), 500–549. https://doi.org/10.1016/0010-0285(86)90008-3

Esteves, I. M., Chang, H., Neumann, A. R., Sun, J., Mohajerani, M. H., & McNaughton, B. L. (2021). Spatial information encoding across multiple neocortical regions depends on an intact hippocampus. *Journal of Neuroscience*, *41*(2), 307–319. https://doi.org/10.1523/JNEUROSCI.1788-20.2020

Ettlinger, G. (1990). "Object Vision" and "Spatial Vision": The Neuropsychological Evidence for the Distinction. *Cortex*, *26*(3), 319–341. https://doi.org/10.1016/S0010-9452(13)80084-6

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The Role of Medial Prefrontal Cortex in Memory and Decision Making. *Neuron*, *76*(6), 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002

Euston, D. R., & McNaughton, B. L. (2006). Apparent encoding of sequential context in rat medial prefrontal cortex is accounted for by behavioral variability. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(51), 13143–13155. https://doi.org/10.1523/JNEUROSCI.3803-06.2006

Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, *318*(5853), 1147–1150. https://doi.org/318/5853/1147 [pii] 10.1126/science.1148979

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Progress in Brain Research*, *155 B*, 3–21. https://doi.org/10.1016/S0079-6123(06)55001-0

Ferrier, D. (1886). *Fucntions of the Brain* (2nd ed.). New York, NY: G.P Putnam's Sons.

Ferrier, D., & Yeo, G. (1884). The Effects of Lesions of Different Regions of the Cerebral Hemispheres. *Proceedings of the Royal Society of London*, (36), 222–224.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal Optical Society of America*, *4*(12). Retrieved from https://www.osapublishing.org/abstract.cfm?uri=josaa-4-12-2379

Földiak, P. (2002). Sparse coding in the primate cortex. *The Handbook of Brain Theory and Neural Networks*, (April 2002), 7. https://doi.org/10.1.1.16.1720

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). How category learning affects object representations: Not all morphspaces stretch alike. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(4), 807–820. https://doi.org/10.1037/a0025836

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*(4), 814–823. https://doi.org/10.1093/cercor/bhs067

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

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *Journal of Neurophysiology*, *88*(2), 929–941. https://doi.org/10.1152/jn.2002.88.2.929

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*(12), 5235–5246. https://doi.org/10.1523/jneurosci.23-12-05235.2003

Freiwald, W., & Tsao, D. (2010). Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System, *330*(April), 481–485. https://doi.org/10.5061/dryad.5t110.Supplementary

Furtak, S. C., Wei Shau-Ming, Agster, K. L., & Burwell, R. D. (2007). Functional Neuroanatomy of the Parahippocampal region in the Rat: The Perirhinal and Postrhinal Cortices. *Hippocampus*, *17*(9), 709–722. https://doi.org/10.1002/hipo.20314

Fuster, J. M. (2000). The Prefrontal Cortex of the Primate: A synopsis. *Psychobiology*, *28*(2), 125–131.

Fuster, J. M. (2001). The prefrontal cortex - An update: Time is of the essence. *Neuron*, *30*(2), 319–333. https://doi.org/10.1016/S0896-6273(01)00285-9

Gainotti, G. (2000). What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: A review. *Cortex*, *36*(4), 539–559. https://doi.org/10.1016/S0010-9452(08)70537-9

Gais, S., Albouy, G., Boly, M., Dang-Vu, T. T., Darsaud, A., Desseilles, M., … Peigneux, P. (2007). Sleep transforms the cerebral trace of declarative memories. *Proceedings of the National Academy of Sciences*, *104*(47), 18778–18783. https://doi.org/10.1073/pnas.0705454104

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Gauthier, I., & Tarr, M. J. (2016). Visual Object Recognition: Do We (Finally) Know More Now Than We Did? *Annual Review of Vision Science*, *2*(1), 377–396.

https://doi.org/10.1146/annurev-vision-111815-114621

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Glickfeld, L. L., Andermann, M. L., Bonin, V., & Reid, R. C. (2013). Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nature Neuroscience*, *16*(2), 219–226. https://doi.org/10.1038/nn.3300

Godsil, B. P., Kiss, J. P., Spedding, M., & Jay, T. M. (2013). The hippocampal-prefrontal pathway: The weak link in psychiatric disorders? *European Neuropsychopharmacology*, *23*(10), 1165–1181. https://doi.org/10.1016/j.euroneuro.2012.10.018

Goldman-Rakic, P. S. (1984). The frontal lobes: Uncharted provinces of the brain. *Trends in Neurosciences*, *7*(11). https://doi.org/10.1016/S0166-2236(84)80147-2

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116–139. https://doi.org/10.1037/0096-3445.130.1.116

Goodale, M., & Milner, D. (1992). Separate visual pathways for perception and action. *Literature and Theology*, *15*(1), 80–92. https://doi.org/https://doi.org/10.1016/0166-2236(92)90344-8

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://doi.org/10.1016/S0042-6989(01)00097-9

Graham, D. J., & Field, D. J. (2007). 3 . 14 Sparse Coding in the Neocortex. *Current Opinion in Neurobiology*, *III*(1972), 181–187.

Gross, C. G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia*, *46*(3), 841–852. https://doi.org/10.1016/j.neuropsychologia.2007.11.009

Güntürkün, O., Koenen, C., Iovine, F., Garland, A., & Pusch, R. (2018). The neuroscience of perceptual categorization in pigeons: A mechanistic hypothesis. *Learning and Behavior*, *46*(3), 229–241. https://doi.org/10.3758/s13420-018-0321-6

Gureckis, T. M., & Goldstone, R. L. (2008). The Effect of the Internal Structure of Categories on Perception. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, (January 2008). https://doi.org/10.4314/jlt.v44i2.71793

Hallock, H. L., Wang, A., & Griffin, A. L. (2016). Ventral midline thalamus is critical for hippocampal–prefrontal synchrony and spatial working memory. *Journal of Neuroscience*, *36*(32), 8372–8389. https://doi.org/10.1523/JNEUROSCI.0991-16.2016

Hauffen, K., Bart, E., Brady, M., Kersten, D., & Hegdé, J. (2012). Creating objects and object categories for studying perception and perceptual learning. *Journal of Visualized Experiments : JoVE*, (69), 1–10. https://doi.org/10.3791/3358

Hebscher, M., & Gilboa, A. (2016). A boost of confidence: The role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. *Neuropsychologia*, (1989), 1–

13. https://doi.org/10.1016/j.neuropsychologia.2016.05.003

Heidbreder, C. A., & Groenewegen, H. J. (2003). The medial prefrontal cortex in the rat: Evidence for a dorso-ventral distinction based upon functional and anatomical characteristics. *Neuroscience and Biobehavioral Reviews*, *27*(6), 555–579. https://doi.org/10.1016/j.neubiorev.2003.09.003

Hélie, S., Turner, B. O., & Cousineau, D. (2018). Can categorical knowledge be used in visual search? *Acta Psychologica*, *191*(September), 52–62. https://doi.org/10.1016/j.actpsy.2018.08.016

Hernandez, A. R., Reasor, J. E., Truckenbrod, L. M., Lubke, K. N., Johnson, S. A., Bizon, J. L., … Burke, S. N. (2017). Medial prefrontal-perirhinal cortical communication is necessary for flexible response selection. *Neurobiology of Learning and Memory*, *137*, 36–47. https://doi.org/10.1016/j.nlm.2016.10.012

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, *146*(3643), 549–551. https://doi.org/10.1126/science.146.3643.549

Hilgetag, C. C., & Goulas, A. (2016). Is the brain really a small-world network? *Brain Structure and Function*, *221*(4), 2361–2366. https://doi.org/10.1007/s00429-015-1035-6

Homa, D., & et al. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, *101*(1), 116–122. https://doi.org/10.1037/h0035772

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418–439. https://doi.org/10.1037/0278-7393.7.6.418

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning & Memory*, *2*(3), 322–330. https://doi.org/10.1037//0278-7393.2.3.322

Horner, A. E., Heath, C. J., Hvoslef-Eide, M., Kent, B. A., Kim, C. H., Nilsson, S. R. O., … Bussey, T. J. (2013). The touchscreen operant platform for testing learning and memory in rats and mice. *Nature Protocols*, *8*(10), 1961–1984. https://doi.org/10.1038/nprot.2013.122

Hubel, D. H., & Wiesel, T. N. (1959). Receptove fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*, 574–591.

Hubel, D., & Wiesel, T. (1964). Receptive Fields and Functional Architecture in two Nonstriate Visual Areas (18 and 19) of the Cat.

Huberman, A. D., & Niell, C. M. (2011). What can mice tell us about how vision works? *Trends in Neurosciences*, *34*(9), 464–473. https://doi.org/10.1016/j.tins.2011.07.002

Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D., & Seamans, J. K. (2012). Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(13), 5086–5091. https://doi.org/10.1073/pnas.1114415109

Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category representations in object-selective cortex. *NeuroImage*, *134*, 170–179. https://doi.org/10.1016/j.neuroimage.2016.04.012

Jarovi, J., Volle, J., Yu, X., Guan, L., & Takehara-Nishiuchi, K. (2018). Prefrontal theta oscillations promote selective encoding of behaviorally relevant events. *ENeuro*, *5*(6). https://doi.org/10.1523/ENEURO.0407-18.2018

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity. *Neuron*, *53*(6), 891–903. https://doi.org/10.1016/j.neuron.2007.02.015

Johnson, L. a, Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2010). Stored-trace reactivation in rat prefrontal cortex is correlated with down-to-up state fluctuation density. *The Journal of Neuroscience*, *30*(7), 2650–2661. https://doi.org/10.1523/JNEUROSCI.1617-09.2010

Karimi-Rouzbahani, H., Bagheri, N., & Ebrahimpour, R. (2017). Invariant object recognition is a personalized selection of invariant features in humans, not simply explained by hierarchical feed-forward vision models. *Scientific Reports*, *7*(1), 1–24. https://doi.org/10.1038/s41598-017-13756-8

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007a). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. https://doi.org/10.1152/jn.00024.2007

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007b). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. https://doi.org/10.1152/jn.00024.2007

Kim, M., Kwak, C., Yu, N., & Kaang, B. (2016). Optimization of the touchscreen paired-associate learning (PAL) task for mice and its dorsal hippocampal dependency. *Animal Cells and Systems*, *20*(5), 229–236. https://doi.org/10.1080/19768354.2016.1221855

Kitamura, T., Ogawa, S. K., Roy, D. S., Okuyama, T., Morrissey, M. D., Smith, L. M., … Tonegawa, S. (2017). Systems Consolidation of a Memory, *78*(April), 73–78. https://doi.org/10.1126/science.aam6808

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, *80*(1), 324–330. https://doi.org/10.1152/jn.1998.80.1.324

Komiyama, T., Sato, T. R., Oconnor, D. H., Zhang, Y. X., Huber, D., Hooks, B. M., … Svoboda, K. (2010). Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature*, *464*(7292), 1182–1186. https://doi.org/10.1038/nature08897

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Kromrey, S., Maestri, M., Hauffen, K., Bart, E., & Hegdé, J. (2010). Fragment-based learning of visual object categories in non-human primates. *PLoS ONE*, *5*(11). https://doi.org/10.1371/journal.pone.0015444

Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. *Psychology of Learning and Motivation - Advances in Research and Theory*, *28*(C), 207–250. https://doi.org/10.1016/S0079-7421(08)60491-0

Kurtz, K. J. (2007). The divergent autoencoder (DIVA) model of category learning. *Psychonomic Bulletin and Review*, *14*(4), 560–576. https://doi.org/10.3758/BF03196806

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260. https://doi.org/10.1037/0033-295X.107.2.227

Lamour, Y., Dutar, P., & Jobert, A. (1984). Cortical projections of the nucleus of the diagonal band of broca and of the substantia innominata in the rat: An anatomical study using the anterograde transport of a conjugate of wheat germ agglutinin and horseradish peroxidase. *Neuroscience*, *12*(2), 395–408. https://doi.org/10.1016/0306-4522(84)90061-7

Laramée, M. E., & Boire, D. (2015). Visual cortical areas of the mouse: Comparison of parcellation and network structure with primates. *Frontiers in Neural Circuits*, *8*(JAN), 1–16. https://doi.org/10.3389/fncir.2014.00149

Lashley, K. S. (1930). The Mechanism of Vision: III. The Comparative Visual Acuity of Pigmented and Albino Rats. *Pedagogical Seminary and Journal of Genetic Psychology*, *37*(4), 481–484. https://doi.org/10.1080/08856559.1930.9944157

Lee, I., & Lee, S. H. (2012). Putting an object in context and acting on it: Neural mechanisms of goal-directed response to contextual object. *Reviews in the Neurosciences*, *24*(1), 27–49. https://doi.org/10.1515/revneuro-2012-0073

Lee, J. Q., Zelinski, E. L., McDonald, R. J., & Sutherland, R. J. (2016). Heterarchic reinstatement of long-term memory: A concept on hippocampal amnesia in rodent memory research. *Neuroscience and Biobehavioral Reviews*, *71*, 154–166. https://doi.org/10.1016/j.neubiorev.2016.08.034

Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, K. (2014). Dimensionality of object representations in monkey inferotemporal cortex. *Neural Computation*, *1872*(10), 1840–1872. https://doi.org/10.1162/NECO

Lesburguères, E., Gobbo, O. L., Alaux-Cantin, S., Hambucken, A., Trifilieff, P., & Bontempi, B. (2011). Early tagging of cortical networks is required for the formation of enduring associative memory. *Science*, *331*(6019), 924–928. https://doi.org/10.1126/science.1196164

Lindh, D., Sligte, I. G., Assecondi, S., Shapiro, K. L., & Charest, I. (2019). Conscious perception of natural images is constrained by category-related visual features. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-12135-3

Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, *72*(1), 95–104. https://doi.org/10.1037/h0023319

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, *4*(5), 401–414. https://doi.org/10.1016/S0960-9822(00)00089-0

Logothetis, Nikos K., & Sheinberg, D. L. (1996). Visual Object Recognition. *Annual Review of Neuroscience*, 577–621. https://doi.org/https://doi.org/10.1146/annurev.ne.19.030196.003045

Lopes-dos-Santos, V., Conde-Ocazionez, S., Nicolelis, M. A. L., Ribeiro, S. T., & Tort, A. B. L. (2011). Neuronal assembly detection and cell membership specification by principal component analysis. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0020996

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Low, R. J., Gu, Y., & Tank, D. W. (2014). Cellular resolution optical access to brain regions in fissures: imaging medial prefrontal cortex and grid cells in entorhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18739–18744. https://doi.org/10.1073/pnas.1421753111

Mao, D., Kandler, S., Mcnaughton, B. L., & Bonin, V. (2017). Sparse orthogonal population representation of Spatial Context in the Retrosplenial Cortex. *Nature Communications*, *8*(243), 1–9. https://doi.org/10.1038/s41467-017-00180-9

Mao, D., Neumann, A. R., Sun, J., Bonin, V., Mohajerani, M. H., & McNaughton, B. L. (2018). Hippocampus-dependent emergence of spatial sequence coding in retrosplenial cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), 8015–8018. https://doi.org/10.1073/pnas.1803224115

Mar, A. C., Horner, A. E., Nilsson, S. R. O., Alsiö, J., Kent, B. A., Kim, C. H., … Bussey, T. J. (2013). The touchscreen operant platform for assessing executive function in rats and mice. *Nature Protocols*, *8*(10), 1985–2005. https://doi.org/10.1038/nprot.2013.123

Markham, K. R., Butt, A. E., & Dougher, M. J. (1996). A Computer Touch-Screen Apparatus for Training Visual Discriminations in Rats. *Training*, *65*(I), 173–182. https://doi.org/10.1901/jeab.1996.65-173

Marr, D. (1970). A Theory for Cerebral Neocortex. *Proceedings of the Royal Society B: Biological Sciences*, *176*(1043), 161–234. https://doi.org/10.1098/rspb.1970.0040

Marr, D. (1971). Simple Memory: A Theory for Archicortex. *Philosophical Transactions of the Royal Society B, 262*. https://doi.org/http://doi.org/10.1098/rstb.1971.0078

Maurer, A. P., Burke, S. N., Diba, K., & Barnes, C. A. (2017). Attenuated activity across multiple cell types and reduced monosynaptic connectivity in the aged perirhinal cortex. *Journal of Neuroscience*, *37*(37), 8965–8974. https://doi.org/10.1523/JNEUROSCI.0531-17.2017

Mayrhofer, J. M., El-Boustani, S., Foustoukos, G., Auffret, M., Tamura, K., & Petersen, C. C. H. (2019). Distinct Contributions of Whisker Sensory Cortex and Tongue-Jaw Motor Cortex in a Goal-Directed Sensorimotor Transformation. *Neuron*, *103*(6), 1034-1043.e5. https://doi.org/10.1016/j.neuron.2019.07.008

McClelland, J. L. (2010). Emergence in Cognitive Science. *Topics in Cognitive Science*, *2*(4), 751–770. https://doi.org/10.1111/j.1756-8765.2010.01116.x

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190–1210. https://doi.org/10.1037/a0033812

McClelland, J. L., & Goddard, N. H. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus*, *6*, 654–665. https://doi.org/10.1002/(SICI)1098-1063(1996)6:6<654::AID-HIPO8>3.0.CO;2-G

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McNaughton, B. L. (2010). Cortical hierarchies, sleep, and the extraction of knowledge from memory. *Artificial Intelligence*, *174*(2), 205–214. https://doi.org/10.1016/j.artint.2009.11.013

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(5), 355–368. https://doi.org/10.1037/0278-7393.7.5.355

Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cognitive Psychology*, *1*(3), 242–299. https://doi.org/10.1016/0010-0285(70)90017-4

Meyers, E. M. (2018). Dynamic population coding and its relationship to working memory. *Journal of Neurophysiology*, *120*(5), 2260–2268. https://doi.org/10.1152/jn.00225.2018

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

Milivojevic, B., Vicente-Grabovetsky, A., & Doeller, C. F. (2015). Insight reconfigures hippocampal-prefrontal memories. *Current Biology*, *25*(7), 821–830. https://doi.org/10.1016/j.cub.2015.01.033

Miller, E. K., & Buschman, T. J. (2007). Rules through Recursion: How Interactions between the Frontal Cortex and Basal Ganglia May Build Abstract, Complex Rules from Concrete, Simple Ones. *Neuroscience of Rule-Guided Behavior*. https://doi.org/10.1093/acprof:oso/9780195314274.003.0022

Minamimoto, T., La Camera, G., & Richmond, B. J. (2009). Measuring and modeling the interaction among reward size, delay to reward, and satiation level on motivation in monkeys. *Journal of Neurophysiology*, *101*(1), 437–447. https://doi.org/10.1152/jn.90959.2008

Minamimoto, T., Saunders, R. C., & Richmond, B. J. (2010). Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron*, *66*(4), 501–507. https://doi.org/10.1016/j.neuron.2010.04.010

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*(1), 57–77. https://doi.org/10.1016/0166-4328(82)90081-X

Mitchnick, K. A., Wideman, C. E., Huff, A. E., Palmer, D., McNaughton, B. L., & Winters, B. D. (2018). Development of novel tasks for studying view-invariant object recognition in rodents: Sensitivity to scopolamine. *Behavioural Brain Research*, *344*(February), 48–56. https://doi.org/10.1016/j.bbr.2018.01.030

Moorman, D. E., & Aston-Jones, G. (2015). Prefrontal neurons encode context-based response execution and inhibition in reward seeking and extinction. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(30), 9472–9477. https://doi.org/10.1073/pnas.1507611112

Morrison, J. H., Molli Ver, M. E., Grzanna, R., & Coyle, J. T. (1979). Noradrenergic innervation patterns in three regions of medial cortex: An immunofluorescence characterization. *Brain Research Bulletin*, *4*(6), 849–857. https://doi.org/10.1016/0361-9230(79)90022-4

Nevid, J. S. (2007). Kant, cognitive psychotherapy, and the hardening of the categories. *Psychology and Psychotherapy: Theory, Research and Practice*, *80*(4), 605–615. https://doi.org/10.1348/147608307X204189

Nielsen, K. J., Logothetis, N. K., & Rainer, G. (2006). Discrimination Strategies of Humans and Rhesus Monkeys for Complex Visual Displays. *Current Biology*, *16*(8), 814–820. https://doi.org/10.1016/j.cub.2006.03.027

Nithianantharajah, J., McKechanie, A. G., Stewart, T. J., Johnstone, M., Blackwood, D. H., St Clair, D., … Saksida, L. M. (2015). Bridging the translational divide: Identical cognitive touchscreen testing in mice and humans carrying mutations in a disease-relevant homologous gene. *Scientific Reports*, *5*(February), 3–7. https://doi.org/10.1038/srep14613

Nosofsky, R. M. (1986). Attention , Similarity , and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Similarity, Frequency, and Category Representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54–65. https://doi.org/10.1037/0278-7393.14.1.54

Nosofsky, R. M. (2011). The generalized context model: an exemplar model of classification. In P. E. M. & A. J. Willis (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511921322.002

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79. https://doi.org/10.1037/0033-295x.101.1.53

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., … Zeng, H. (2014). A mesoscale connectome of the mouse brain. *Nature*, *508*(7495), 207–214. https://doi.org/10.1038/nature13186

Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., & Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, *466*(7306), 617–621. https://doi.org/10.1038/nature09178

Okamura, J. Y., Yamaguchi, R., Honda, K., Wang, G., & Tanaka, K. (2014). Neural substrates of view-invariant object recognition developed without experiencing rotations of the objects. *Journal of Neuroscience*, *34*(45), 15047–15059. https://doi.org/10.1523/JNEUROSCI.1898-14.2014

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*(4), 481–487. https://doi.org/10.1016/j.conb.2004.07.007

Öngür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, *10*(3), 206–219. https://doi.org/10.1093/cercor/10.3.206

Op de Beeck, H., Wagemans, J., & Vogels, R. (2003). The Effect of Category Learning on the Representation of Shape: Dimensions Can Be Biased but not Differentiated. *Journal of Experimental Psychology: General*, *132*(4), 491–511. https://doi.org/10.1037/0096-3445.132.4.491

Pachitariu, M., Stringer, C., Schröder, S., Dipoppa, M., Rossi, L. F., Carandini, M., & Harris, K. D. (2016). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *BioRxiv*, 061507. https://doi.org/10.1101/061507

Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY*, *54A*(1), 197–235. https://doi.org/10.1080/0272498004200008

Pan, X., & Sakagami, M. (2012). Category representation and generalization in the prefrontal cortex. *European Journal of Neuroscience*, *35*(7), 1083–1091. https://doi.org/10.1111/j.1460-9568.2011.07981.x

Pan, X., Sawa, K., Tsuda, I., Tsukada, M., & Sakagami, M. (2008). Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. *Nature Neuroscience*, *11*(6), 703–712. https://doi.org/10.1038/nn.2128

Pandya, D. N., & Yeterian, E. H. (1990). Prefrontal cortex in relation to other cortical areas in rhesus monkey: Architecture and connections. In H. B.M. Uylings, J. P. . Vand Eden, M. . De Bruin, & M. G. . Fenstra (Eds.), *Progress in Brain Research* (Vol. 85, pp. 63–94). Elsevier. https://doi.org/10.1016/S0079-6123(08)62676-X

Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*, *105*(October 2016), 177–183. https://doi.org/10.1016/j.neuropsychologia.2017.03.033

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97. https://doi.org/10.1038/nature08103

Perez-Orive, J. (2002). Oscillations and Sparsening of Odor Representations in the Mushroom Body. *Science*, *297*(5580), 359–365. https://doi.org/10.1126/science.1070502

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*(3), 329–342. https://doi.org/10.1007/BF00239352

Peters, G. J., David, C. N., Marcus, M. D., & Smith, D. M. (2013). The medial prefrontal cortex is critical for memory retrieval and resolving interference. *Learning and Memory*, *20*(4), 201–209. https://doi.org/10.1101/lm.029249.112

Piaget, J. (1923). Les fonctions du langage de deux enfants de six ans. In *Langage et pensée chez l'enfant* (3e Ed., pp. 14–46). Neuchâtel: Delachaux et Niestlé. Retrieved from http://www.fondationjeanpiaget.ch/fjp/site/textes/VE/JP23_Langag_pensee_chap1_fonctionslang.pdf

Pnevmatikakis, E. A., Soudry, D., Gao, Y., Machado, T. A., Merel, J., Pfau, D., … Paninski, L. (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron*, *89*(2), 285. https://doi.org/10.1016/j.neuron.2015.11.037

Poggio, T., & Riesenhuber, M. (2000). Models of object recognition. *Nature Neuroscience*, 1199–1204. https://doi.org/10.1038/81479

Posner, M. I., & Keele, S. W. (1968). On the Genesis of Abstract Ideas. *Journal of Experimental Psychology: Animal Behavior Processes*, *77*(3), 353–362. https://doi.org/10.1037/0097-7403.29.3.c2

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2 PART 1), 304–308. https://doi.org/10.1037/h0028558

Preuss, T. M. (1995). Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered. *Journal of Cognitive Neuroscience*, *7*(1), 1–24. https://doi.org/10.1162/jocn.1995.7.1.1

Quillian, M. R. (1966). Semantic Memory. *Bolt Bernek and Newman Inc.* Cambridge, MA:

Advanced Research Projects Agency (ARPA). Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/641671.pdf

Quillian, M. R. (1967). Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430. https://doi.org/10.1002/bs.3830120511

Ragozzino, M. E., & Kesner, R. P. (1998). The effects of muscarinic cholinergic receptor blockade in the rat anterior cingulate and prelimbic/infralimbic cortices on spatial working memory. *Neurobiology of Learning and Memory*, *69*(3), 241–257. https://doi.org/10.1006/nlme.1998.3823

Range, F., Aust, U., Steurer, M., & Huber, L. (2008). Visual categorization of natural stimuli by domestic dogs. *Animal Cognition*, *11*(2), 339–347. https://doi.org/10.1007/s10071-007-0123-2

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

Richards, B. a, Xia, F., Santoro, A., Husse, J., Woodin, M. a, Josselyn, S. a, & Frankland, P. W. (2014). Patterns across multiple memories are identified over time. *Nature Neuroscience*, *17*(7), 981–986. https://doi.org/10.1038/nn.3736

Richler, J. J., & Palmeri, T. J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 75–94. https://doi.org/10.1002/wcs.1268

Richler, J. J., Wilmer, J. B., & Gauthier, I. (2017). General object recognition is specific: Evidence from novel and familiar objects. *Cognition*, *166*(September), 42–55. https://doi.org/10.1016/j.cognition.2017.05.019

Riga, D., Matos, M. R., Glas, A., Smit, A. B., Spijker, S., & Van den Oever, M. C. (2014). Optogenetic dissection of medial prefrontal cortex circuitry. *Frontiers in Systems Neuroscience*, *8*(December), 1–19. https://doi.org/10.3389/fnsys.2014.00230

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. https://doi.org/10.1016/S0022-5371(73)80056-8

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, *179*(1), 3–20. https://doi.org/10.1016/0006-8993(79)90485-2

Rolls, E. T. (2016). Pattern Completion and Pattern Separation Mechanisms in the Hippocampus. In P. A. Jackson, A. A. Chiba, R. F. Berman, & M. E. Ragozzino (Eds.), *The Neurobiological Basis of Memory: A System, Attribute, and Process Analysis* (pp. 77–113). Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-15759-7

Rolls, E. T., & Milward, T. (2000). A Model of Invariant Object Recognition in the Visual System: Learning Rules, Activation Functions, Lateral Inhibition, and Information-Based Performance Measures, *2572*, 2547–2572. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.7487&rep=rep1&type=pdf

Rolls, E. T., & Treves, A. (1990). The relative advantages of sparse versus distributed encoding

for associative neuronal networks in the brain. *Network: Computation in Neural Systems*, *1*(4), 407–421. https://doi.org/10.1088/0954-898X/1/4/002

Room, P., Russchen, F. T., Groenewegen, H. J., & Lohman, A. H. M. (1985). Efferent connections of the prelimbic (area 32) and the infralimbic (area 25) cortices: An anterograde tracing study in the cat. *Journal of Comparative Neurology*, *242*(1), 40–55. https://doi.org/10.1002/cne.902420104

Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*(4), 532–547. https://doi.org/10.1016/0010-0285(75)90021-3

Rosch, E., & Mervis, C. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502. https://doi.org/10.1037/0096-1523.2.4.491

Rose, J., & Woosley, C. (1948). Structure and Relations of Limbic Cortex and Anterior Thalamic Nuclei in Rabbit and Cat. *The Journal of Comparative Neurology*, *98*(3), 279–347. https://doi.org/10.1002/cne.900890307

Rosselli, F. B., Alemi, A., Ansuini, A., & Zoccolan, D. (2015). Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. *Frontiers in Neural Circuits*, *9*(March), 1–22. https://doi.org/10.3389/fncir.2015.00010

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience*, *30*(25), 8519–8528. https://doi.org/10.1523/JNEUROSCI.4837-09.2010

Royer, J., Blais, C., Gosselin, F., Duncan, J., & Fiset, D. (2015). When less is more: Impact of face processing ability on recognition of visually degraded faces. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(5), 1179–1183. https://doi.org/10.1037/xhp0000095

Rumelhart, D. E., & Norman, D. A. (1973). Active semantic networks as a model of human memory. *Proceedings of the 3rd International Joint Conference on Artificial Intelligence*, (October), 450–457. Retrieved from http://portal.acm.org/citation.cfm?id=1624830

Rumelhart, D. E., & Ortony, A. (1977). The Representation of Knowledge in Memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the Acquisition of Knowledge* (1st ed., pp. 99–135). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315271644-10

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*, 3–30.

Sakagami, M., & Pan, X. (2007). Functional role of the ventrolateral prefrontal cortex in decision making. *Current Opinion in Neurobiology*, *17*(2), 228–233. https://doi.org/10.1016/j.conb.2007.02.008

Sakagami, M., Pan, X., & Uttl, B. (2006). Behavioral inhibition and prefrontal cortex in decision-making. *Neural Networks*, *19*(8), 1255–1265. https://doi.org/10.1016/j.neunet.2006.05.040

Sakagami, M., & Tsutsui, K. I. (1999). The hierarchical organization of decision making in the primate prefrontal cortex. *Neuroscience Research*, *34*(2), 79–89. https://doi.org/10.1016/S0168-0102(99)00038-3

Salatas, H., & Bourne, L. E. (1974). Learning conceptual rules: III. Processes contributing to rule difficulty. *Memory & Cognition*, *2*(3), 549–553. https://doi.org/10.3758/BF03196919

Salz, D. M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M. W., & Eichenbaum, H. (2016). Time cells in hippocampal area CA3. *Journal of Neuroscience*, *36*(28), 7476–7484. https://doi.org/10.1523/JNEUROSCI.0087-16.2016

Schaeffer, B., & Wallace, R. (1969). Semantic similarity and the comparison of word meanings. *Journal of Experimental Psychology*, *82*(2), 343–346. https://doi.org/10.1037/h0028287

Schneider, G. E. (1969). Two Visual Systems. *Science*, (February), 895–902.

Seamans, J. K., Lapish, C. C., & Durstewitz, D. (2008). Comparing the prefrontal cortex of rats and primates: Insights from electrophysiology. *Neurotoxicity Research*, *14*(2–3), 249–262. https://doi.org/10.1007/BF03033814

Seger, C. A., & Miller, E. K. (2010). Category Learning in the Brain. *Annual Review of Neuroscience*, *33*(1), 203–219. https://doi.org/10.1146/annurev.neuro.051508.135546

Seger, C. A., & Peterson, E. J. (2013). Categorization=decision making+generalization. *Neuroscience and Biobehavioral Reviews*, *37*(7), 1187–1200. https://doi.org/10.1016/j.neubiorev.2013.03.015

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and Memorization of Classifications. *Psychological Monographs: General and Applied*, *75*(13). https://doi.org/10.1037/h0093825

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320. https://doi.org/10.1038/415318a

Smith, E. E. (1967). Effects of Familiarity on Stimulus Recognition and Categorization. *Journal of Experimental Psychology*, *74*(3), 324–332. https://doi.org/10.1037/h0021274

Smith, J. D., & Minda, J. P. (199b). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(6), 1411–1436. https://doi.org/10.1037/0278-7393.24.6.1411

Smith, J. D., Redford, J. S., & Haas, S. M. (2008). Prototype Abstraction by Monkeys (Macaca mulatta). *Journal of Experimental Psychology: General*, *137*(2), 390–401. https://doi.org/10.1037/0096-3445.137.2.390

Snyder, H. R., & Munakata, Y. (2010). Becoming self-directed: Abstract representations support endogenous flexibility in children. *Cognition*, *116*(2), 155–167. https://doi.org/10.1016/j.cognition.2010.04.007

Sporns, O., & Bullmore, E. T. (2014). From connections to function: The mouse brain connectome atlas. *Cell*, *157*(4), 773–775. https://doi.org/10.1016/j.cell.2014.04.023

Sripati, A. P., & Olson, C. R. (2010). Responses to compound objects in monkey inferotemporal cortex: The whole is equal to the sum of the discrete parts. *Journal of Neuroscience*, *30*(23), 7948–7960. https://doi.org/10.1523/JNEUROSCI.0016-10.2010

Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology*, *83*(3 PART 1), 508–510. https://doi.org/10.1037/h0028846

Sutherland, R. ., Sparks, F., & Lehmann, H. (2010). Modest Proposal for the Situation of Systems Consolidation, *48*(8), 2357–2369. https://doi.org/10.1016/j.neuropsychologia.2010.04.015.Hippocampus

Sutherland, R. J., & Lehmann, H. (2011). Alternative conceptions of memory consolidation and the role of the hippocampus at the systems level in rodents. *Current Opinion in Neurobiology*, *21*(3), 446–451. https://doi.org/10.1016/j.conb.2011.04.007

Szabo, M., Deco, G., Fusi, S., Del Giudice, P., Mattia, M., & Stetter, M. (2006). Learning to attend: Modeling the shaping of selectivity in infero-temporal cortex in a categorization task. *Biological Cybernetics*, *94*(5), 351–365. https://doi.org/10.1007/s00422-006-0054-z

Tafazoli, S., Di Filippo, A., & Zoccolan, D. (2012). Transformation-Tolerant Object Recognition in Rats Revealed by Visual Priming. *Journal of Neuroscience*, *32*(1), 21–34. https://doi.org/10.1523/JNEUROSCI.3932-11.2012

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., … Fernandez, G. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, *103*(3), 756–761. https://doi.org/10.1073/pnas.0507774103

Talpos, J. C., Winters, B. D., Dias, R., Saksida, L. M., & Bussey, T. J. (2009). A novel touchscreen-automated paired-associate learning (PAL) task sensitive to pharmacological manipulation of the hippocampus: A translational rodent model of cognitive impairments in neurodegenerative disease. *Psychopharmacology*, *205*(1), 157–168. https://doi.org/10.1007/s00213-009-1526-3

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139. https://doi.org/10.1146/annurev.ne.19.030196.000545

Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The Ubiquity of Small-World Networks. *Brain Connectivity*, *1*(5). https://doi.org/10.1089/brain.2011.0038

Thierry, A. M., Blanc, G., Sobel, A., Stinus, L., & Glowinski, J. (1973). Dopaminergic terminals in the rat cortex. *Science*, *182*(4111), 499–501. https://doi.org/10.1126/science.182.4111.499

Todd Maddox, W., Gregory Ashby, F., & Bohil, C. J. (2003). Delayed Feedback Effects on Rule-Based and Information-Integration Category Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(4), 650–662. https://doi.org/10.1037/0278-7393.29.4.650

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*(6754), 699–703. https://doi.org/10.1038/44372

Townsend, J. T., & Ashby, F. G. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154–179. Retrieved from http://www.indiana.edu/~psymodel/papers/ashtow86.pdf

Treves, A., & Rolls, E. T. (1991). What Determines the Capacity of autoassociative memories in the brain? *Network*, *2*, 371–397.

Tripathi, A., Schenker, E., Spedding, M., & Jay, T. M. (2016). The hippocampal to prefrontal cortex circuit in mice: a promising electrophysiological signature in models for psychiatric disorders. *Brain Structure and Function*, *221*(4), 2385–2391. https://doi.org/10.1007/s00429-015-1023-x

Troje, N. F., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999). Categorical learning in pigeons: The role of texture and shape in complex static stimuli. *Vision Research*, *39*(2), 353–366. https://doi.org/10.1016/S0042-6989(98)00153-9

Tronel, S., Feenstra, M. G. P., & Sara, S. J. (2004). Noradrenergic action in prefrontal cortex in the late stage of memory consolidation. *Learning and Memory*, *11*(4), 453–458. https://doi.org/10.1101/lm.74504

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., … Morris, R. G. M. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science*, *333*(6044), 891–895. https://doi.org/10.1126/science.1205274

Tulving, E. (1972). Episodic and semantic memory. In T. & W. & D. Ed. (Eds.), *Episodic and Semantic Memory: Organization of Memory* (Vol. 1, pp. 381–403). New York, NY: Academic Press. https://doi.org/10.1017/S0140525X00047257

Tversky, A. (1977). Psychological Review. *Journal of Mental Science*, *84*(4), 327–351. https://doi.org/https://doi.org/10.1037/0033-295X.84.4.327

Tversky, A., & Itamar, G. (1978). Studies of Similarity. *Cognition and Categorization*, 79–98.

Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., & Nikolic, D. (2009). Neural synchrony in cortical networks: history, concept and current status. *Frontiers in Integrative Neuroscience*, *49*(13), 3662–3669. https://doi.org/10.3389/neuro.07

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*(7), 682–687. https://doi.org/10.1038/nn870

Uylings, Harry B.M., Groenewegen, H. J., & Kolb, B. (2003). Do rats have a prefrontal cortex? *Behavioural Brain Research*, *146*(1–2), 3–17. https://doi.org/10.1016/j.bbr.2003.09.028

Uylings, Harry B.M., & Van Eden, C. G. (1991). Chapter 3 Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Progress in Brain Research*, *85*(C), 31–62. https://doi.org/10.1016/S0079-6123(08)62675-8

Van De Werd, H. J.J.M., Rajkowska, G., Evers, P., & Uylings, H. B. M. (2010). Cytoarchitectonic and chemoarchitectonic characterization of the prefrontal cortical areas in the mouse. *Brain Structure and Function*, *214*(4), 339–353. https://doi.org/10.1007/s00429-010-0247-z

Van De Werd, Henri J.J.M., & Uylings, H. B. M. (2014). Comparison of (stereotactic) parcellations in mouse prefrontal cortex. *Brain Structure and Function*, *219*(2), 433–459. https://doi.org/10.1007/s00429-013-0630-7

Van Eden, C. G., & Buijs, R. M. (2000). Functional neuroanatomy of the prefrontal cortex: Autonomic interactions. *Progress in Brain Research*, *126*, 49–62. https://doi.org/10.1016/S0079-6123(00)26006-8

Van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Vinje, W. E., & Gallant, J. L. (2000). Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision. *Science*, *287*(5456), 1273–1276. https://doi.org/10.1126/science.287.5456.1273

Vinken, K., Vermaercke, B., & Op de Beeck, H. P. (2014). Visual categorization of natural movies by rats. *Journal of Neuroscience*, *34*(32), 10645–10658. https://doi.org/10.1523/JNEUROSCI.3663-13.2014

Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part 1: Behavioural study. *European Journal of Neuroscience*, *11*(4), 1223–1238. https://doi.org/10.1046/j.1460-9568.1999.00530.x

Vogels, R. (1999b). Categorization of complex visual images by rhesus monkeys. Part 2: Single Cell Study. *European Journal of Neuroscience*, *11*(4), 1239–1255. https://doi.org/10.1046/j.1460-9568.1999.00530.x

Wallis, G., & Rolls, E. T. (1997). Invariant Face and Object Recognition in the Visual System. *Progress in Neurobiology*, *51*, 167–194.

Wang, G., Obama, S., Yamashita, W., Sugihara, T., & Tanaka, K. (2005). Prior experience of rotation is not required for recognizing objects seen from different angles. *Nature Neuroscience*, *8*(12), 1568–1575. https://doi.org/10.1038/nn1600

Wang, Q., Gao, E., & Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. *Journal of Neuroscience*, *31*(5), 1905–1918. https://doi.org/10.1523/JNEUROSCI.3488-10.2011

Wang, Q., Sporns, O., & Burkhalter, A. (2012). Network analysis of corticocortical connections

reveals ventral and dorsal processing streams in mouse visual cortex. *Journal of Neuroscience*, *32*(13), 4386–4399. https://doi.org/10.1523/JNEUROSCI.6063-11.2012

Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual Behavior in Pigeons: Categories, Subcategories, and Pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(3), 235–246. https://doi.org/10.1037/0097-7403.14.3.235

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(June), 440–442. https://doi.org/10.1038/30918

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Transient subcortical connections of inferior temporal areas TE and TEO in infant macaque monkeys. *Journal of Comparative Neurology*, *352*(2), 213–226. https://doi.org/10.1002/cne.903520205

Weiskrantz, L., & Saunders, R. C. (1984). Impairments of Visual Object. *Brain*, 1033–1072. https://doi.org/10.1093/brain/107.4.1033

Willmore, B., & Tolhurst, D. J. (2001). Characterizing the sparseness of neural codes. *Network (Bristol, England)*, *12*(3), 255–270. https://doi.org/10.1080/713663277

Winters, B. D., & Reid, J. M. (2010). A Distributed Cortical Representation Underlies Crossmodal Object Recognition in Rats. *The Journal of Neuroscience* , *30*(18), 6253–6261. https://doi.org/10.1523/JNEUROSCI.6073-09.2010

Winters, B. D., Saksida, L. M., & Bussey, T. J. (2008). Object recognition memory: Neurobiological mechanisms of encoding, consolidation and retrieval. *Neuroscience and Biobehavioral Reviews*, *32*(5), 1055–1070. https://doi.org/10.1016/j.neubiorev.2008.04.004

Wixted, J. T., Squire, L. R., Jang, Y., Papesh, M. H., Goldinger, S. D., Kuhn, J. R., … Steinmetz, P. N. (2014). Sparse and distributed coding of episodic memory in neurons of the human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), 9621–9626. https://doi.org/10.1073/pnas.1408365111

Wolbers, T., & Buchel, C. (2005). Dissociable Retrosplenial and Hippocampal Contributions to Successful Formation of Survey Representations. *Journal of Neuroscience*, *25*(13), 3333–3340. https://doi.org/10.1523/JNEUROSCI.4705-04.2005

Wright, A. A., & Katz, J. S. (2007). Generalization Hypothesis of Abstract-Concept Learning: Learning Strategies and Related Issues in Macaca mulatta, Cebus apella, and Columba livia. *Journal of Comparative Psychology*, *121*(4), 387–397. https://doi.org/10.1037/0735-7036.121.4.387

Wutz, A., Loonis, R., Roy, J. E., Donoghue, J. A., & Miller, E. K. (2018). Different Levels of Category Abstraction by Different Dynamics in Different Prefrontal Areas. *Neuron*, 1–11. https://doi.org/10.1016/j.neuron.2018.01.009

Xia, F., Richards, B. A., Tran, M. M., Josselyn, S. A., Takehara-Nishiuchi, K., & Frankland, P. W. (2017). Parvalbumin-positive interneurons mediate neocortical-hippocampal interactions that are necessary for memory consolidation. *ELife*, *6*, 1–25. https://doi.org/10.7554/eLife.27868

Yasushi Miyashita. (1988). Neuronal correlate of visual associative long-term memory in the

primate temporal cortex. *Nature*, 20.

Zaitsev, A. V., Povysheva, N. V., Gonzalez-Burgos, G., Rotaru, D., Fish, K. N., Krimer, L. S., & Lewis, D. A. (2009). Interneuron diversity in layers 2-3 of monkey prefrontal cortex. *Cerebral Cortex, 19*(7), 1597–1615. https://doi.org/10.1093/cercor/bhn198

Zoccolan, D., Oertelt, N., DiCarlo, J. J., & Cox, D. D. (2009). A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences, 106*(21), 8748–8753. https://doi.org/10.1073/pnas.0811583106