

SPATIOTEMPORAL DYNAMICS OF MULTIPLE MEMORY SYSTEMS DURING
NATURALISTIC CATEGORIZATION

by

KYLE K. MORGAN

A DISSERTATION

Presented to the Department of Psychology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2019

DISSERTATION APPROVAL PAGE

Student: Kyle K. Morgan

Title: Spatiotemporal Dynamics of Multiple Memory Systems During Naturalistic Categorization

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Psychology by:

Dasa Zeithamova	Chairperson
Don Tucker	Core Member
Mike Posner	Core Member
Ian Greenhouse	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School
-----------------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2019

© 2019 Kyle K. Morgan

DISSERTATION ABSTRACT

Kyle K. Morgan

Doctor of Philosophy

Department of Psychology

June 2019

Title: Spatiotemporal Dynamics of Multiple Memory Systems During Naturalistic Categorization

How we learn and remember has been the focus of centuries worth of psychological research. Category learning has been a convenient domain to study these concepts, where modern imaging evidence has shown that the brain uses multiple memory systems that specialize in learning specific category structures. However, a question still remains about whether these systems alternate or operate alongside one another to maximize performance in everyday tasks. The primary goals of this dissertation were to characterize the mechanisms associated with category learning, and understand the extent to which different memory systems are recruited within a single task. Three studies providing spatial and temporal distinctions between learning-related changes in the brain and category-dependent memory systems are presented. The results from these experiments support the notion that exemplar memorization, rule-based, and perceptual similarity-based categorization are flexibly recruited to optimize performance during a single task. We conclude that these three methods, along with their underlying memory systems, aid in the development of expertise, but their engagement may depend on the level of familiarity with a category. Characterizing the conditions under which these systems are recruited will play an important role in future studies that will facilitate their engagement to enhance learning. This dissertation includes previously published

and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Kyle K. Morgan

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene (Graduate)
University of Oregon, Eugene (Undergraduate)

DEGREES AWARDED:

Doctor of Philosophy, Cognitive Neuroscience, 2019, University of Oregon
Master of Science, Psychology, 2014, University of Oregon
Bachelor of Science, Psychology, 2012, University of Oregon

AREAS OF SPECIAL INTEREST:

Cognitive Neuroscience
Learning & Memory
Multiple Memory System Theory
Electroencephalography
Functional Magnetic Resonance Imaging
Machine Learning
Cognitive Training
Accelerated Learning

PROFESSIONAL EXPERIENCE:

Graduate Employee, University of Oregon, 2014-Present
Department of Psychology

Researcher, Philips Neuro, 2017-2018

Researcher, Electrical Geodesics, Inc., 2012-2018

Research Assistant, University of Oregon, 2011-2012
Neuroinformatics Center

Research Assistant, Portland State University, 2009-2010
Department of Psychology

GRANTS, AWARDS, AND HONORS:

Promising Scholar Award, University of Oregon Graduate School, 2015

Honorable Mention, National Science Foundation Graduate Research Fellowship, 2015

Diversity Excellence Scholarship, University of Oregon, 2010-2012

PUBLICATIONS:

Morgan, K. K., Tucker, D. M., & Luu, P. (2018). Understanding the neural mechanisms in rapid recognition of football formations. In R.A. Carlstedt & M. Balconi (Eds.), *Handbook of sport neuroscience and psychophysiology* (Ch. 8). Abingdon, UK: Routledge.

Hou, J., Morgan, K. K., Tucker, D. M., et al. (2016). An improved artifacts removal method for high dimensional eeg. *Journal of Neuroscience Methods*, 268, 31-42.

Morgan, K. K., Luu, P., & Tucker, D. M. (2016). Changes in p3b latency and amplitude reflect expertise acquisition in a football visuomotor learning task. *PLoS ONE*, 11(4), e0154021.

Kuo, C., Luu, P., Morgan, K. K., et al. (2014). Localizing movement-related primary sensorimotor cortices with multi-band frequency changes and functional mri. *PLoS ONE*, 9(11), e112103.

Song, J., Morgan, K. K., Turovets, S., et al. (2013). Anatomically accurate head models and their derivatives for dense array eeg source localization. *Funct Neurol Rehabil Ergon*, 3(2-3), 275-293.

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Professors Zeithamova and Tucker for their assistance in the preparation of this manuscript. Special thanks are due to Dr. Phan Luu, whose consistent mentorship throughout my research career has been paramount to the success of this dissertation. I also thank the remaining members of my dissertation committee, Professors Posner and Greenhouse, for their valued input to the design of the presented research. Finally, the research in this dissertation was only made possible by the dedicated staff in the Brain Electrophysiology Laboratory (BEL), as well as a generous grant from the Lewis Center for Neuroimaging (LCNI) at the University of Oregon.

To my family.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
<u>Introduction</u>	1
<u>Distinct Mechanisms Across Stages of learning</u>	2
<u>Distinct Mechanisms Across Different Category Structures</u>	6
<i>Exemplar Memorization</i>	7
<i>Perceptual Similarity Categorization</i>	9
<i>Rule-Based Categorization</i>	11
<u>Multiple Memory Systems Within a Single Task</u>	11
<i>Neuroimaging</i>	12
<u>Overview of Dissertation</u>	14
II. TRACKING THE ONSET OF AUTOMATICITY IN A FOOTBALL CATEGORY LEARNING TASK.....	15
<u>Introduction</u>	15
<u>Method</u>	19
<i>Participants</i>	19
<i>Task</i>	19
<i>Procedure</i>	21
<i>Learning Criterion</i>	22
<i>EEG Recording and Post-Processing</i>	23
<i>EEG Source Analysis</i>	23
<u>Results</u>	24

Chapter	Page
<i>Behavioral</i>	24
<i>Learning Effects</i>	24
<i>Training Effects</i>	26
<i>Event Related Potentials</i>	28
MFN	29
<i>MFN Learning Effects</i>	29
<i>MFN Training Effects</i>	29
P3b	30
<i>P3b Learning Effects</i>	31
<i>P3b Training Effects</i>	32
<i>P3b Correlations to Behavior</i>	33
<i>Source Analysis</i>	36
MFN	36
P3b	36
<i>Pre-Learning</i>	36
<i>Post-Learning</i>	37
<u>Discussion</u>	38
<i>Learning and Practice</i>	39
<i>Event-Related Potentials</i>	40
MFN	40
P3b	42
<i>Source Analysis</i>	44

Chapter	Page
<i>P3b</i>	44
<i>Limitations</i>	45
<i>Conclusions</i>	46
III. DISSOCIATING MULTIPLE MEMORY SYSTEMS DURING FOOTBALL	
FORMATION CATEGORIZATION.....	47
<u>Introduction</u>	47
<u>Method</u>	50
<i>Participants</i>	50
<i>Task</i>	50
<i>fMRI</i>	52
<i>Procedure</i>	53
<u>Results</u>	54
<i>Behavioral</i>	54
<i>Univariate Analysis</i>	56
<i>Training Runs</i>	56
<i>Generalization Run</i>	59
<i>Multi-Voxel Pattern Analysis</i>	61
<u>Discussion</u>	62
<i>Univariate fMRI</i>	62
<i>MVPA</i>	64
<i>Limitations</i>	65
<i>Conclusions</i>	66

Chapter	Page
IV. EVALUATING TEMPORAL DYNAMICS OF CATEGORIZATION	68
<u>Introduction</u>	68
<u>Method</u>	72
<i>Participants</i>	72
<i>Task</i>	72
<i>Procedure</i>	74
<i>Learning Criterion</i>	75
<i>EEG Recording and Post-Processing</i>	75
<u>Results</u>	75
<i>Behavioral</i>	75
<i>Event-Related Potentials</i>	76
<i>MFN</i>	76
<i>P3b</i>	77
<i>LIAN</i>	78
<i>Machine Learning</i>	79
<i>Data Preparation</i>	79
<i>Whole-Brain SVM Classification</i>	81
<i>Region-Based SVM Classification</i>	82
<u>Discussion</u>	85
<i>ERP Analysis</i>	85
<i>MFN</i>	85
<i>P3b</i>	86

Chapter	Page
<i>LIAN</i>	87
<i>ML Analysis</i>	90
<i>Limitations</i>	91
<i>Conclusions</i>	92
V. GENERAL DISCUSSION	94
<u>Review of Empirical Results from Dissertation</u>	94
<u>Characteristics of Multiple Memory Systems Used in Category Learning</u>	96
<i>Exemplar Memorization</i>	96
<i>Rule-Based Categorization</i>	97
<i>Interactions Between Exemplar Memorization and Rule-Based Categorization</i>	98
<i>Perceptual Similarity Categorization</i>	100
<i>Category Learning Strategies as a Function of Expertise</i>	102
<u>Practical and Theoretical Implications</u>	104
<u>Future Directions</u>	106
REFERENCES CITED.....	108

LIST OF FIGURES

Figure	Page
1. Diagram of events in a single trial for experiment 1	20
2. Graphs of trials to learn by trial type and error rate by trial type and learning condition.	25
3. Graphs of error rates and reaction times across training days	27
4. Channel montages used to quantify the MFN and P3b	28
5. Voltage maps and waveforms of the MFN from representative channels for pre- and post-learning conditions	29
6. Voltage maps and waveforms of the MFN from representative channel for all training days	30
7. Voltage maps and waveforms of the P3b from representative channels for pre- and post-learning conditions	31
8. Voltage maps and waveforms of the P3b from representative channels for all training days	33
9. RT/P3b latency ratio plots	36
10. Source estimates for the P3b during the learning session	37
11. Source estimates for the P3b across training sessions	38
12. Diagram of events in a training for experiment 2	51
13. Diagram of events in a generalization trial for experiment 2	52
14. Categorization accuracy by run	54
15. Confusability matrix	55
16. Generalization accuracy graph	56
17. Univariate contrasts of Rule > Similarity (Red) and Similarity > Rule (Blue) during training	57

Figure	Page
18. Univariate contrasts of Rule > Similarity (Red) and Similarity > Rule (Blue) during generalization	59
19. ROI-based MVPA classification accuracy	62
20. Diagram of events in a trial during training for experiment 3	73
21. Electrode montages used for the LIAN, MFN, and P3b ERP components	76
22. Voltage map and representative waveform of the MFN.....	77
23. Voltage map and representative waveform of the P3b	78
24. Voltage maps and representative waveforms of the left and right LIAN.....	79
25. Electrode montages used to define regions during machine-learning analysis	80
26. Whole-brain classification accuracy over time on an experimental trial.....	82
27. Region-based classification accuracy over time and correlated with behavioral performance	84

LIST OF TABLES

Table	Page
1. Cluster location and size for Rule > Similarity contrast in blocks 4, 5, and 6.....	58
2. Cluster location and size for Similarity > Rule contrast in blocks 4, 5, and 6.....	58
3. Cluster location and size for Rule > Similarity contrast in generalization block.....	60
4. Cluster location and size for Similarity > Rule contrast in generalization block.....	60

CHAPTER I

INTRODUCTION

From Morgan, K. K., Tucker, D. M., & Luu, P. (2018). Understanding the neural mechanisms of memory in rapid recognition of football formations. In R.A. Carlstedt & M. Balconi (Eds.), *Handbook of sport neuroscience and psychophysiology* (Ch. 8). Abingdon, UK: Routledge.

Introduction

My long-term goal is to develop targeted interventions to improve the learning process by facilitating engagement of an optimal memory system for a given task. However, in order to affect learning, we would first need to understand learning and have a map that describes the associated brain areas. But what is learning and what neural mechanisms support it? The answer is not as clear as it may seem.

Category learning has been a productive means to study learning and memory, and refers to the development of the ability to recognize common features among different categories of objects (Bruner, Goodnow, & Austin, 1967). Research using category learning has definitively outlined the mechanisms associated with different learning stages (Shiffrin & Schneider, 1977). On the other hand, we know that different tasks engage completely separate memory systems that are optimized for the learning involved – even for seemingly similar tasks such as in categorization (Zeithamova & Maddox, 2006; Knowlton & Squire, 1993; Gabrieli, 1998; Smith, Patalano, & Jonides, 1998). This makes it difficult to uniquely attribute changes in brain activity to either distinct learning systems or representations of the distinct mechanisms associated with different task sets.

The main goal of this dissertation is to understand the degree to which distinct learning and memory systems may be recruited within the same task. If valid, these characterizations could prove useful for both basic memory research and learning enhancement research. In the memory research domain, it would provide evidence that the brain can flexibly recruit more than one memory system to optimize performance on a single task. Up until this point, this has only been a theory and has yet to be tested. These results would also open an avenue for understanding how expertise within these systems develops alongside one another, and whether that development fits general models of learning progression. In respect to learning enhancement, understanding which systems are engaged and when they are engaged could be a critical component for generating interventions (such as in brain stimulation) that are tailored to specific memory systems – both in their spatial distribution as well as ensuring these interventions are correctly coupled to the timing of each system’s engagement in a task.

Distinct Mechanisms Across Stages of Learning

As a person transitions from being a novice to expertly performing a task, the neural processes initially used to acquire and perform the task disengage; allowing more cognitive resources to be available for other functions (Shiffrin & Schneider, 1977). Fitts & Posner’s (1967) classic model of skill acquisition describes three distinct learning stages: the cognitive, associative, and autonomous stages. The cognitive stage is a period in which the task goals are being established, and the participant explicitly relies on this goal to develop a sequence of actions to accomplish the desired outcome. Learning within this stage often relies on declarative knowledge. When a sequence of actions has been established, the participant transitions into the associative stage that is a period where the

actions are refined. Attention is allocated to specific details of the action sequence, and the participant may explore other solutions (sometimes even throwing out an entire action within the sequence altogether) in order to optimize performance. The final stage is the autonomous stage where learning improvements are slower, but the participant practices their actions in order to make the process more routinized and automatic.

Learning curves across a variety of tasks reflect the general model proposed by Fitts and Posner. Learning in the first two stages occurs rapidly, but towards the end of the associative phase and into the autonomous stage performance improvements slow dramatically. An inversion of the learning curve can be used to describe the amount of working memory and controlled attention required to complete the task over time. Under the Fitts and Posner model, the first two learning stages rely more heavily on explicit control, but as the task sequence becomes more automatic this type of control gives way to a more routinized form of control. Other similar models suggest that learning is a combination of calculated/explicit processes and memory retrieval, and as memories build up there is less reliance on explicit processes and a greater reliance upon streamlined memory retrieval during automatization (Logan, 1988).

A more succinct model of learning has been created to describe the contrasting reliance upon executive functions and has only two stages: early and late. Under the dual-stage model, the early stages of learning from Fitts, Posner, and Logan are combined into a single stage marked as a heavy reliance on controlled processes, which require a person to be actively attentive, and are limited by working memory capacity. In contrast, the late stage (the analogue for the autonomous stage of Fitts and Posner) is defined by its lack of reliance on controlled processes, reflected as automated performance, and is not limited

by working memory capacity and can be carried out subconsciously under the right context (Shiffrin & Schneider, 1977).

Modern imaging evidence have delineated distinct brain networks that are involved in the two learning stages (Chein & Schneider, 2005). The frontal lobe is responsible for the executive monitoring of unfamiliar stimuli; a process that is integral to the early stages of learning. By contrast, cortical regions in the posterior corticolimbic system, which are responsible for habit learning, are engaged when subjects demonstrate proficient performance in the late stages of learning (Chein & Schneider, 2005; Gabriel, Burhans, Talk, & Scalf, 2002). These posterior corticolimbic structures consolidate information and, with sufficient practice, enable performance to be more automated and habitual, removing the need for executive control.

Finer details about how the brain learns to recognize categories are best framed within Schneider's dual processing theory. Information about objects and events, and the context or location under which they occurred are processed in two streams in the cerebral cortex (Schneider, 1969). Within this model, the sensory pathways (e.g. primary visual cortex) take information in from the outside world and help us form an initial identification of an event or object, and then send this information up to the parietal lobe (Ungerleider & Mishkin, 1982; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003). We refer to this stream as the dorsal, or "where" pathway. This pathway specializes in the spatial analysis of stimuli and organizes holistic attention that eventually leads to impulsive actions. There is also a second pathway that is responsible for the identification of "what" event or object is being presented, and this information is processed by the ventral limbic system -- parahippocampal gyrus, piriform, and entorhinal cortex (with the addition of

the amygdala, in humans) – and is referred to as the ventral processing stream (Ungerleider & Mishkin, 1982; Keel, Ivry, Mayr, Hazeltine, & Heuer, 2003). Information from both streams converge at the hippocampus, which is a structure situated in the medial temporal lobe (MTL) that plays a key role in organizing input to link memories by their contextual representation (Luu et al., 2011). Once processing commences within the hippocampus, the output returns to the cortical areas from which the inputs originated (dorsal or ventral). In the dorsal pathway an additional structure, the medial prefrontal cortex (mPFC), selects the memory from the hippocampal feedback, whereas the striatum aids in memory selection for the ventral pathway. This feedback structure allows the hippocampus to organize memory retrieval based off “what” occurred or “where” something occurred, and makes it an essential mechanism for memory retrieval.

The early/deliberate stage of learning is responsible for forming the context under which a someone acquires the relevant information dictating category membership, and requires controlled processing from frontal regions, whereas the automatic phase marks a reduction in frontal engagement (reflected as a reduction in the need for controlled attention) and an increase in activity from more posterior regions where the context is simply monitored. Put more clearly: the early stage is a time where the brain requires more attentional resources to build-up the contextual blueprint that binds inputs and outputs – where we know posterior corticolimbic structures play a large role in associating the two. The lack of context in the early stage leaves little work for this system to do, but as that context forms with practice, the role of posterior/automaticity system becomes increasingly important to the point where controlled processes are no longer needed (Donchin & Coles, 1988; Polich, 2007; Luu et al., 2011). This allows the

person to perform the learned action without a substantial cognitive load so that they can focus their attention elsewhere. This classic anterior/posterior shift seen under most concept learning conditions can be useful for structuring interventions that may speed up the learning process. When using brain stimulation as an example, it might be beneficial to target an intervention over frontal control areas during the initial training phase, and then slowly fade the stimulation towards the posterior as the subject acquires the task.

A relevant downside to the classic anterior/posterior framework is that it is not an all-encompassing model that describes how expertise develops in different memory systems. Although it has been a historically accurate way of describing learning within several systems, there are other memory systems that may have a differential reliance on the brain mechanisms discussed in this section. Likewise, it is possible that multiple memory systems could be engaged simultaneously during a task, either in competition or working in conjunction to optimize learning. In order to maximize any benefit subjects would have from a targeted intervention, it would be optimal to take a multiple systems approach and develop a framework that best describes how the brain behaves under a more varied range of task conditions.

Distinct Mechanisms Across Different Category Structures

Humans possess multiple learning and memory systems, varying in speed of acquisition, flexibility, and the degree of cognitive resources they require, optimizing learning in a wide range of situations (Knowlton & Squire, 1993; Gabrieli, 1998; Smith, et al., 1998). To provide compelling evidence for the existence of multiple systems, prior research has focused on creating tasks which exaggerate the preferential recruitment of one system over another. In addition, evidence from these studies has shown that

performance is hindered when participants fail to engage the memory system optimal for a given category structure. However, detecting which memory system is actually engaged typically requires analyzing behavioral or neuroimaging data averaged across many trials. Identifying signatures of distinct memory systems that would be detectable on short time-scales would allow us to better understand how each system contributes to performance, and also allow us to understand how these systems fit within the expertise development framework.

Categorization is a convenient domain in which to explore how task parameters dictate the recruitment of different memory systems. Placing things into categories is an essential and frequent part of our daily lives, and is fundamental to many tasks used in the laboratory setting. The human categorization systems serve the purpose of optimizing the speed and accuracy of categorization under a variety of different conditions. Historically, there has been a multitude of tasks employed to explore the specific memory system tied to each in isolation (Ashby & Maddox, 2005). However, these tasks can be reduced into three types that are relevant to the context of this dissertation: exemplar memorization, rule-based, and perceptual similarity-based tasks. We will review the general structure of these tasks along with the memory systems engaged during each type.

Exemplar Memorization

Some of the most classic studies of human memory have been centered on the declarative memorization of individual objects. Within the context of category learning, exemplar memorization refers to the memorization of individual examples of a category (i.e. exemplars) along with their category membership (Medin & Schaffer, 1978;

Hintzman 1984; Nosofsky, Clark, & Shin, 1989; Nosofsky & Palmeri, 1997). Theories for exemplar memorization assume that stored memories of specific exemplars viewed during training form the representation of the category. These stored exemplars become a reference for which new probing stimuli are compared against. Under these assumptions, subjects are able to master a combination of well-structured and poorly-structured categories, and learn items that constitute exceptions to the rules governing category membership (Medin & Schaffer, 1978).

The most well-known task that drove the development of exemplar-based models utilized a category structure known as the 5-4 (Medin & Schaffer, 1978; Medin, Altom, & Murphy, 1984; Medin & Smith 1981; Nosofsky 1992; Palmeri & Nosofsky, 1995). In this task, there are 5 exemplars that represent Category A and 4 exemplars that represent Category B. In Category A, four exemplars share three features with the Category A prototype (a representative example of the category), and one exemplar that shares two features. This results in one ambiguous exemplar and four exemplars that are very similar to the prototype. Two of the exemplars in Category B share only 2 features with the Category B prototype, resulting in half of the exemplars in this category being relatively ambiguous. Subjects are trained to recognize members of these categories, and then tested using seven additional items that measure their ability to generalize the category structures. Computational models that strictly model exemplar memorization are able to predict performance on this task better than models describing other methods of categorization (Nosofsky, 1992; Nosofsky, Kruschke, & McKinley, 1992; Palmeri & Nosofsky, 1995). In other tasks that utilize purely arbitrary category labels, it has also been shown that exemplar memorization is the preferred method of categorization (Lei &

Shansheng, 2003). However, in tasks that utilize a random category structure, it is arguable whether exemplar memorization truly qualifies as a formal categorization task at all (Ashby & Waldron, 2000).

Early learning within the exemplar categorization system relies heavily on working memory and frontal control regions. With extended training, the presentation of a probe stimulus prompts the retrieval of the category label associated with the exemplar in the absence of attention (Shiffrin and Schneider, 1977). These retrieval processes are centered on the hippocampus (Nosofsky & Zaki, 1998). Exemplar memorization can be particularly useful when there are only a few exemplars to remember, or when categories are poorly structured (Minda & Smith, 2001; Erickson & Kruschke, 1998). However, in the case that there are discoverable and reliable features determining category membership, it may be more efficient to use a memory system that develops quicker than exemplar memorization.

Perceptual Similarity Categorization

Perceptual similarity categorization involves placing stimuli into categories based off relevant information from other stimuli that are similar to the probe (Ashby & Ell, 2001; Smith & Minda, 2001). One example of perceptual similarity tasks are prototype learning tasks, which require the participant to learn the prototype of a single category, such as a face, and classify other like-stimuli into this category. The stimuli that belong to this category share several common features with the prototype (such as a nose, chin, cheeks, and eyes in our face example) with only a few minor distortions of other features that do not represent the category as a whole (such as hair length). Participants

must extract these common features from the stimuli in the category in order to learn the category prototype (Rosch, 1975; Rosch, 1978).

The process of extracting features from examples within a category typically relies on the perceptual memory system which is mediated by visual cortex (Aizenstein et al., 2000; Ashby & Casale, 2003). However, Zeithamova and colleagues (2008) have revealed that the way in which perceptual similarity tasks are structured plays a major role in the recruitment of the perceptual memory system during categorization. For example, if participants are tasked with learning the prototype of a single category (category A) against examples that do not belong to a category (non-A), then the perceptual memory system is engaged. But, if the participant is tasked with learning two different prototypes and must directly compare examples of each category to one another (category A vs. category B) then this form of prototype categorization relies more heavily on explicit memory, mediated by parietal and frontal control regions. Results from this study suggest that the brain areas which optimize learning even within the same categorization domain are sensitive to small variations in task parameters.

Learning and subsequent processing that occurs within the perceptual similarity system is relatively quick and does not heavily rely on working memory (Waldron & Ashby, 2001; Maddox, Ashby, Ing, & Pickering, 2004; Maddox, Filoteo, Hejl, & Ing, 2004; Zeithamova & Maddox, 2006; Smith & Kemler Nelson, 1984). The perceptual similarity system is crucial for making rapid judgements about category membership, but falls short in its ability to classify objects when within-category similarity is low or between category similarity is high (Nosofsky, 1986). It makes sense that the shortcomings of this memory system are supported by the strengths of another system.

Rule-Based Categorization

Perhaps the most well-known method of categorization is that of a rule-based strategy. In rule-based categorization, the participant is tasked with discovering an explicit rule that dictates category membership. In order for the task to qualify as a classic “rule-based” task, the rule must be easy to verbalize, which differs from other types of categorization that rely on implicit forms of memory subserving strategies that can be difficult to articulate. Common rule-based tasks involve an array of features each with their own variations (dimensions), such as different colors or shapes within an image, that can be used to define members of each category. Rule discovery is commonly achieved through explicit reasoning or hypothesis testing, whereby the participant tests different rules based off the different dimensions of the features within the stimuli until they learn the correct one (Ashby, Alfonso-Reese, Turken, & Waldron, 1998).

Hypothesis testing relies heavily on working memory and controlled attention, which are supported by the working memory system in prefrontal cortex and caudate nucleus (Ashby & Ell, 2001). The working memory system, within the context of rule-based categorization, allows participants to focus on individual diagnostic dimensions while ignoring irrelevant features within the task. This allows for accurate categorization when within-category variance is high and between-category variance is low. However, when compared to the perceptual similarity system, rule-based categorization is cognitively expensive and sensitive to distractions (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006).

Multiple Memory Systems Within a Single Task

To show dissociable signatures of the memory systems discussed earlier, traditional memory system research has been conducted using tasks that aim to engage each system in isolation and then demonstrating that aggregate behavior and neural data are more consistent with properties of one system over another. However, the composition of natural categories contains elements of exemplar memorization, rule-, *and* perceptual-based systems, suggesting people may be switching between systems within a single task. For example, relying on perceptual similarity to categorize mammals may be quick and effortless for many examples of the category, but would likely sacrifice accuracy when categorizing bats or dolphins – mammals that may be mistakenly classified as birds or fish when using perceptual similarity. An overarching goal of my research is to devise ways to improve the learning process by understanding and exploiting the underlying neural processes that dictate success in each task. This goal begins with an empirical investigation of how the brain learns under realistic conditions, while respecting the potential for the brain to utilize different memory systems that detract from a-priori models of typical learning.

Neuroimaging

To achieve a comprehensive picture of the memory systems involved in a realistic learning task, we must describe which systems are engaged (“where” in the brain these systems are located) and the temporal dynamics under which these systems are employed (“when” these systems are engaged). Brain activity is commonly measured using two noninvasive neuroimaging methods: Electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI). EEG measures the electrical activity of brain cells by placing a network of electrodes on the scalp. With enough (256) channels,

we call this dense array or “high density” EEG (dEEG) (Tucker, 1993). When EEG is recorded during a task, small changes in voltage can correlate with specific operations within the task, called an Event Related Potential (ERP). Small experimental manipulations or changes in behavior can have a measurable impact on ERPs and, as we will discuss in the next chapter, can allow us to track learning. The pros of recording EEG is that it picks up brain activity with high temporal resolution, down to one millisecond. However, EEG can be non-specific and records the activity of tens of thousands of brain cells at once. Moreover, due to differences in cell structure in different brain parts, EEG can only record activity from the cerebral cortex, yet we know that the majority of memory systems involve subcortical areas in some way.

Some of the spatial shortcomings of EEG are made-up for using fMRI. The brain is constantly being fed oxygen through its matrix of vasculature, and when neurons fire oxygen is stripped from hemoglobin (a protein in red blood cells that carries oxygen) until subsequent cardiac events occur to resupply the brain with oxygenated hemoglobin. fMRI can detect the subtle difference between oxygenated and de-oxygenated hemoglobin while a subject performs a task, and when we superimpose the map of where oxygen exchange is occurring over a structural image of the brain, we get a map of brain activity. This map is only limited by the presence of blood vessels, which is luckily very dense, and can image activity in deeper brain regions than EEG. However, blood flow in the brain is substantially slower than the electrical events happening between neurons, and fMRI is stuck measuring activity 5-10 seconds after a neural event has occurred. This makes it difficult for fMRI to tease-apart brain activity that occurs below the ~7 second timescale in complex tasks that involve several steps to complete the task. Given the

spatial shortcomings of EEG and the temporal limitations of fMRI, this dissertation will employ both methods to understand the spatial and temporal dynamics of multiple memory systems under realistic learning conditions.

Overview of Dissertation

In Chapter II of this dissertation we will focus on using Electroencephalography (EEG) to examine the mechanisms associated with the different stages of learning within a categorization task. We will then discuss the compatibility of our results with the general models of how the brain behaves during the different stages of learning. Chapter III will build off of the findings in Chapter II by using functional Magnetic Resonance Imaging (fMRI) to determine the extent to which multiple distinct memory systems are engaged during the task. We will follow up in Chapter IV by using the results in Chapter III to guide an effort to dissociate between different memory systems in EEG. We will use several different techniques, including standard Event Related Potential (ERP) analysis and data-driven machine learning, to understand the time course by which different memory systems are engaged. We will conclude the dissertation in Chapter V with a discussion of how single tasks may be supported by different cognitive and neural systems, along with a look forward into using the brain as a guide for making targeted interventions to improve the learning process.

CHAPTER II

TRACKING THE ONSET OF AUTOMATICITY IN A FOOTBALL CATEGORY LEARNING TASK

From Morgan, K.K., Luu, P., Tucker, D.M., (2016.) Changes in p3b latency and amplitude reflect expertise acquisition in a football visuomotor learning task.

PLoS One, 11, e0154021. <https://doi.org/10.1371/journal.pone.0154021>

Introduction

Earlier we discussed a general neurophysiological framework for the dual-stage model of learning, whereby the early stage of learning is marked by a reliance on frontal control regions, and the late stage of learning relies on more posterior consolidation areas. The Medial Frontal Negativity (MFN) and P300 Event Related Potentials (ERPs) obtained through EEG are of particular interest for marking the transition from early to late learning. The MFN is a stimulus-locked medial frontal component with its primary sources in the Anterior Cingulate Cortex (ACC) (Luu et al., 2011). The ACC plays a major role in error monitoring and attention during reward-based learning (Bush et al., 2002). In theory, this would make the MFN an ideal component for indexing effortful control seen in the early stage of category learning.

The P300 can vary in its topographic distribution as well as the conditions under which it is evoked, and it is now recognized that there is a family of P300 components (Squires, Squires, & Hillyard, 1975). The P3a, which has a mediofrontal scalp distribution, is commonly evoked during a 3-stimulus oddball task when participants are exposed to infrequent, novel (non-target) stimuli (Ebmeier et al., 1995; Friedman,

Simpson, & Hamberger, 1993; Katayama & Polich, 1998). Of more relevance to the current study is the P3b, which is traditionally found over more parietal scalp sites, and occurs within the same oddball task but in response to stimuli that require an action (such as a response or silent count) (Katayama & Polich, 1998).

Conventionally, the P3a is thought to reflect the attentional shift caused by the mismatch between a novel stimulus in a series of expected stimuli, whereas the P3b reflects the match between a stimulus and the voluntarily sustained attentional trace (Näätänen, 1990). However, this popular theory for the P3b and voluntary attention cannot fully explain the results of several previous studies which showed a linear increase in P3b amplitude correlating with the acquisition of a response mapping to the point of expert performance (Donchin & Coles, 1988; Knight, 1996; Polich, 2004; Pineda, Westerfield, Kronenberg, & Kubrin, 1997; Luft, Takase, & Bhattacharya, 2014). It is generally accepted that attention decreases with expertise, and thus if the P3b were a reflection of controlled attention a decrease in amplitude when participants approach expert performance is expected (Anderson, 1982; Anderson, 1983; Anderson, 1993; Fitts & Posner, 1967). Results from previous studies in our lab, wherein P3b amplitude continued to increase as participants transitioned from novice to more automated performance, is more consistent with the context updating theory of the P3b (Polich, 2007; Luu et al., 2011).

Under context-updating theory, the P3b indexes the updating and/or confirmation of the context under which an action is learned and performed on a trial-by-trial basis (Katayama & Polich, 1998; Donchin & Coles, 1988; Polich, 2007). The context can be information pertaining to the rules of a task, or even the environment under which

knowledge was acquired. Relevant to the dual-stage model of learning, the early stage aids in the formation of this context (not indexed by the P3b) and the posterior corticolimbic system maintains it. Once the context is formed, the constant maintenance and recall of this information helps to guide a person toward selecting the correct action in response to a stimulus quickly and efficiently. The P3b reflects the time-course by which the context is updated and the processing resources that were available when the context was referenced (Luu et al., 2011). It is important to note, however, that the P3b does not directly reflect memory retrieval (evidenced by its time-course), but is instead an indirect correlate. The sources of the P3b remains to be definitively resolved. However, results from scalp EEG source localization studies as well as data from human intracranial EEG (iEEG) and animal studies revealed common P3b sources: the parietal lobe, PCC, medial temporal lobe, and superior temporal sulcus (Halgren et al., 1994; Halgren et al., 1995a; Halgren et al., 1995b; Baudena, Halgren, Heit, & Clarke, 1995; Smith et al., 1990; Brankack, Seidenbecher, & Muller-Gartner, 1996; Shin, 2011; Kahana, Seelig, & Madsen, 2001).

Previously we performed three dense-array EEG (dEEG) studies focused on the dual-stage theory of learning using an arbitrary categorization association task (Luu, Tucker, & Stripling, 2007; Luu, Shane, Pratt, & Tucker, 2009; Luu et al., 2011). In these studies, we used a Go/No-Go task that required participants to learn arbitrary category structures to form an appropriate action (Wise & Murray, 2000). The participants were tasked with associating a simple visual stimulus (numbers) with a specific button press on a key pad. Our results demonstrated that increases in both P3b *and* MFN amplitude reflected performance improvements as participants achieved task proficiency and

reached behavioral automaticity in the late learning stage. This result was contrary to the anterior-posterior shift that commonly describes a transition between novice to expert performance.

The goals of the present study were to extend the previous findings by examining the MFN and P3b's relations to behavioral performance measures across the stages of learning in a more realistic learning task and determine how well our results fit into the general anterior-posterior framework describing transitions in learning stages. To pursue these questions, we tracked the MFN and P3b ERP components as our participants were subjected to a multi-day, modified Go/No-Go task that is similar to the cognitive training program used by the varsity football team at the University of Oregon to help new players acquire the playbook. In this task, participants were presented with defensive football formations as viewed from the quarterback's perspective. Participants were responsible for acquiring the proper stimulus-response mappings that help them determine which defensive formations require input from the quarterback (target formations, or "Go trials"), and which formations do not require any intervention (non-target formations, or "No-Go trials").

We hypothesized that our participants would be proficient in the task by the end of the first day of training, and that an increase in P3b amplitude and a decrease in MFN amplitude would mark the onset of expertise in the task. We also hypothesized that the onset of full automated performance and cognition would occur during the first day of training, and that changes in the P3b would parallel the performance improvements that occur during this stage (e.g. reductions in errors and reaction times). With further training in the subsequent days, we hypothesized that the P3b would continue to track small

performance improvements after the task became well learned, with the MFN continuing to decrease in amplitude as effortful control continues to decrease under automaticity.

Method

Participants

Fifteen right-handed participants were recruited from the University of Oregon Human Subjects Pool (eight males, seven females), with ages between 18 and 41 years ($M = 23$, $SD = 6$). All participants had normal or correct-to-normal vision, had no history of head trauma or seizures, and were not consuming medication that could affect their EEG. Participants were pre-screened online for their experience with football in order to reduce the chance of contextual familiarity confounding differences in skill acquisition rate. Only the participants who were comfortable recognizing variations in defensive and offensive football formations (e.g. participants who had a history of playing football, or were an avid fan of the game) were qualified to participate. Before each session, participants provided informed written consent and filled out several mood questionnaires. The mood questionnaires were not used for analysis, but were collected as part of a standard lab procedure in the case that they might be useful if a participant displayed adverse behavior during the study. Data from all participants who completed all 3 days of the study were included in our analyses. The research protocol was approved by the University of Oregon and Electrical Geodesics, Inc. (EGI) institutional review boards, and the study took place in the Brain Electrophysiology Laboratory at EGI.

Task

The task used in this study was adapted to resemble the cognitive training program used at the University of Oregon to aid new football recruits in learning the

playbook and familiarizing themselves with an opponent's playing style (Axon Sports, LLC, Phoenix, AZ). Likewise, the paradigm was a modified version of a traditional go/no-go discrimination task (Newman, Patterson, Howland, & Nichols, 1990). On each trial, 1 of 8 defensive formations were presented centrally on a 43 cm (diagonal) computer monitor for 1500 ms. Half of the formations were randomly selected as "go" stimuli, and the other half were designated as "no-go" stimuli. The formations were presented at random, with the restriction that a formation could not be presented twice in a row. Participants had to either press, or refrain from pressing, a key on a keypad when a formation was presented. For the go stimuli, participants had to learn to respond with the appropriate digit on the correct hand for each stimulus. The participants were given four digits to respond with (digits I and II of both hands), and each of the four go stimuli were mapped onto a specific digit. Each formation was presented on the screen for 1500 ms or until a key-stroke was made. Immediately following a response (or non-response), contingent feedback was shown for 10 s or until the participant made another keystroke. Upon feedback termination, a fixation mark was shown for the duration of the inter-trial interval of 1500 - 2500 ms before the next formation was presented.

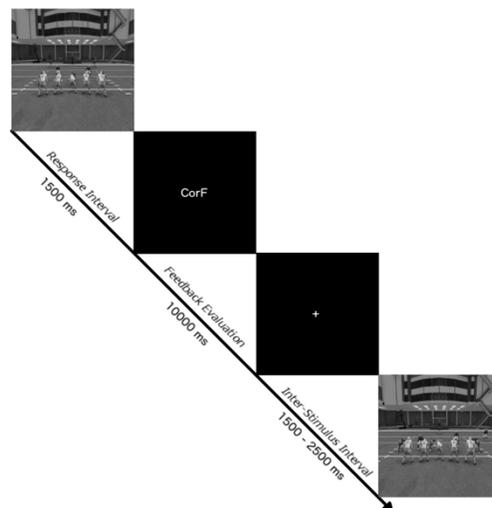


Fig. 1. Diagram of events in a single trial for experiment 1. Formations were shown for 1500 ms or until a key-stroke was made. Immediately following a response (or non-response), contingent feedback was shown for 10 s or until the participant made another keystroke. Upon feedback termination, a fixation mark was shown for the duration of the inter-trial interval of 1500 - 2500 ms before the next formation was presented

Immediately after each trial, specific feedback about performance on that trial was provided (Fig. 1).

The feedback given to the participant were designed to provide them all of the information needed to learn the response mapping. In sum, there were six types of

feedback: ErrorGo (error of omission), ErrorNG (error of commission), Correct (correct response to a go stimulus but made with the wrong hand), CorH (correct response to a go stimulus with the correct hand but wrong digit), CorNG (correct withholding of a response to a no go stimulus), CorF (correct response to a go stimulus with the correct hand and digit). Feedback were presented for 10 s, or until the participant terminated the feedback with a button press. Upon termination of the feedback, the next trial began between 1500 and 2500 ms later (Fig. 1).

To motivate participants to continuously improve on the task, they were made aware that they would be compensated based off of their study performance. To track performance, point values were assigned to each contingent feedback: correct responses (CorNG and CorF) would earn them eight points, whereas errors (ErrorNG and ErrorGo) would lose them eight points. Partly-correct responses (Correct and CorH) would lose a participant four and two points. Participants were informed that they would be able to track their score across each block, and that their final score will determine how much they are compensated on each day. Participants started with a score of zero, and were explicitly told how their point total will be converted into their compensation rate (\$15 - 45).

Procedure

Following the informed consent process, participants were fitted with a 256-channel HydroCel Geodesic Sensor Net (HCGSN) and placed 55 cm in front of the computer monitor. A chin-rest was used to minimize head movements and keep the distance to the monitor fixed. Participants were explicitly told that there were 8 defensive formations in this study, and that they must learn which formations require a specific key

stroke and which formations require them to make no response at all. To add relevant context to the learning environment, the act of pressing a button corresponded to the quarterback's decision to "hike" the football in response to a target formation. Similarly, an inaction corresponded to the quarterback's decision to not hike the ball, and instead could be assimilated to halting gameplay (e.g. calling a timeout, or pausing to change a play at the last second). The response feedback that would help teach the participant to make the correct decision were explained clearly on a piece of paper, and participants were allowed to look over the feedback for several minutes.

Once the participant could demonstrate an understanding of the feedback to the research assistant, a short practice block consisting of 30 trials followed. Formations used in the practice block were not used in the actual experimental blocks. For the experiment, 8 blocks of 100 stimuli (800 trials per session) were used. Each participant underwent 3 training sessions, and each session was scheduled exactly 48 hours apart within the same week (Monday, Wednesday, and Friday). The practice block was only given during the first session, and on average each session lasted around 2.5 H. All participants displayed proficiency in the task within the first session, and were compensated an average of \$40 for each session.

Learning Criterion

To simplify the analysis process, we used the fixed-number of consecutive responses method (FCCR) in order to determine when a participant had sufficiently acquired the response mapping as we have done in the past (Luu et al., 2007). With this method, a subject fulfilled the learning criterion when they could make four correct responses (or non-responses) in a row for each stimulus. Because the time before this

learning criterion was met is a period where participants could not differentiate between whether they needed to withhold or make a correct hand-finger response for a given stimulus, all trials preceding the fulfillment of this criterion were included in a "pre-learning" condition (this includes all trials where errors were committed, for both Go and No-go stimuli). However, because we are only concerned with how a subject acquires and demonstrates a response *mapping* and not response *inhibition*, only the go-trials where the participant provided a fully correct response (CorF) were included in a "post-learning" condition after the learning criterion was fulfilled.

EEG Recording and Post-Processing

The dEEG was recorded using a 256-channel HydroCel Geodesic Sensor Net and the data were amplified using a Net Amps 400 Amplifier (Electrical Geodesics, Inc., Eugene, OR). Recordings were referenced to Cz and impedances were maintained below 50 k Ω . dEEG was bandpass filtered (0.1 - 100 Hz) upon being sampled at 250 s/s with a 16-bit analog-to-digital converter.

After recording, signals were filtered between .1 - 30 Hz bandpass and segmented into 1200 ms long segments time-locked to the onset of each stimulus (segments extended 200 ms before and 1000 ms after the stimulus onset). Segments containing eyeblinks, muscle tension, major eye movements, or large head movements with 10 or more channels exceeding an absolute voltage threshold of 140 μ V were excluded from a participant's average. Segments containing minor eye movements (saccades) were not fully rejected given the lack of overlap between the latency and distribution of the saccades with the latency and location of the MFN and P3b. All data were re-referenced to the average reference for analysis.

EEG Source Analysis

Source analysis was performed using GeoSource (version 2.0) software (Electrical Geodesics, Inc., Eugene, OR). The software relies on the MRI and CT scan of a single subject (Colin 27) to construct an atlas model of the brain and head that is used to estimate the sources of scalp EEG. The brain (gray and white matter) and cerebrospinal fluid (CSF) are segmented as they appear in the MRI, whereas the skull and skin surfaces are characterized from the CT. These two volumes are then co-registered together. Once registered, the gray matter tissue is parceled into 7 mm voxels which serve as individual source locations with three orthogonal orientations, resulting in 2,394 triples sources. Following the construction of the head model, averaged 256 sensor-locations are then registered to the scalp surface.

A Finite Difference Method (FDM) is used to compute an estimate of how current propagates from the sources in the cortical gray matter to the scalp where EEG is measured. Conductivity values used in the FDM were: 0.25 S/m for the brain, 1.8 S/m for CSF, 0.018 S/m for skull, and 0.44 S/m for scalp (Ferree, Eriksen, & Tucker, 2000). The local autoregressive average (LAURA) constraint was used to compute inverse source estimates (Grave de Paralta Menendez et al., 2004).

Results

Behavioral

Learning Effects

We refer to "learning effects" as effects occurring within the learning process during the first session. A paired samples t-test was run on the number of trials it took each participant to learn the response-mappings, separated by stimulus type ("Go" vs "No

Go"). A significant effect was found, $t(14) = 5.3, p < .001$, such that Go stimuli took longer to acquire than No Go stimuli. A summary of this effect can be found in Fig. 2.

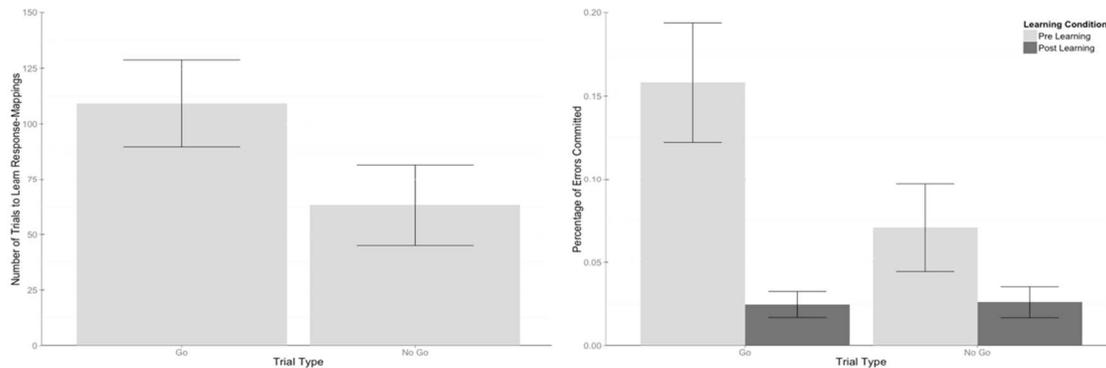


Fig. 2. Graphs of trials to learn by trial type and error rate by trial type and learning condition. (Left) One average, Go stimuli took significantly longer to learn than No Go stimuli. (Right) Error rates for both types of stimuli significantly decreased after the learning criterion was met. However, participants made more errors with Go stimuli than No Go stimuli during the learning period. Once the learning criterion was met, there were no differences in the number of errors committed between the stimulus types.

For the error rate calculation, an error was quantified as an error of commission for No Go trials, and any response or non-response that was not fully correct for Go trials (i.e. errors of omission, correct responses to a Go stimulus with the incorrect hand, and correct responses to a Go stimulus with the correct hand but incorrect digit). For each participant, errors were counted for the period before and after the learning criterion was met during the first session only, as all participants acquired the task during the first session.

In a repeated measures ANOVA which used trial type and learning condition (pre-learning and post-learning) as within-subject factors, significant main effects of trial type ($F(1,14) = 30.72, p < .001$), and learning condition ($F(1, 14) = 66.11, p < .001$) were found. The nature of these effects show that error rates (collapsing across trial type) were significantly lower after the learning criterion was fulfilled, and that more errors were committed for Go stimuli than No Go stimuli before participants acquired the response-

mapping (Fig 2.). Additionally, a significant interaction between these two factors was found, demonstrating that there was no difference between Go and No Go errors after the learning criterion was satisfied, $F(1, 14) = 35.3, p < .001$.

Training Effects

All participants sufficiently acquired the task during the first half of the first session and did not commit enough errors during sessions 2 and 3 to define a secondary or tertiary learning period. Due to this, we labeled days 2 and 3 as full training sessions throughout all of our analyses, where we assume most correct responses performed during these days were a result of a participant's knowledge and expertise in the task, and not due to chance as they may have been during the pre-learning period during day 1. We refer to "training effects" as effects occurring after participants satisfied the learning criterion during sessions 1 - 3, accordingly. When computing training effects, only the post-learning data from day 1 were used for comparison.

Trial type and training session (Days 1 - 3) served as within-subject factors in a repeated measures ANOVA which evaluated error rates across days. Significant main effects for trial type ($F(1, 14) = 18.74, p < .001$) and training session ($F(1.23, 17.18) = 31.49, p < .001$, Greenhouse-Geisser corrected) were found. The effects show that error rates decreased with practice, and more errors were committed for Go stimuli than No Go stimuli. A significant interaction between trial type and practice session was also found, $F(1.31, 18.31) = 7.44, p = .009$ (Greenhouse-Geisser corrected), which showed more errors committed for Go trials compared to No Go trials on days 1 and 3, but no difference on day 2, (Fig. 3). When collapsing across trial type, significant linear ($F(1, 14) = 53.955, p < .001$) and quadratic ($F(1, 14) = 9.02, p = .006$) trends were found (Day

1 $M = 14\%$, Day 2 $M = 5.6\%$, and Day 3 $M = 4.2\%$). A comparison of the means shows significant differences in the errors committed on the first day compared to the second and third days ($t(28) = 6.27, p < .001$ & $t(28) = 7.35, p < .001$, respectively), but no significant difference in errors between the second and third days ($t(28) = 1.07, p = .54$).

For the reaction time (RT) analysis, RTs for trials where participants made a

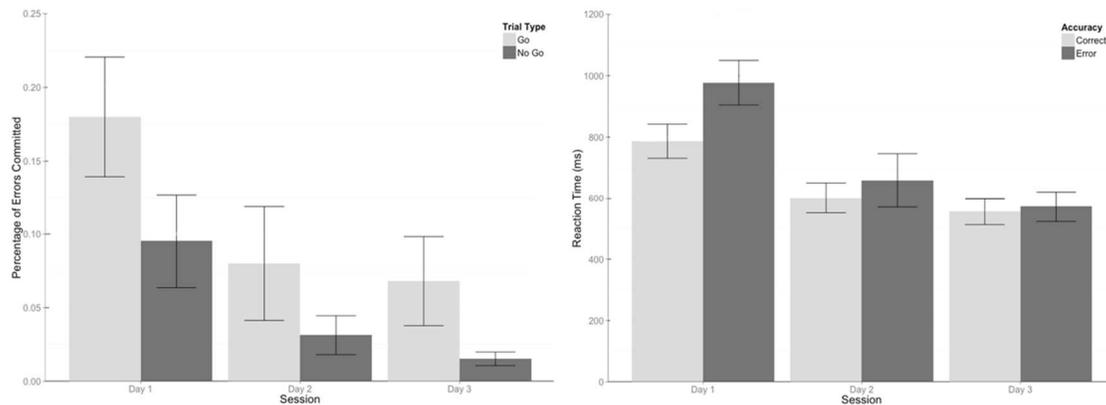


Fig. 3. Graphs of error rates and reaction times across training days. (Left) Error rates decreased with practice, and the number of errors were greater for Go stimuli than No Go stimuli for all sessions except for day 2. (Right) Reaction times (RT) decreased with training. In addition, RTs for correct responses were quicker than those for incorrect responses. However, there was only a difference in RT for correct and incorrect responses during the first training session, suggesting that errors made during this day may differ in nature than those committed during subsequent training days. This interaction corresponds to the session where learning was achieved for all participants.

correct response (CorF) were labeled as "Correct", and RTs for trials where participants made an incorrect response (errors of commission, correct responses to a Go stimulus committed with the incorrect hand, and correct responses to a Go stimulus committed with the correct hand but incorrect digit) were labeled as "Errors" in an accuracy category.

Accuracy and training session were included as within-subject factors. Significant main effects of accuracy ($F(1, 14) = 65.77, p < .001$) and training session ($F(1, 28) = 164.85, p < .001$), along with an interaction between the two ($F(2, 28) = 22.70, p = .009$) were found. A mean inspection shows that RTs were significantly shorter for correct responses compared to errors (Correct $M = 647.93$, Errors $M = 735.67$). However, RT

differences for trial accuracy were only significantly different during the first training session (Fig. 3). This interaction suggests that the nature of errors committed during the first training session (where learning occurred) may differ than those which occurred in the remaining training days. Collapsing across accuracy, RTs decrease in a significant linear trend, $F(1, 14) = 295.72, p < .001$, Day 1 $M = 881.88$, Day 2 $M = 629.59$, Day 3 $M = 563.95$. Significant differences in all pair-wise comparisons of these RTs were found, Day 1 v. Day 2: $t(28) = 13.65, p < .001$, Day 1 v. Day 3: $t(28) = 17.20, p < .001$, and Day 2 v. Day 3: $t(28) = 3.60, p = .004$.

Event Related Potentials

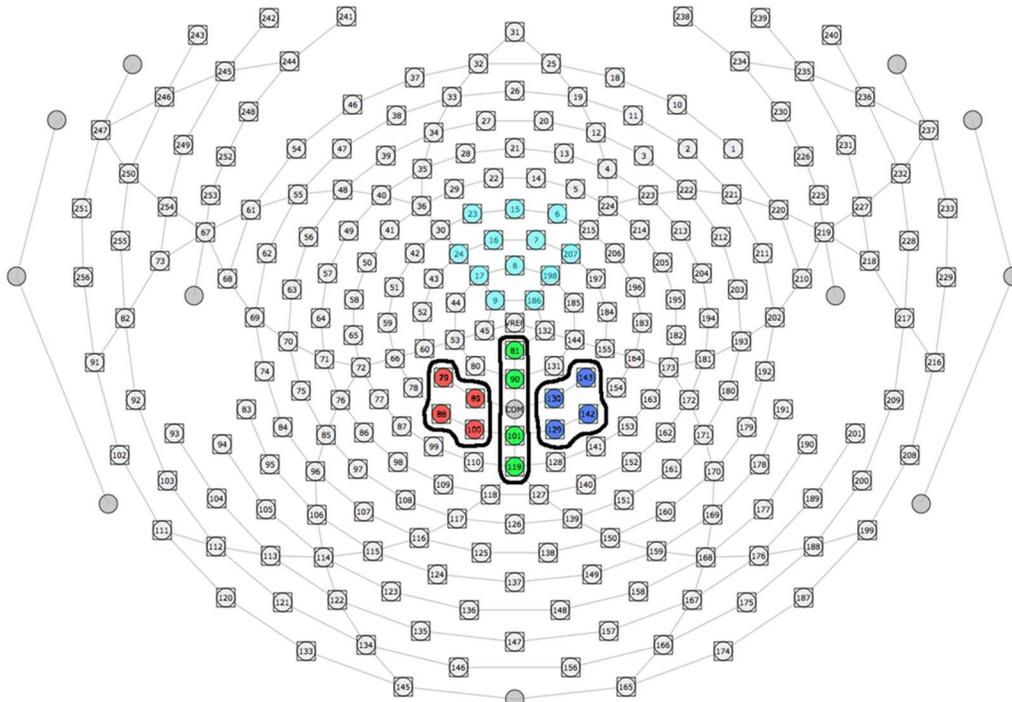


Fig. 4. Channel montages used to quantify the MFN and P3b. Electrodes used to quantify the MFN are highlighted in cyan. 3 distinct groups of channels were used to separate the P3b component by laterality (left = red, midline = green, blue = right).

MFN

For the MFN analysis, a cluster of electrodes that best represent the medial frontal distribution of the component were chosen (Fig 4.). To quantify the MFN, an adaptive mean amplitude corresponding to 20 ms before and 20 ms after the maximum negative peak amplitude in a window extending from approximately 180 – 300 ms after stimulus onset was computed for the MFN electrode cluster (blue windows in Fig. 5 and Fig. 6). The MFN was referenced to the nearest positive peak (P200) around 150 – 200 ms after stimulus onset. This method was applied to each individual participant and condition.

MFN Learning Effects. A paired samples t-test was run to evaluate differences in MFN amplitude before and after learning occurred during the first training session. The test revealed a significant effect, such that the amplitude of the MFN decreased after participants acquired the task (pre-learning $M = -5.85 \mu\text{V}$, post-learning $M = -4.59 \mu\text{V}$), $t(14) = 3.76$, $p = .002$.

This effect is illustrated in Fig. 5.

MFN Training Effects. To evaluate MFN training effects, only post-learning MFN amplitudes were used from day 1 and compared across the subsequent 2 days of training.

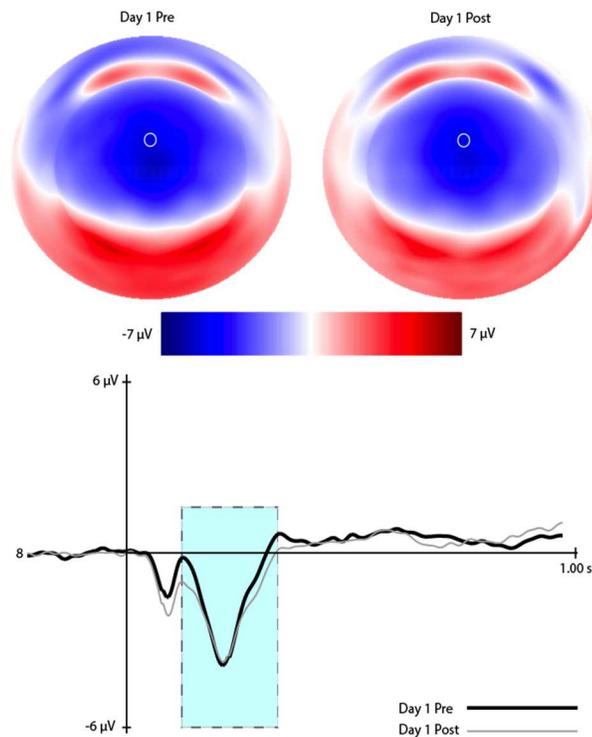


Fig. 5. Voltage maps and waveforms of the MFN from representative channels for pre- and post- learning conditions. (Top) Voltage distributions of the MFN across the scalp for both learning conditions. White circles represent the location of the representative channel shown in the bottom of the figure. (Bottom) MFN waveform (blue window) displayed for both learning conditions.

A repeated measures ANOVA which used training day as a within-subjects factor revealed no significant effect of training day on MFN amplitude, $F(2, 28) = 1.3, p = .29$. However, looking at the means shows a slight increase in MFN amplitude across days, as opposed to the hypothesized decrease (Day 1 $M = -4.59 \mu\text{V}$, Day 2 $M = -5.16 \mu\text{V}$, Day 3 $M = -5.26 \mu\text{V}$), Fig. 6.

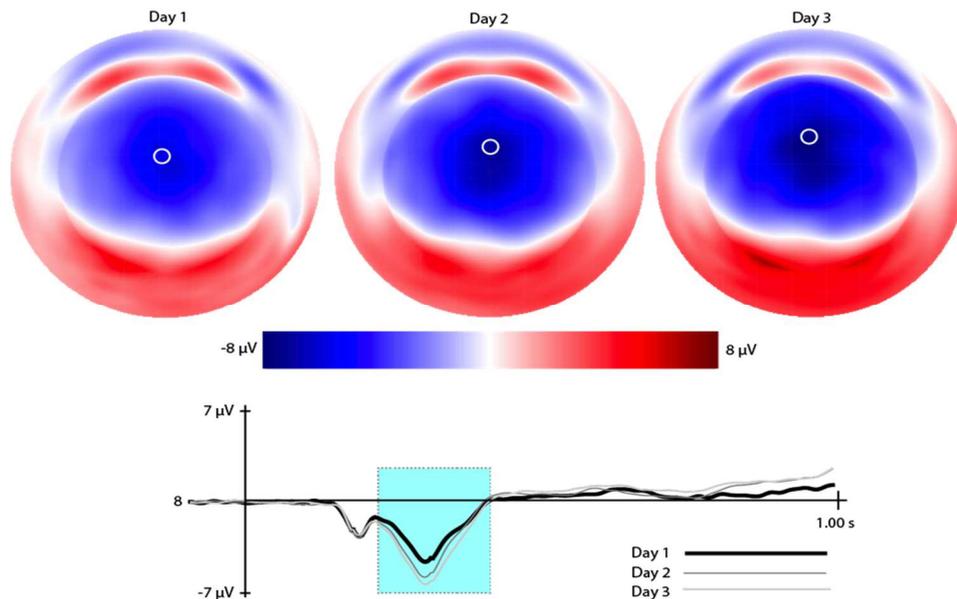


Fig. 6. Voltage maps and waveforms of the MFN from representative channel for all training days. (Top) Voltage distributions of the MFN across the scalp for all training days. Clear differences in negative energy can be seen around the medial frontal region. White circle represents the location of the representative channel shown on the bottom of the figure. (Bottom) MFN waveform (blue window) displayed by representative channel from each training day. Clear amplitude differences can be seen across training sessions.

P3b

For the P3b analysis, three sets of channels corresponding to laterality (left, midline, and right) were used to evaluate differences in P3b scalp topography based off of similar electrode sites chosen in our previous studies (Fig 4.) (Luu et al., 2007; Luu et al., 2011). To quantify the P3b, an adaptive mean amplitude corresponding to 22 ms before and after the maximum peak amplitude in a window extending from approximately 450 - 950 ms after stimulus onset was computed for each separate channel

grouping (red windows in Fig. 7 and Fig. 8). This method was applied to each individual participant and condition, so that small variations in the latency of the P3b were considered.

P3b Learning Effects. A repeated measures ANOVA was performed with laterality (left, middle, and right) and learning condition (pre-learning and post-learning) as within-subjects factors. A significant main effect of learning condition was found, ($F(1, 14) = 94.43, p < .001$). No other main effects or interactions were found. The analysis shows that P3b amplitude was much larger after learning had occurred, and did not differentiate by hemisphere (Fig. 7).

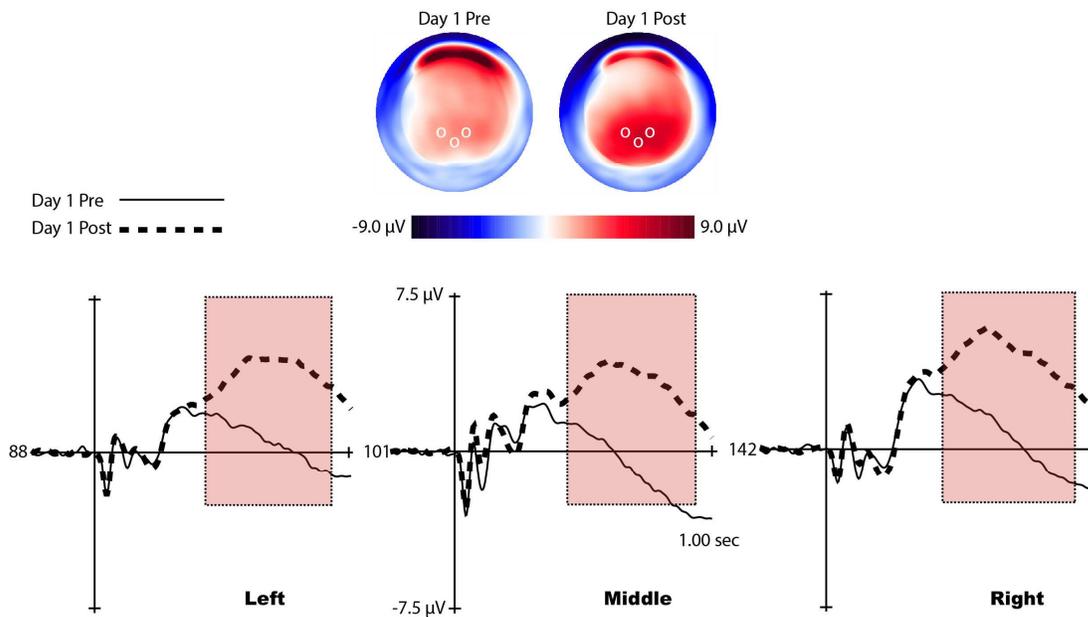


Fig. 7. Voltage maps and waveforms of the P3b from representative channels for pre- and post-learning conditions. (Top) Voltage distributions of the P3b across the scalp for both learning conditions. Clear differences in positive energy can be seen around the occipital region. White circles represent the location of the representative channels shown in the bottom of the figure. (Bottom) P3b waveform (red window) displayed by representative channels from each laterality condition. Clear amplitude differences can be seen between learning conditions.

P3b Training Effects. To evaluate P3b training effects, only post-learning P3b amplitude measurements were used from day 1. Training session (1, 2, and 3) and laterality (left, middle, and right) were included as within-subjects factors. A significant main effect of training session ($F(2, 28) = 7.35, p = .002$) was identified, no other main effects or interactions reached statistical significance. A trend analysis reveals a significant linear trend in P3b amplitude with practice when controlling for laterality, $F(1, 14) = 14.42, p < .001$. An inspection of our means shows that this trend is positive (Day 1 $M = 5.63 \mu\text{V}$, Day 2 $M = 6.67 \mu\text{V}$, Day 3 $M = 7.30 \mu\text{V}$), (Fig. 8).

Differences in peak P3b latency were computed through identifying the largest positive peak between 450 - 950 ms after stimulus onset and recording the segment time of the maximum amplitude (red windows in Fig. 8). Peak latency was not computed for the pre-learning P3b waveform during the first session because the shape of the P3b did not present a reliable "peak" to accurately perform the analysis. Instead, peak latency was found for the post-learning condition on session 1 and the subsequent training days.

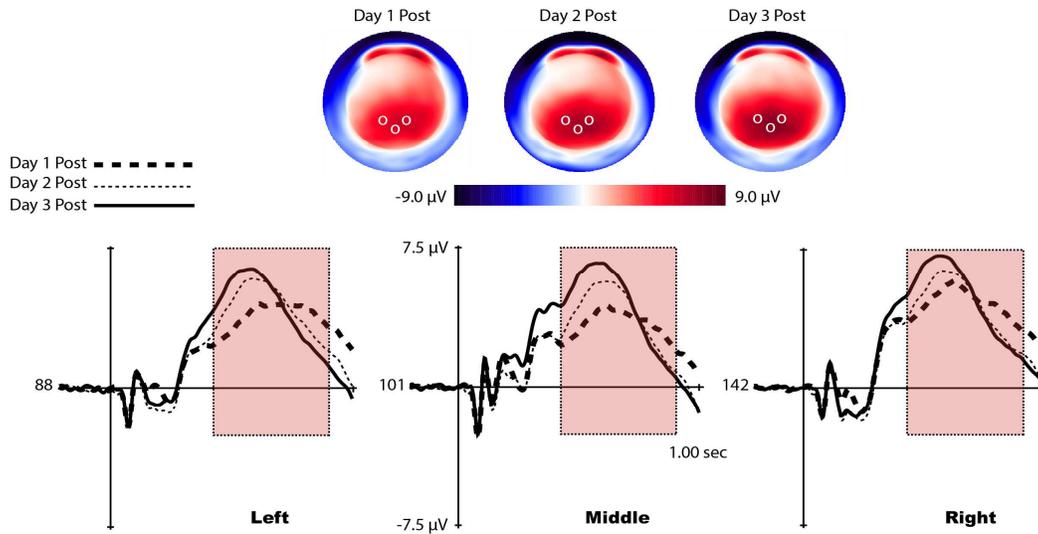


Fig. 8. Voltage maps and waveforms of the P3b from representative channels for all training days. (Top) Voltage distributions of the P3b across the scalp for all training days. Clear differences in positive energy can be seen around the occipital region. White circles represent the location of the representative channels shown in the bottom of the figure. (Bottom) P3b waveform (red window) displayed by representative channels from each laterality condition. Clear amplitude differences can be seen across training sessions.

Laterality and training session served as within-subject factors. A significant main effect of training session was found, $F(1.08, 15.15) = 9.93, p = .006$ (Greenhouse-Geisser corrected). No other main effects or interactions reached significance, suggesting that P3b latency did not differ as a function of topography. Our training session effect shows a decrease in P3b peak latency with training (Day 1 $M = 703.22$, Day 2 $M = 620.53$, Day 3 $M = 585.91$), (Fig. 8). A trend analysis reveals that our latency decreased in a significant linear fashion, $F(1,14) = 18.87, p < .001$.

P3b Correlations to Behavior. Reaction times in our study appeared to provide the most convincing evidence of when a participant achieved expertise in the task among all other behavioral measures. Large decreases in RT were observed when a participant fulfilled the learning criterion, and they continued to decrease slowly with practice. Additionally, RTs provided a convincing parallel to the decreases in errors across training days, which

is the most commonly used measure of task performance. Given the reliability of RT, we focus on correlating our electrophysiological data with RT only.

The ratio of RT to the peak latency of the P3b (RT/P3b latency ratio) has been used as a measure of automated cognition (Lalonde, 1981; Donchin & Coles, 1988; Kramer & Strayer, 1988). Traditionally, the latency of the peak of the P3b is used to measure the amount of time a participant took to evaluate a stimulus, whereas their reaction time is a combined measure of how long it takes for the participant to evaluate, select, and execute a response to that stimulus. Fundamentally, the latency of the P3b cannot support this theory. Our own interpretation of this measure is consistent with context-updating, where the P3b reflects the updating or confirmation of contextually relevant information surrounding a stimulus. When evaluating changes in the RT/P3b latency ratio over time, significant reductions in this ratio indicate that response selection (RT) is moving closer to the updating or restoration of contexts; a process that closely follows response evaluation. This reduction indicates that responses come to fruition quicker as a result of automated cognition associated with the very late stages of learning (Kramer & Strayer, 1988). We would expect any significant reductions in the RT/P3b ratio to occur close to the fulfillment of the learning criterion during the initial training session, followed by a stabilization of the ratio across training days if the participants sufficiently acquired the task and reached automated cognition within the predicted time frame.

Due to the unreliable nature of interpreting single-trial ERPs, post-learning trials from the first training session were grouped into 4 equal bins for each participant. Separate ERPs were computed for each bin, resulting in 4 average, reliable ERPs per

participant. The peak of the P3b was computed by locating the largest positive peak between 450-950 ms after stimulus onset and recording the time of the maximum amplitude. The RT for each bin was then divided by the peak of the P3b for that bin (Fig. 7). A significant effect for bin number was found in a repeated-measures ANOVA, $F(3, 42) = 3.79, p = .02$. Polynomial contrasts revealed a significant linear trend, $F(1, 14) = 7.71, p = .008$, and a pairwise comparison using Tukey's method shows that there was only a difference in RT/P3b between the first bin and the third and fourth bins ($t(42) = 2.87, p = .04$, and $t(42) = 2.84, p = .03$, respectively). Results from this analysis suggest that the RT/P3b ratio decreased over time, and then stabilized toward the end of the first training session.

To help us evaluate whether automated cognition improved with extended training, the RT/P3b ratio for all post-learning correct responses during the first session were compared to the RT/P3b ratios of the subsequent training sessions. In a repeated-measures ANOVA which included training session as a within-subjects factor, a significant main effect for training session was found, $F(1.22, 16.03) = 8.62, p = .007$ (Greenhouse-Geisser corrected). Within this effect, a significant linear trend was discovered, such that the RT/P3b ratio further decreased with extended training, $F(1, 14) = 15.11, p < .001$. However, a post-hoc pairwise comparison of means (Tukey's HSD) showed that the biggest drops in RT/P3b ratio can be seen when comparing the first training session to the second and third ($t(28) = 3.21, p = .009$, and $t(28) = 3.89, p = .002$, respectively). No significant difference between the RT/P3b latency ratio between session 2 and 3 were found, which suggests the development of automated cognition on the task peaked following the initial training session (Fig. 9).

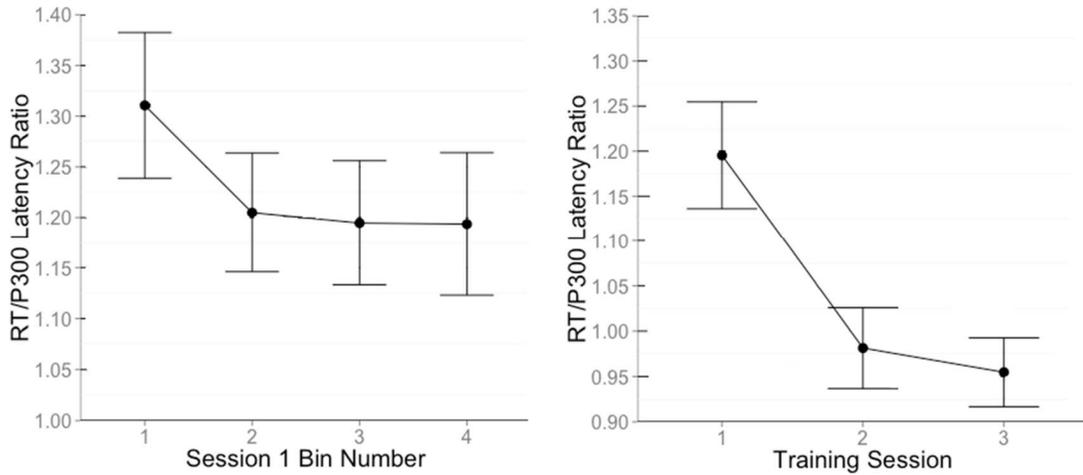


Fig. 9. RT/P3b latency ratio plots. (Left) RT/P3b latency ratio displayed for four consecutive bins during the first training session. The ratio decreases as participants progress through the first training session, and then levels-off as they reach the end. (Right) RT/P3b latency ratio for all three training sessions. The RT/P3b latency ratio drops significantly following the end of the first training session, but stabilizes with extended training.

Source Analysis

MFN

Due to ocular artifact contamination during the time course of the MFN, source analysis for the MFN is not reliable.

P3b

Pre-Learning. Source localization was performed using the LAURA constraint and a regularization constant of 10^{-3} on grand-averaged data of all 15 participants. Sources were obtained for the timepoint displaying the largest P3b amplitude for both pre- and post-learning conditions for day 1 (585 ms) (Fig. 7). Our analysis suggests sources of the P3b in Cuneus and Precuneus (BA7), Lingual Gyrus (BA18), and Fusiform Gyrus (BA37) for the pre-learning condition, and similar sources in the post-learning condition with the addition of Cingulate Gyrus (BA31) and Posterior Cingulate Cortex (BA23). The general absence of cingulate cortex activity in the pre-learning condition is important, as it reflects the lack of P3b presence during the early learning stage (yellow circles, Fig. 10).

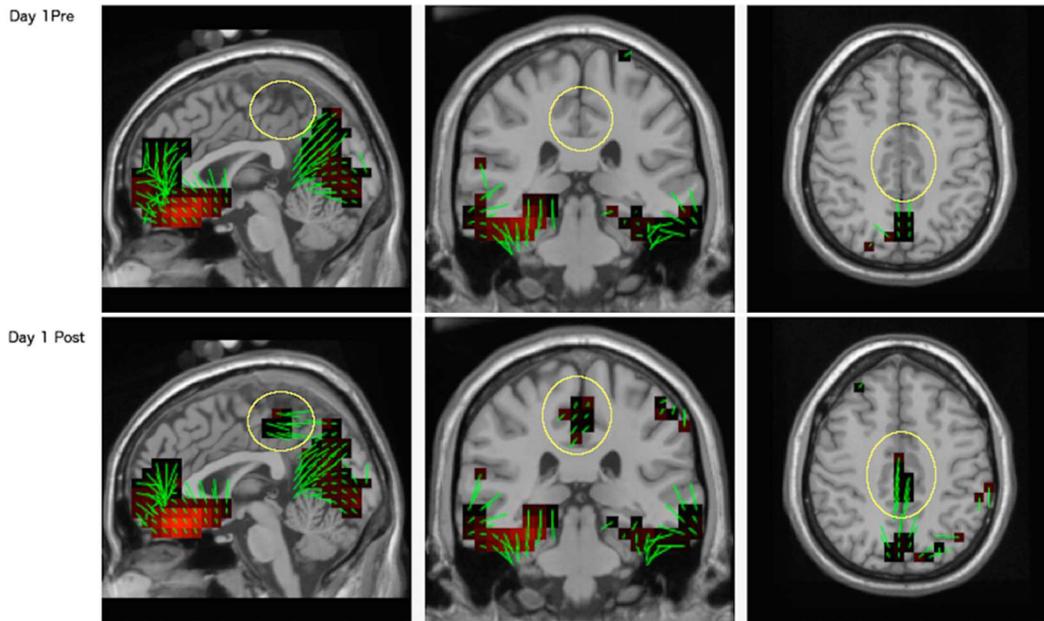


Fig. 10. Source estimates for the P3b during the learning session. Activity in posterior cingulate cortex (yellow circles) is absent during the period before participants acquired the response mappings, and is visible immediately after. Green lines at each voxel correspond to orientation vectors pointing in the positive direction.

Post-Learning. The post-learning P3b derived from the first training session was used in comparison to the P3b's obtained throughout the entirety of the subsequent training days. Sources of the maximum peak of the P3b (585 ms) are displayed in Fig. 9. All sources overlap with those found in Fig. 8, however the amount of cingulate cortex activity appears to differ as a function of training day and P3b amplitude (yellow circles, Fig. 11). Specifically, the PCC demonstrates greater engagement with practice, whereas the remaining sources do not reflect this increase.

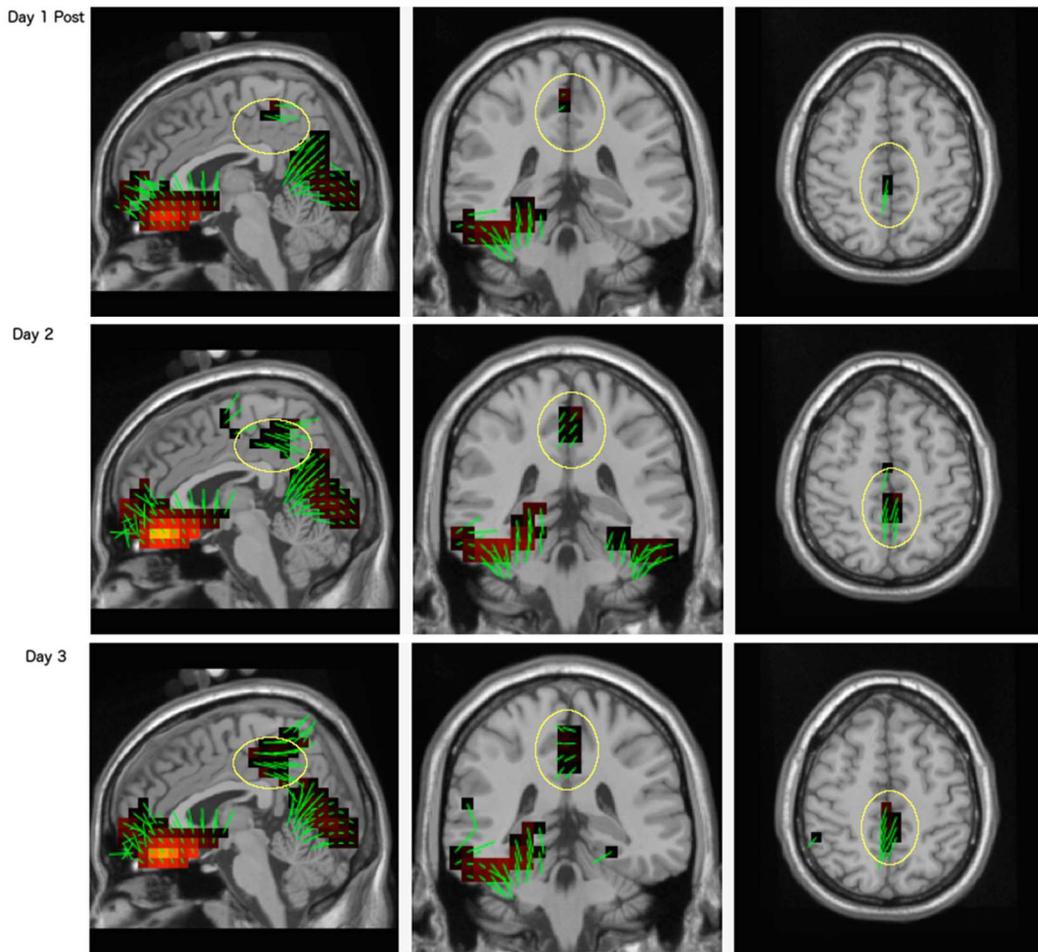


Fig. 11. Source estimates for the P3b across training sessions. Activity in posterior cingulate cortex (yellow circles) increases with training.

Discussion

The main goals of this study were to extend the relation between the MFN and P3b to learning of a realistic category learning task, in addition to framing the results within the dual model of learning and development of expertise. The results showed that participants learned the task, and their performance continued to improve with training. As participants' performance improved, MFN and P3b amplitude along with P3b latency closely tracked the improvements. Consistent with the canonical frontal-posterior shift during learning, the MFN and P3b marked the onset of learning in the first session with

the MFN amplitude decreasing with learning and the P3b amplitude increasing. Across training days, behavior and P3b amplitude continued to show a positive linear trend, which is consistent with our previous results. However, MFN amplitude *increased* with extended practice. Although this effect was not statistically significant, it poses questions for the role of the ACC during extended training and how our results fit into the general framework

Learning and Practice

The Go stimuli took significantly longer for our participants to acquire compared to No Go stimuli. This result can be explained by the fact that there were more ways to make an incorrect response for Go stimuli (4) compared to No Go (1). The larger variety of errors for Go compared to No Go can also explain the large margin of errors seen for the Go Stimuli over No Go stimuli in the initial training phase (pre-learning condition from session 1). However, once participants fulfilled the learning criterion, the difference in errors between our stimulus types diminished. This interaction is indicative of a successful acquisition of category associations.

The dual stage model of learning is supported by the behavioral data. Specifically, the magnitude of error rate and RT reductions between the first and second sessions and the much smaller difference in these two behavioral measures between the second and third sessions suggest a transition between early and late learning systems occurred (Shiffrin & Schneider, 1977; Chein & Schneider, 2005; Gabriel et al., 2002). Unlike error rates, which showed no statistical difference between the second and third sessions, a significant RT reduction was observed during this interval. During the first training session, incorrect response RTs were significantly longer than fully correct

response RTs. This difference decreased with training, as shown by a statistically significant training session x accuracy interaction. This implies that the nature of errors committed on the first day, when participants were learning the rules of the task, were not the same as those committed in the extended training days. The early learning stage is a time where controlled cognitive processing is most prominent, reflecting the trial-and-error strategy associated with early learning (Logan, 1988). The shorter RTs associated with errors after learning suggest that a different mode of performance is engaged, being supported by the later learning system. Late learning stage errors, in the present task, likely reflects impulsive responses because they are associated with faster RTs than correct responses, even though this difference did not reach statistical significance.

Event-Related Potentials

MFN

Our initial hypothesis for the MFN was based off well-replicated fMRI findings that showed decreased ACC activity as cognitive control decreased in later stages of learning. During the first training session our hypothesis was supported, such that the MFN (indexing ACC engagement) decreased in amplitude as participants reached proficiency in the task. However, a mean comparison across training days showed that, after learning occurred, the MFN *increased* in amplitude from day to day. Although not statistically significant perhaps due to a small sample size, the evidence for the increase with extended training was consistent and matched that of our previous findings (Luu et al., 2010). Yet, in our previous studies, the MFN increased in the initial training session as well, so why would the MFN detract from these findings in only the first session?

Folstein and Van Petten (2008) have demonstrated that there are two types of MFN (or N2) components, with one that indexes cognitive control and the other indexing the detection of novelty or mismatch. If the MFN seen in the current experiment was consistent with the cognitive-control MFN, we would have seen ours decrease across all training days. In contrast, if the MFN seen in the current experiment was consistent with mismatch or novelty monitoring we would expect it to increase. Previously we had formed a theory around this type of MFN which suggested the mid-ACC continuously monitors actions in relation to task parameters once the context for the response mapping has been solidified with training. It could be that the MFN seen in the initial training session is more reflective of the dorsal ACC establishing effortful control, and the MFN increase seen across days is a separate MFN altogether which indexes the mid-ACC's role in action monitoring. One piece of supporting evidence for an increasing MFN amplitude with extended training comes from recent work with mice that has shown frontal areas become more engaged *after* learning has occurred. During a visual change detection task, Stern et al. (2018) demonstrated that visual cortex responded to changing stimuli during the naïve learning stage while frontal areas did not. With training, responses from visual cortex stayed the same while the average response rates of frontal areas increased along with synchronized global engagement of the entire cortex in response to task-relevant stimuli.

An alternative theory for the oscillating MFN seen in our study could be that the MFN is indeed indexing cognitive control *or* action monitoring, but the nature of the stimuli used in our task may be making it improbable to differentiate between which one is present at a given time. The stimuli used in our task consisted of football defensive

formations that had a varying degree of spatial configurations. The strategies our participants employed to learn the response mappings varied dramatically based off of a debriefing questionnaire. Some participants used standard rote memorization, whereby the participants relied on constant repetition of the stimuli, while others developed their own explicit counting rules, and some attempted to compare the overall similarity of the formation shape to other members of the same category. Some participants used multiple strategies that depended on the category. It is possible that these strategies, although sharing a common goal (that is to learn the response mapping), are recruiting different memory systems each with a different reliance upon frontal cortical areas. For example, a visual similarity strategy has been shown to not rely as heavily on PFC compared to a strategy involved rule-application (Zeithamova & Maddox, 2006). The unreliable trend of the MFN could be the result of averaging the activity of multiple memory systems together. Yet, because the distribution of stimuli into each category was randomized and the strategies used by our participants varied so widely we cannot definitively test this theory. Relatedly, more variable reaction times during the first day of training could have led to an initial attenuation of the MFN amplitude during the first day of training. As behavior became more routinized and RT variance stabilized with extended training, the amplitude of the MFN increased. Reconciling these nuances will be essential for future interpretations of the MFN and how it relates to expertise.

P3b

Consistent with our previous results, P3b amplitude increased with learning and extended practice (Luu et al., 2007; Luu et al., 2009; Luu et al., 2011). In other studies, such as that performed by Barcelo et al., (2000), P3b amplitude was found to increase

with performance on a Wisconsin card sorting task and decreased whenever participants were required to learn a new rule. We hypothesize that the amplitude of the P3b observed in the current study and in previous studies is consistent with the dual stage learning model. Under this model, we propose that the P3b reflects activity of a cortical network which forms a representation of an environmental context that is consolidated with practice, and the involvement of this network increases in the later learning stage in order to reinforce actions based on an action context (Donchin & Coles, 1988; Katayama & Polich, 1998).

In the first session when the early learning system is expected to be strongly engaged, the P3b was not apparent until participants demonstrated that the task was acquired. In a previous study we showed a large P3b that is time-locked to the onset of the feedback in the early learning stage and an absence of the stimulus-locked P3b. Once participants acquired our previous task, the feedback-locked P3b diminished, and was followed by the presence of a stimulus-locked P3b (Luu, et al., 2003). In the present study, even with the appearance of the P3b in the first session after learning, the dramatic P3b amplitude increase as well as the reduced latency observed with continued training suggest that the late learning system becomes progressively more engaged. In the present study we infer that, when the context was formed during the transition to the late learning stage, the stimuli themselves became a part of the context representation such that actions are now supported as part of the context.

The process of context updating, wherein action is integrated to be part of the context, helps us interpret the observed reduction of the RT/P3b latency ratio. In the present study, RT/P3b latency ratio was found to decrease sharply with training and

gradually stabilized with extended practice. It has been shown that RT is dissociable from P3b latency, with P3b latency being indicative of evaluation speed whereas RT is the behavioral output affected by multiple cognitive processes. The RT/P3b latency ratio decrease with training suggests that the action is now more closely integrated with the contextual representation, requiring less involvement of other cognitive or brain processes.

Source Analysis

P3b

Previously, sources of the P3b have been localized to ventrolateral prefrontal cortex, posterior parietal cortex, temporoparietal junction, and inferior temporal cortex using combined fMRI and EEG (Bledowski et al., 2004; Bledowski et al., 2006). Although no statistical analyses were run, our source analysis in the present study suggest similar sources for the P3b in the lateral and medial areas of parietal cortex (BA7). In addition, we found sources in Lingual Gyrus (BA18) that are consistent with Positron Emission Tomography (PET) results which showed an increase in Lingual Gyrus and Parahippocampal Gyrus involvement during visuomotor mapping (Toni & Passingham, 1999). Our findings of strong cingulate gyrus and posterior cingulate cortex sources (BA31 and BA23) overlap with the P3b sources found in the human iEEG literature, non-human vertebrate literature, and our previous attempts at source analysis discussed earlier (Halgren et al., 1994; Halgren et al., 1995a; Halgren et al., 1995b; Baudena et al., 1995; Smith et al., 1990; Brankack, et al., 1996; Shin, 2011; Kahana et al., 2001). An interesting result from this study is the lack of posterior cingulate cortex involvement before learning occurred, but the clear presence of all other sources for the P3b during

that time period. As learning developed, and with extended practice, the PCC appeared to increase in activation while the remaining sources were relatively stable (i.e. they did not follow the linear activation pattern shown in the PCC). We interpret this finding as a reflection of the representation and constant updating of action contexts carried out by the posterior dorsal corticolimbic system in the context-updating model of the P3b.

Limitations

One limitation of the current study is the presence of saccades and other ocular artifacts in the EEG. In order to closely mimic the field of view a quarterback would realistically have on the playing field, the stimuli used in our task were large and caused our participants to horizontally scan each picture for identifying features. The presence of artifacts prevented us from subjecting our data to whole-brain analyses such as Principle Components Analysis (PCA), which would have helped us identify other potential components that correlate with the learning seen in our study. Ocular artifacts also prevented us from performing source analysis on the MFN. Source analysis would have been useful for determining whether the sources of the MFN shifted as a result of training (e.g. from anterior to middle cingulate cortex), and in turn would help us further explore the reasoning behind the MFN's linear increase across days 2 and 3.

Another limitation of our study is that EEG only measures cortical activity, whereas there are numerous other studies which suggest several subcortical structures are essential to the learning process. This shortcoming forces us to rely on ERPs that may not directly reflect the processes we are interested in, for example we are interested in memory retrieval, but the P3b is an indirect way of measuring this component. There is a substantial amount of empirical evidence that suggest subcortical and cortical systems

work synchronously during the early and late learning stages (Toni & Passingham, 1999; Groll, de Lange, Verstraten, Passingham, & Toni, 2006; Brovelli, Laksiri, Nazarian, Meunier, & Bossaoud, 2008). Similarly, it is known that some memory systems rely more heavily on subcortical structures than others, and thus if our hypothesis that the inconsistent trend of the MFN seen in our study is the work of averaging activity from different memory systems together, we would not be able to explore all components of those systems with EEG alone.

Conclusions

The goal of the present study was to investigate the changes in the dorsal posterior corticolimbic system and medial frontal areas as participants achieve learning and skilled performance in a realistic categorization task. By using a task that is relevant to participants' background and by leveraging information of the brain responses with behavioral measures, we were able to support previous findings and extend them to more realistic learning situations, while at the same time raising more questions about the role of frontal control regions during categorization of realistic stimuli. Results from this study serve as motivation for the study in the next chapter which will utilize fMRI to further understand the cortical and subcortical sources involved in our category learning task.

CHAPTER III

DISSOCIATING MULTIPLE MEMORY SYSTEMS DURING FOOTBALL FORMATION CATEGORIZATION

The experiment and analyses described in this chapter were developed by myself and Dr. Zeithamova. Dr. Zeithamova contributed substantially to this work by developing a set of guidelines for univariate and multivariate fMRI analyses. I was the primary contributor to the experimental design, performed the analyses under Dr. Zeithamova's guidance, and did all the writing.

Introduction

Prior research has shown that people are able to employ the memory system that is better suited for a given category structure. For example: there is behavioral and neural evidence for engaging a rule-based system when categorizing stimuli separable by a unidimensional rule, while engaging a similarity-based system for categories better separable by overall perceptual similarity (Rips, 1989; Smith & Sloman, 1994). As discussed in Chapter I, common methods for dissociating between memory systems require multiple tasks where each system is engaged in isolation. But in everyday tasks, it is likely that perceptual and formal categorization rules may both be drawn upon, perhaps on a case-by-case basis (Ashby et al., 1998; Poldrack et al., 2001). To date, no experiment has used neuroimaging to definitively demonstrate switching between the rule-based and perceptual similarity system in a single task.

The successful switching between memory systems has only been inferred through behavioral studies. Rips (1989) pioneered the first attempt at understanding if people can engage different strategies (supported by different memory systems) to

perform categorization of a single stimulus set. In his task, subjects were presented with two object categories along with a description of a test object where the description only included a value on a single dimension, such as the object's diameter. Subjects were told ahead of time that the test objects were between the extreme values of the two categories on every trial. The categories always included one "variable" category where objects within this category can vary along the relevant dimension (such as *pizza*, as they come in different sizes) and one "fixed" category where objects within this category cannot vary along the dimension (such as a *quarter*). If subjects were asked to apply a rule to govern category membership, subjects overwhelmingly placed the test objects into the "fixed" category, whereas if they were asked to make a similarity judgement they were more likely to place the test object into the "variable" category. The results from this experiment are the only evidence that people can utilize rule-based and similarity-based categorization to classify items within a single task, but made no attempts at correlating these behavioral differences with the memory systems supporting each strategy.

Recall that the rule-based categorization system generates and applies explicit rules for stimulus classification through hypothesis testing. On each trial, the learner selects a strategy or rule, and if feedback to their response indicates this rule was correct then the rule is recycled on the next trial. If the feedback indicates an incorrect response, the learner must generate a new rule and continue the cycle until they discover a rule that maximizes performance. Hypothesis testing requires participants to consciously attend to the rule they are applying and interpret the feedback which will inform their rule selection on subsequent trials (Ashby et al., 1998; Maddox, Ashby et al., 2004; Zeithamova & Maddox, 2006; Waldron & Ashby, 2001). Neuroimaging studies of rule-

based categorization have shown that this system reliably engages dorsal lateral prefrontal cortex along with caudate nucleus – structures that are key to executive attention (Lombardi et al., 1999; Rao et al., 1997; Rogers et al., 2000; Ashby & Ell, 2001). In addition, patients with lesions to one or both of these regions perform poorly on rule-based categorization tasks while having little to no impairment on tasks that engage perceptual similarity (Brown & Mardsen, 1988; Robinson et al., 1980).

The perceptual similarity system is commonly engaged when subjects view a stimulus and rely on the retrieval of information from other stimuli that are similar to the probe to make a categorization decision (Medin & Shaffer, 1978). This system is also employed during A/Non-A prototype learning, where subjects are trained on exemplars from a single category and then asked whether a probe stimulus belongs to the trained category (Category A) or some other category (Category Non-A) (Zeithamova et al., 2008). The former strategy suggests subjects rely on relevant features abstracted from related stimuli to perform classification, whereas in prototype learning subjects compare the probe stimulus to a representative example (prototype) of the category. Regardless, the perceptual similarity system involves posterior visual areas, and does not rely as heavily on the working memory system used for rule-based categorization (Aizenstein et al., 2000; Reber, Stark, & Squire, 1998a; Reber, Stark, & Squire, 1998b; Reber & Squire, 1999).

Our results from Chapter II serve as motivation for the current experiment, where we believe that subjects flexibly switched between explicit rule application and perceptual similarity analysis to categorize the football formations. However, the spatial resolution of EEG and the arbitrary category structure used in the previous task made it

difficult to resolve this speculation. The goal of the current experiment is to examine the extent to which two different memory systems can be engaged during a single task. To evaluate this question, we will use the same categorization task from Chapter I, except we will structure the categories such that members of the same category share a common set of features instead of the arbitrary association used previously. Formations from two categories will be visually similar to each other and require subjects to discover an explicit counting rule to accurately categorize these formations. One category of formations will be visually distinct from the other two, where subjects should theoretically rely on perceptual similarity when classifying members of this category. We hypothesize that the category structure used in this task will force subjects into flexibly changing between the rule-based and perceptual similarity-based system on a trial-by-trial basis, and this alternation will be measurable with fMRI.

Method

Participants

Eleven right-handed subjects between the ages of 18 and 30 ($M = 24.2$) were recruited from the University of Oregon Human Subjects Pool to participate in this experiment (5 males, 6 females). Subjects had no self-reported neurological or psychiatric conditions, as well as no MRI contraindications. Subjects were compensated \$35 for their participation and the protocol was approved by the Electrical Geodesics, Inc. and University of Oregon IRBs.

Task

The task used was designed to interchangeably recruit a rule-based or similarity-based memory system to categorize 3 categories of football defensive formations. Two

formation categories were very similar to each other and one category was visually distinct from the other two. For the two similar categories, subjects needed to discover an explicit counting rule in order to categorize members of these two groups reliably: One category of formations displayed three people on the line of scrimmage, while the other had four. For the visually distinct category, subjects could rely on a simple visual similarity analysis to recognize members of this category as there were significantly more players on the line of scrimmage (six). This forced subjects to focus their attention to the line of scrimmage while ignoring irrelevant players positioned elsewhere on the field.

Every category had three formations, each sharing the defining number of players on the line of scrimmage for that category, for a total of nine formations used throughout the experiment. On each training trial, subjects were shown a random formation for 2.5 s and were tasked with pressing a button on a keypad to place the formation into one of the three categories during the 2 s window (Fig. 12). Corrective feedback was given to the subject immediately after making their response and was on the screen for 1.5 s. The inter-trial-interval was optimized for event-related-design using Optseq2 software and varied between 2 and 8 s (Dale, 1999). Each formation was shown six times during each training block and there were six total training blocks.

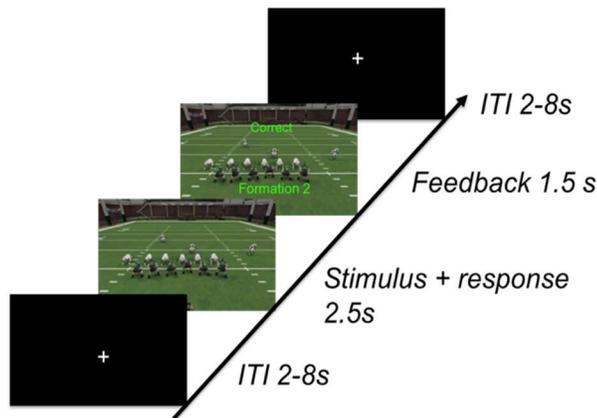


Fig. 12. Diagram of events in a training trial for experiment 2. Formations were shown for 2.5 s. Immediately following a response, contingent feedback was shown for 1.5 s. Upon feedback termination, a fixation mark was shown for the duration of the inter-trial interval of 2 – 8 s before the next formation was presented

A generalization block was implemented at the end of the experiment to test the subject's ability to generalize the strategies they acquired during training. During this block, the nine old formations were intermixed with nine new formations that belonged to the learned category structures. Each stimulus was randomly shown one at a time and was on the screen for 2 s while the subject used a response pad to categorize the stimulus. No feedback was given during this block, and instead a black fixation screen was shown for 10 s before the next stimulus was presented – resulting in a total fixed trial length of 12 s (Fig. 13). Each new and old stimulus was shown only once during the generalization block.

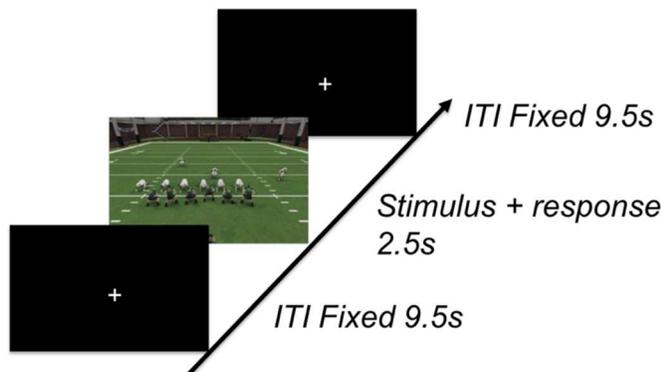


Fig. 13. Diagram of events in a generalization trial for experiment 2.

Formations were shown for 2.5 s regardless of when a subject made a response. No feedback was given. Instead, a fixation cross appeared for a fixed 9.5 s until the next formation was shown.

fMRI

MRI data was acquired with a 3T Siemens Skyra. A high-resolution T1-weighted MPRAGE was acquired for co-registration and normalization before the task was administered (TR = 2.5 s, TE = 3.41 ms, flip angle = 7°, matrix size = 256x256, FOV = 256 mm, 1 mm isotropic). Whole-brain fMRI was acquired using a gradient-echo EPI pulse sequence: TR = 2 s, TE = 26 ms, 100x100 matrix, FOV = 200 mm, 72 oblique axial slices, no skip, 2 mm isotropic voxels, GRAPA factor 2, multiband factor 3.

Procedure

Before coming to the scanning center, subjects were pre-screened over the telephone to ensure eligibility. Upon arrival at the center, a structural T1 scan was acquired followed by an exposure block with simultaneous scanning. During this block, subjects were asked to passively look at the screen and refrain from pressing any buttons. No other context or instructions were given. Each of the nine training formations were shown one at a time for 2 s each before a fixed 10 s ITI. Each formation was shown 4 times for 36 total trials. Following the exposure block, subjects were read instructions for the experiment. They were told how many formations there would be in the experiment, along with the set number of categories the formations belonged to. Their job was to figure out which formations belong to each category by pressing the buttons on their response pad and utilizing the corrective feedback. A brief practice test (un-scanned) was given where they learned to categorize unrelated formations. After practice, six training blocks were run with brief breaks in-between, and after training the subjects sat through another exposure block where they passively viewed each stimulus. To end the experiment, subjects went through the generalization block, given only the instructions that they were going to go through a final block with no feedback. They were not told whether there would be novel formations in this block. Subjects were asked to write-down their strategies in a debrief questionnaire for categorizing the formations before receiving compensation and leaving the center.

Results

Behavioral

Due to noise caused by motion during scanning, only 10 out of the 11 subjects were retained for analysis. When evaluating performance for each stimulus category across training runs, we can clearly see that subjects performed better at categorizing the visually distinct category compared to the two visually similar categories (Fig. 14). However, this performance difference was only present for the first 3 training runs, and by the end of run 4 subjects were able to accurately identify members of all categories equally.

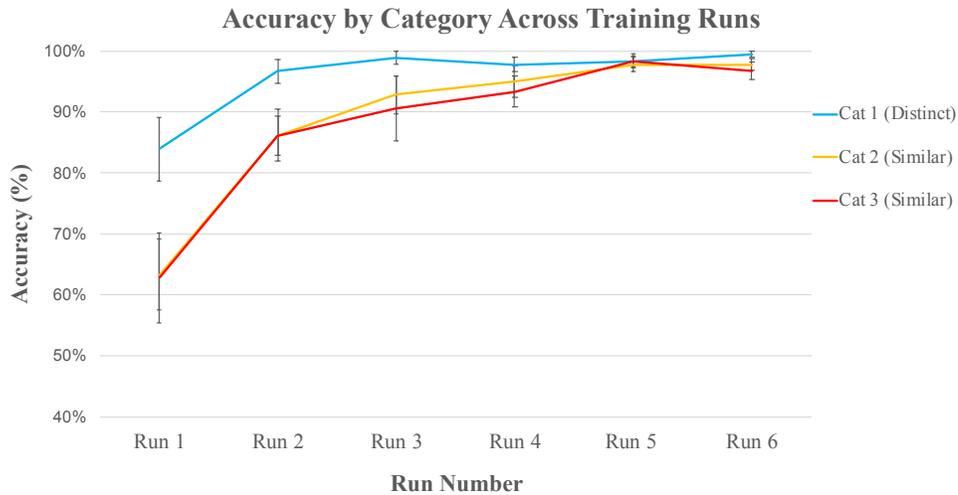


Fig. 14. Categorization accuracy by run. Subjects accurately categorized the visually distinct category much quicker than the two visually similar categories. Accuracy for the visually similar categories peaked between runs 4 and 5, which we infer is the time at which subjects discovered the counting rule.

A confusion matrix shows that subjects commonly mixed up the two visually similar categories when making errors. By block 4, subjects limited their confusion, and we can infer that this was the point at which most subjects discovered the explicit counting rule which allowed them to differentiate between members of the two categories (Fig. 15).

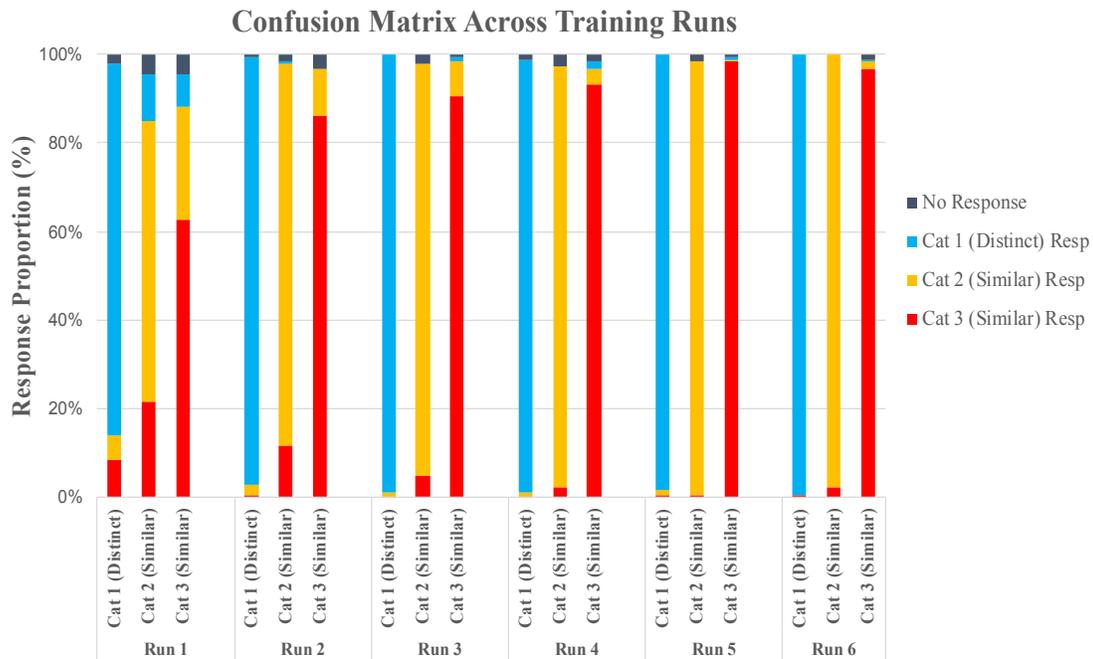


Fig. 15. Confusability matrix. During the first 3 training blocks, subjects commonly confused the two visually similar categories for one another. By run 4, subjects were able to accurately dissociate between these two categories. Subjects rarely confused any other category when classifying formations in the visually distinct category.

The generalization run was used to ensure subjects acquired the correct categorization strategies by testing their ability to generalize their strategies to novel stimuli belonging to the same categories used in training. On average, subjects completed the generalization run with 92% accuracy for the visually distinct category and 88% accuracy for the visually similar categories (Fig. 16). Had subjects been relying on the

declarative recall of individual stimuli throughout training, their performance in the generalization run would have been closer to 50%.

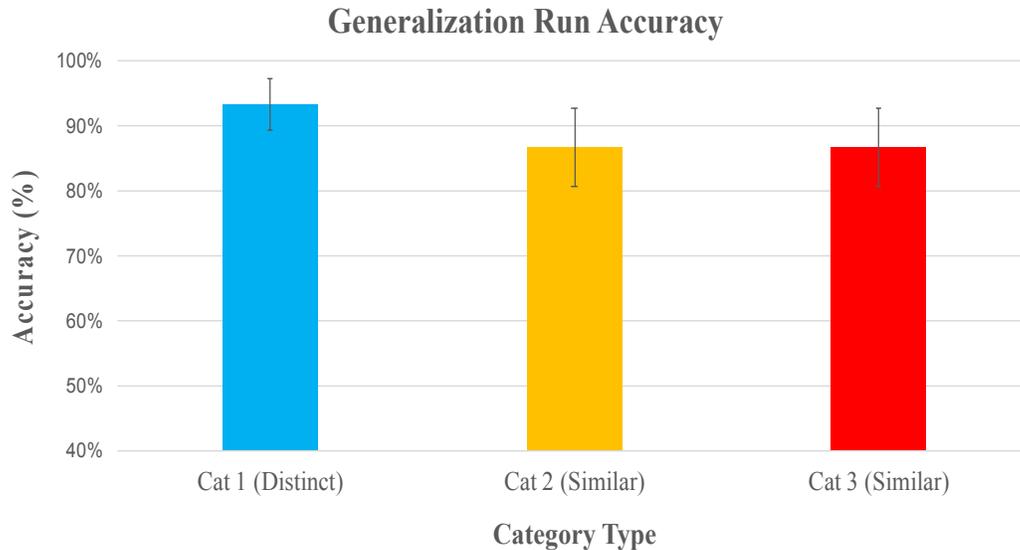


Fig. 16. Generalization accuracy graph. During the generalization block, performance was highest for the visually distinct category – which included novel and old stimuli. Accuracy was slightly lower for the two visually similar categories, although not statistically significant.

Univariate Analysis

Training Runs

Data from each training run and each participant was analyzed separately at a first level analysis. Visually distinct and rule-separable correct trials were modeled separately as two predictors. Each category stimulus onset time was convolved with a hemodynamic response function and entered into a general linear model with their temporal derivatives to estimate beta weights. Data from run 4, 5, and 6 (the runs after subjects could perform the task with proficiency) were combined at a second-level using fixed-effects analysis. A group analysis was then run, with contrasts showing areas that were more engaged during visually distinct trials over rule-separable trials, and vice-versa (Fig. 17).

Individual voxels were considered active when reaching $Z > 1.9$ and surviving a whole-brain cluster size threshold set at $p < .05$ (Worsley, 2001).

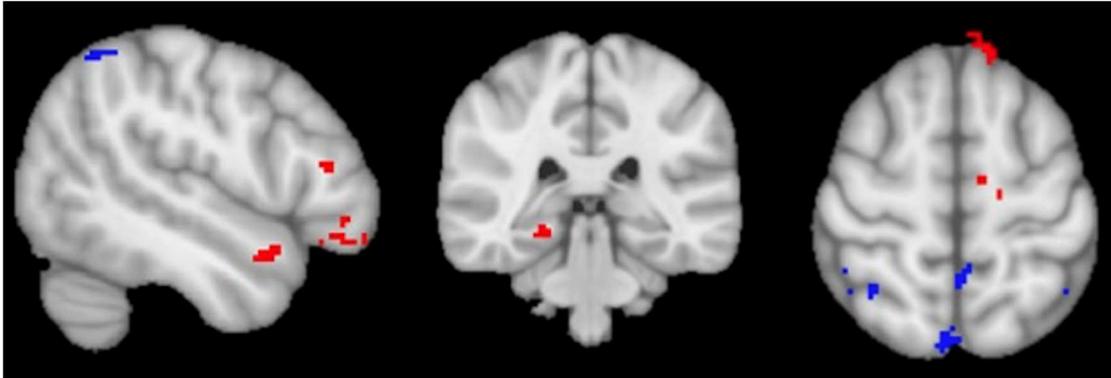


Fig. 17. Univariate contrasts of Rule > Similarity (Red) and Similarity > Rule (Blue) during training. Red: dorsal lateral and inferior frontal areas along with hippocampus were engaged significantly more during rule application compared to perceptual similarity analysis. Blue: Fusiform gyrus and lateral occipital cortex were engaged significantly more during perceptual similarity analysis compared to rule application.

Consistent with our hypotheses, the superior and inferior frontal gyri were engaged significantly more on rule-separable trials compared to trials categorizable using perceptual similarity (red clusters, Fig. 17). The right hippocampus, a region associated with declarative recall, was also engaged during rule-separable trials. For perceptual similarity trials, the lateral occipital cortex and fusiform gyrus were engaged significantly more when compared to rule-separable trials (blue clusters, Fig. 17). A summary of the top 11 regions associated with each condition is displayed in Table 1 and Table 2.

Table 1

Cluster location and size for Rule >Similarity contrast in blocks 4, 5, and 6

Location	Cluster Size (Voxels)	Z-Value	X	Y	Z
L. Sup. Fr. Gyrus	58	2.79	-54	44	-10
L. IFG	50	2.95	-50	30	14
L. Sup. Fr. Gyrus	38	2.72	-12	40	56
L. Sup. Fr. Gyrus	34	2.47	-16	56	38
R. Hippocampus	26	2.88	22	-34	-10
L. Sup. Temp. Gyrus	25	2.67	-50	10	-16
R. Fusiform Gyrus	25	3.04	40	-44	-20
L. Suppl. Motor Cortex	24	2.72	-10	-12	56
R. Lateral Occipital Cortex	22	2.42	58	-64	24
Brain Stem	22	2.63	6	-22	-28
R. Mid. Temp. Gyrus	20	2.56	40	-58	2

Table 2

Cluster location and size for Similarity >Rule contrast in blocks 4, 5, and 6

Location	Cluster Size (Voxels)	Z-Value	X	Y	Z
R. Lateral Occipital Cortex	519	3.16	6	-74	36
R. Lateral Occipital Cortex	154	2.87	34	-62	62
L. Fusiform Gyrus	106	3.17	-20	-66	-18
L. Lateral Occipital Cortex	98	2.83	-36	-56	38
R. IFG	89	3.25	20	56	-6
L. Post. Cingulate Gyrus	70	2.88	-8	-40	48
R. Lateral Occipital Cortex	55	2.47	20	-88	38
R. Fusiform Gyrus	52	2.62	20	-54	-16
L. Middle Frontal Gyrus	49	2.77	-38	34	18
R. Occipital Pole	41	2.34	20	-104	-10
Brain Stem	40	3.05	22	-32	-42

Generalization Run

Similar to the preprocessing workflow used for the training runs, data from each subject was analyzed separately at a first-level analysis. Visually distinct and rule-separable trials that involving new stimuli used only in the generalization block and old stimuli used during training were modeled as four separate predictors (i.e. Novel-Rule, Novel-Similarity, Old-Rule, Old-Similarity). Each category stimulus onset time was convolved with a hemodynamic response function and entered into a general linear model with their temporal derivatives to estimate beta weights. A group analysis was then run, with contrasts showing areas that were more engaged during visually distinct trials over rule-separable trials, and vice-versa (Fig. 18). Individual voxels were considered active when reaching $Z > 1.9$ and surviving a whole-brain cluster size threshold set at $p < .05$ (Worsley, 2001).

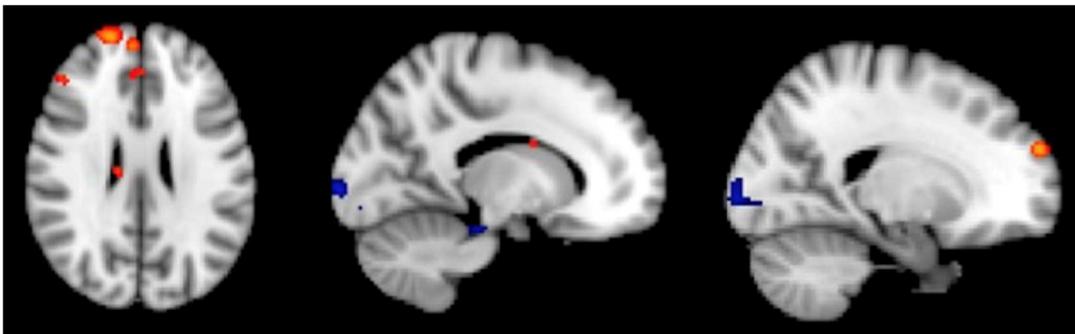


Fig. 18. Univariate contrasts of Rule > Similarity (Red) and Similarity > Rule (Blue) during generalization. Red: Frontal control regions were engaged significantly more during the rule-based trials compared to similarity-based trials during generalization. A cluster over caudate nucleus was also found. Blue: Similarity-based trials relied more heavily on lateral occipital cortex compared to trials separable by a counting rule.

Results from our univariate analysis show that the left caudate nucleus, left superior frontal gyrus, and left inferior frontal gyrus were engaged significantly more on rule-based trials compared to perceptual similarity trials (Fig. 18). Caudate nucleus, instead of hippocampus, is one of the only observable differences between the training and generalization contrasts for this condition. A list of the top 11 clusters from this

contrast are listed in Table 3. In addition, the lateral occipital cortex and right fusiform gyrus were engaged significantly more for perceptual similarity trials over rule-separate trials during generalization (Fig. 18). A summary of the top 11 clusters is shown in Table 4.

Table 3

Cluster location and size for Rule >Similarity contrast in generalization block

Location	Cluster Size (Voxels)	Z-Value	X	Y	Z
L. Caudate Nucleus	290	3.5	-8	-10	24
Cerebellum	129	3.22	16	-72	-28
Cerebellum	125	3.43	32	-80	-22
Cerebellum	90	3.18	4	-50	-10
L. Sup. Frontal Gyrus	88	3.21	-28	6	64
L. Lateral Occipital Cortex	73	3.17	-26	-78	50
R. Lateral Occipital Cortex	71	3.08	40	-74	42
L. Inf. Frontal Gyrus	67	3.22	-42	22	4
Cerebellum	63	3.3	-26	-90	-26
L. Sup. Frontal Gyrus	59	3.17	-42	46	20
Brain Stem	58	2.92	14	-16	-38

Table 4

Cluster location and size for Similarity >Rule contrast in generalization block

Location	Cluster Size (Voxels)	Z-Value	X	Y	Z
R. Lateral Occipital Cortex	1922	4	18	-100	6
R. Fusiform Gyrus	335	3.41	12	-72	-2
R. Inf. Frontal Gyrus	213	3.11	62	6	12
L. Postcentral Gyrus	144	3.35	-40	-26	54
L. Sup. Temporal Gyrus	143	3.49	68	-24	28
Cerebellum	113	3.09	-20	-72	-52
L. Fusiform Gyrus	100	2.88	38	-54	-24
R. Mid. Temporal Gyrus	99	3.13	66	-40	2
R. Mid. Frontal Gyrus	82	3.25	32	18	30
R. Mid. Temporal Gyrus	79	3.37	54	-6	-28
R. Angular Gyrus	70	3.7	56	-46	30

Multi-Voxel Pattern Analysis

In order to perform region-based Multi-Voxel Pattern Analysis (MVPA), cortical and subcortical regions of interest (ROIs) were automatically defined using Freesurfer software for each participant (Dale, Fischl, & Sereno, 1999; Fischl et al., 2002). ROIs included superior parietal lobe, anterior cingulate cortex (ACC), medial orbitofrontal cortex (MOFC), inferior parietal lobe, inferior frontal gyrus (IFG), and fusiform gyrus. Data from each participant was then modeled using a separate regressor for each trial to construct a betaseries representing each trial (Rissman, Gazzaley, & D'Esposito, 2004). Each betaseries was smoothed ($\sigma = 3$) before being registered to a normalized space using Advanced Neuroimaging Tools (ANTs) toolbox (Avants et al., 2011). A Support Vector Machine (SVM) classifier was trained on data from each individual ROI for 5 out of 6 training runs, and performed a binary classification of category membership (rule-based vs similarity-based categories) on the 6th run (Mumford, Turner, Ashby, & Poldrack, 2012). Leave-one-run-out cross validation was performed and an average classification accuracy was obtained for every subject.

Due to a site-wide data loss, 9 out of 11 subjects were used for MVPA. A one-sample t-test was used against a baseline value of .5 (50% chance) to determine each ROIs ability to classify between rule-based and similarity-based conditions. The IFG ($M = .66$; $t(8) = 4.23$, $p = .003$), inferior parietal lobe ($M = .70$; $t(8) = 3.65$, $p = .007$), superior parietal lobe ($M = .76$; $t(8) = 5.8$, $p < .001$), MOFC ($M = .58$; $t(8) = 3.3$, $p = .011$), and fusiform gyrus ($M = .62$; $t(8) = 3.75$, $p = .006$) all predicted category membership with statistically significant accuracy (Fig. 19). The ACC ($M = .58$) failed to accurately differentiate between categories, $t(8) = 2.02$, $p = .078$.

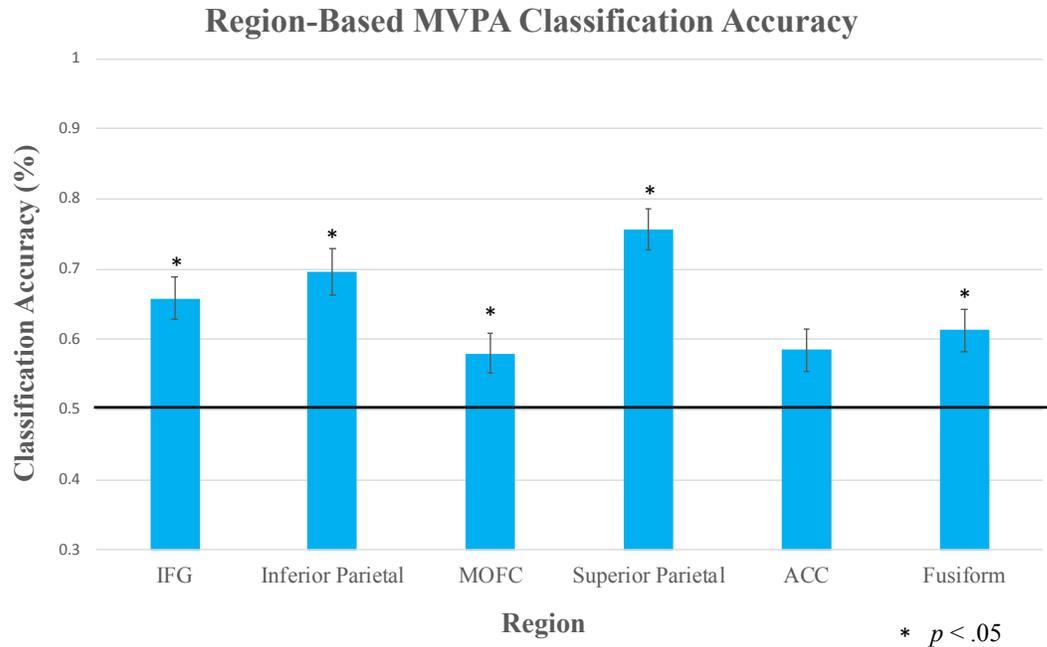


Fig. 19. ROI-based MVPA classification accuracy. The IFG, inferior parietal cortex, MOFC, superior parietal cortex, and fusiform gyrus were able to classify between our two conditions with significantly above-chance accuracy. Amongst these regions, the superior and inferior parietal cortices provided the most reliable classification. The ACC did not reach statistical significance.

Discussion

The main goal of this experiment was determine the extent to which people engage multiple memory systems during a single categorization task. In-line with our hypotheses, the results showed that once subjects acquired the formations in the task, the rule-based and perceptual similarity system were both engaged but it depends on the formation presented. These two systems were also engaged during a test block where subjects were forced to generalize the categorizations strategies they developed during training. For the machine-learning analysis, regions in the superior and inferior parietal lobes, along with MOFC, fusiform, and IFG successfully dissociated between conditions in the task.

Univariate fMRI

The categories in this experiment were designed such that they require subjects to discover a counting rule to differentiate between two visually similar categories and utilize a perceptual similarity strategy to identify members of a visually distinct category. Our subjects' performance on the generalization block support the assumption that they would recruit the proper strategies. Specifically, they would not have been able to accurately categorize novel formations into the trained categories had they relied exclusively on declarative recall of individual formations.

The superior and inferior frontal gyri were more active during the categorization of visually-similar trials compared to visually distinct trials. These regions are a part of the working memory system, where it is inferred that they are responsible for orienting attention and establishing executive control (Lombardi et al., 1999; Rao et al., 1997; Rogers et al., 2000; Ashby & Ell, 2001). In our experiment, subjects focused their attention toward the players on the line of scrimmage, where they were required to count each player if the formation belonged to one of the two visually similar categories. These results support the findings of past research that have established a greater working memory requirement for the rule-based system compared to the perceptual similarity system.

Interestingly, caudate nucleus, a region that is integral to rule application, did not reach a level of significance for the rule-based condition during training. Instead, a cluster over the hippocampus had the highest level of activation during training – a region that is well-known for its role in declarative recall (Eichenbaum, 2000). Given that subjects only needed to attend to a single feature within each stimulus to perform categorization (the number of players on the line of scrimmage), it is possible that

subjects utilized the rule for a short period of time during training but relied more on declarative recall of the few relevant players. However, when encountering novel formations in the generalization block that belong to the categories acquired during training, subjects were forced into applying the counting rule, and thus the strong presence of caudate nucleus during generalization could reflect a more consistent reliance on rule application.

In support of our hypothesis, robust activation of the lateral occipital cortex was present for the visually distinct category when compared to the visually similar categories. This held true throughout training and extended into the generalization block. The lateral occipital cortex has been well-established as the main region governing perceptual similarity categorization (Aizenstein et al., 2000; Reber et al., 1998a; Reber et al., 1998b; Reber & Squire, 1999). Perceptual similarity categorization can be carried-out with minimal working memory resources, and is optimized for instances with low within-category similarity (Nosofsky, 1986). The absence of the working memory system when subjects viewed members of the visually distinct category further support our conclusion that this category engages the perceptual similarity system. The novelty of our findings is rooted in our demonstration that subjects switch between the perceptual similarity-based or rule-based memory systems based on the structure of a given stimulus.

MVPA

Our region-based MVPA showed that frontal and parietal regions provided the most reliable classification between the rule-based and perceptual similarity-based categories. From our univariate analyses, we saw that the rule-based condition relies heavily on these regions, and supports previous findings that this form of categorization

requires a higher degree of attentional resources. On the other hand, MVPA provides a more sensitive measure of these conditional effects. More specifically, MVPA provides an avenue to detect more subtle differences between our conditions that lie within the activity patterns of single regions – information that is sometimes subtracted-out by traditional analyses (Haynes & Rees, 2006). These small activation patterns can potentially code for task-relevant information that is important to *both* memory systems in our experiment.

Frontoparietal regions are well-known for their importance to cognitive control, with an essential component being selective-attention to information that is relevant to the task (Desimone & Duncan, 1995; Miller & Cohen, 2001). However, non-human primate experiments have demonstrated that activity in frontal and parietal regions can code for an array of different task-relevant features, such as representations of individual stimuli, rule selection, and response selection (Asaad, Rainer, & Miller, 1998; Freedman & Assad, 2006; White & Wise, 1999). Follow-up studies in humans have shown similar dissociations between stimulus sets and rules using MVPA (Bode & Hanes, 2009; Haynes et al., 2007). These components are essential to the rule- *and* perceptual similarity-based systems, and we can interpret the reliable dissociation between our two conditions within these regions as reflecting these concepts.

Limitations

A major limitation in this study was the small sample size of 11 subjects. This experiment was only made possible with the generous funding of a pilot grant from our scanning center. Unfortunately, attempts at securing additional funding to acquire a more interpretable sample size were unsuccessful. A larger sample size would have allowed us

to split subjects based off the strategies they used during the experiment. It would have been ideal to have a large enough sample to understand how each memory system relates to small variations in strategy. However, the reliability of the effects seen in our current sample (i.e. the small variance) leads us to believe that increasing the sample size would only make the presented results more robust.

Another limitation in our study is the inability to track strategies on a case-by-case basis. The differences between the training and generalization contrasts make us wonder if subjects used rule-based categorization for only a short amount of time during training before switching to declarative recall. Future studies should prompt subjects to provide their categorization strategies at the end of each block, as opposed to at the end of the experiment. This would allow us to perform a better segmentation of our data and separate the different regions associated with strategy changes within a single subject.

Relatedly, the specific rules utilized in this experiment were counting rules that are associated with numerical cognition. We did not empirically test the generalizability of the results for our rule-based condition to other rules such as those used to differentiate between shapes or colors. Yet, the contrasts conducted in our fMRI analyses did not reveal posterior parietal cortex, the canonical area associated with numerical cognition, as differentiating between the rule-based and perceptual similarity-based conditions (Hubbard, Piazza, Pinel, & Dehaene, 2005; Arsalidou & Taylor, 2011; Mandelbaum, 2013). This may suggest that our results are not specific to numerical cognition.

Conclusions

The goal of the present study was to examine if people can engage multiple memory systems to categorize stimuli in a single task. Overall, the results from this

experiment support our hypotheses, and we conclude that people can switch between a rule-based and similarity-based memory system to optimize the categorization of a stimulus. Our results support the findings of decades worth of neuroimaging research that have established the differences between memory system, but expands on them by looking at how these systems are employed under more realistic learning conditions. This is an important precedent, as it establishes the need to take a more dynamic approach toward creating interventions to enhance learning in the real-world. Unfortunately, the spatial shortcomings of fMRI prohibit us from further investigating how these are selected and employed on timescales well below that of the hemodynamic response – information that would be useful for optimizing the timing of an intervention. The proceeding experiment in Chapter IV aims to augment the findings of the current experiment by attempting to describe the temporal dynamics of successful memory system engagement.

CHAPTER IV

EVALUATING TEMPORAL DYNAMICS OF CATEGORIZATION

Introduction

In Chapter III we demonstrated that the rule-based and similarity-based categorization systems are differentially engaged during our categorization task, but the results did not provide any details about the timing under which these systems are employed. As discussed in Chapter I, fMRI operates on a time-scale that makes it difficult to separate temporal information between multiple processes that happen within the hemodynamic response window, such as orienting attention or selecting a rule to categorize a stimulus in our football task. In EEG, a recording method with exceptional temporal resolution, several ERPs exist which index some of the defining cognitive processes of the rule-based and similarity-based memory systems; the latencies of which could be used to extract information about the temporal dynamics of these two systems. However, at the time of this dissertation, no work has been done to describe the differences between these two systems using EEG.

Recall that the Medial Frontal Negativity (MFN) is an ERP component with a latency window of 180 – 300 ms after stimulus onset. From what we saw in Chapter II, the amplitude of the MFN is largest in the early stages of learning during visuomotor association tasks, and theories from similar experiments suggest the amplitude of this component reflects the amount of controlled attention required to complete the task (Bush et al., 2002). The rule-based memory system relies more heavily on working memory compared to the perceptual similarity system (Zeithamova et al., 2008). The overlap in function and general location of the MFN with the known frontal control regions

associated with the rule-based memory system makes it an ideal marker for indexing the attentional aspect of this system.

In addition to frontal cortical areas, the engagement of the dorsal striatum is reliably seen in fMRI during tasks that require participants to recall declarative information, such as an explicit rule seen in Chapter III (Lombardi et al., 1999; Rao et al., 1997; Rogers et al., 2000; Ashby & Ell, 2001). It is becoming increasingly clear that the dorsal striatum is an essential component of the working memory system (Lewis et al., 2004), where it is hypothesized to regulate the information contained in working memory (Frank, Loughry, & O'Reilly, 2001; Gruber, Dayan, Gutkin, & Solla, 2006). Although EEG cannot measure subcortical activity in the striatum, intracranial recordings from rats suggests that hippocampal theta rhythms show reliable coherence with striatal theta during declarative recall (DeCoteau et al., 2007). Due to the laminar cell structure of the hippocampus, hippocampal theta is measurable in humans using non-invasive EEG, and can be seen during working memory tasks (Tesche & Karhu, 2000). Theories for the hippocampus's role in WM are similar to the ones proposed for the dorsal striatum, but are more commonly framed with respect to the context under which information was acquired. More specifically, hippocampal theta (as it relates to WM) has been hypothesized to index the monitoring of contextual information based off the items present in WM (Tesche & Karhu, 2000). This theory for hippocampal theta is almost identical to the context-updating theory used to describe the significance of the P3b ERP component discussed in Chapter II. It is no coincidence that a strong source of the P3b has been reliably traced to the medial temporal lobe, including the hippocampus, and has been historically used as an additional measure for hippocampal theta rhythms

(Brankack, et al., 1996; Shin, 2011; Kahana et al., 2001). We would expect the amplitude of the P3b to correlate with the retrieval of declarative information, which is an essential function for rule-based *and* similarity-based categorization – particularly if participants rely on declarative recall as opposed to a perceptual similarity strategy discussed in Chapter III.

A third component that could potentially dissociate between the two memory systems is the Lateral Inferior Anterior Negativity (LIAN). The LIAN is a lesser-known bilateral component that has shown clear dissociations between the recognition of spatial targets and digit targets in a visuomotor association task (Luu et al., 2007). Specifically, the amplitude of the right LIAN is anticorrelated with acquiring the ability to recognize spatial configurations and shows no changes when targets invoke the phonological loop. However, the amplitude of the left LIAN is positively correlated with learning to recognize phonological targets, and is insensitive to acquiring an ability to perform spatial analyses. The primary source of these components is inferred to be the Inferior Frontal Gyrus (IFG), but it is worth mentioning again that the LIAN is rarely discussed in the literature, where it doesn't receive any mention outside of its role in visuomotor learning. Due to the lack of work being conducted on how the MFN, P3b, and LIAN relate to different memory systems, it would be worthwhile to employ methods that have a higher sensitivity for detecting a dissociation between these systems when compared to standard ERP/univariate analyses.

The use of machine learning in fMRI to evaluate the spatial representations of memories has seen a surge in popularity over the past two decades (Haxby et al., 2001; Haynes, 2015; Tong & Pratte, 2012; Norman, Plyn, Detre, & Haxby, 2006). Yet, the field

of EEG has only recently entered into the machine learning domain, and to date no work has utilized machine learning to evaluate when, on a given trial, the pattern representation of specific memories becomes reliable. Common methods for defining the spatial representation of memories in fMRI requires that activity in each voxel is averaged across a full trial which extends several seconds. However, as you will see later in this chapter, this technique can be applied to EEG where a spatial pattern is derived from the raw voltages of several groups of electrodes for individual samples (1-4 ms) within a trial. This results in a series of pattern representations, and by analyzing the classification accuracy of these patterns across time we can derive the earliest point at which a participant selects a memory system that is optimized for categorizing specific types of stimuli.

The goal of this chapter is to examine the earliest point we can accurately differentiate between a participants' selection of two different memory systems when viewing a stimulus. In addition, we are interested in evaluating overlap between the spatial distribution of any dissociating ERP components and the location of the memory systems described with fMRI in Chapter III. To accomplish this, we will have participants perform the same task used in Chapter III but this time they will perform the task during an EEG recording session. Having already established that this task reliably recruits the rule-based and similarity-based memory systems in the previous chapter, we will interpret standard ERP amplitude differences between categories as a reflection of the memory system chosen for each category. The onset of these ERP differences coupled with results from a machine learning analysis will lead us into a theory that describes the temporal dynamics of selecting and employing these two systems.

Method

Participants

Forty-four right-handed participants were recruited from the University of Oregon Human Subjects Pool (22 males, 22 females), with ages ranging between 18 and 39 years old ($M = 19.5$, $SD = 3.2$). All participants had normal or corrected-to-normal vision, had no history of head trauma or seizures, and were not consuming medication that could affect their EEG. Participants were pre-screened online for their experience with football in order to reduce the chance of contextual familiarity confounding differences in skill acquisition rate. Only those subjects that were comfortable recognizing football defensive formations were allowed to participate. The research protocol was approved by the University of Oregon and Brain Electrophysiology Laboratory Company (BELCO) institutional review boards, and the study took place in the laboratory of BELCO.

Task

The task used in this study was an EEG analogue of the fMRI task used in the previous chapter. Stimuli in this task consisted of three categories of football defensive formations with two categories being very visually similar to each other and one category being visually distinct from the other two. For the two similar categories, subjects needed to discover an explicit counting rule in order to categorize members of these two groups reliably: One category of formations displayed three people on the line of scrimmage, while the other had four. For the visually distinct category, subjects could rely on a simple visual similarity analysis to recognize members of this category. For the visually distinct category, six players were placed on the line of scrimmage. Within each category, all players were shuffled around the field of view with the exception of the players on the

line of scrimmage, as the number of players on the line dictated category membership (Fig. task breakdown). This forced subjects to focus their attention to the line of scrimmage over time while ignoring irrelevant players positioned elsewhere on the field.

Every category had three formations, each sharing the defining number of players on the line of scrimmage for that category, for a total of nine formations used throughout the experiment. On a given trial, participants were randomly shown one of the nine formations for 2,000 ms and were instructed to place the stimulus into one of the three categories by pressing a button on a response box within the stimulus exposure window. Once they made a response, the stimulus disappeared and the subject was presented with a corrective feedback screen which indicated whether they were correct along with text describing the correct category for the stimulus (Fig. 20). The feedback was on the screen for 1,500 ms, after which a fixation cross with a variable inter-stimulus-interval was shown for 2,000 – 3,000 ms (Fig. task breakdown). The task was divided into eight training blocks consisting of 90 trials (or 10 exposures per stimulus) per block, which totaled 80 exposures of every stimulus throughout training.

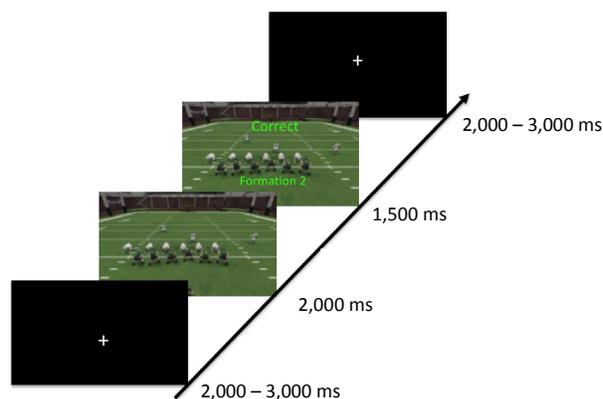


Fig. 20. Diagram of events in a trial during training for experiment 3. Formations were shown for 2 s. Immediately following a response, contingent feedback was shown for 1.5 s. Upon feedback termination, a fixation mark was shown for the duration of the inter-trial interval of 2 – 3 s before the next formation was presented

After the final training block, a generalization block was used which tested each subject's ability to apply any rules they developed during training to novel stimuli.

During the training block, a mixture of the nine training stimuli and nine novel stimuli belonging to the same categories were used. Subjects were not told that the generalization block would include novel stimuli. No feedback was given to the participants after pressing a button to categorize each formation. Instead, a black screen was shown for 1,500 ms after a response was made before the fixation cross appeared to begin the next trial. Each old and novel stimulus (18 total stimuli) was shown 5 times for a total of 90 trials in the generalization block.

Procedure

Following the informed consent process, participants were fitted with a 256-channel HCGSN EEG net and placed 55 cm in front of the computer monitor. A chin-rest was used to minimize head movements and keep the distance to the monitor fixed for every participant. Participants were explicitly told that there were nine defensive formations in this study belonging to three categories, and that they must learn which formations go into each category. The response feedback that would help teach the participant to make the correct decision was explained clearly, and participants were allowed to ask questions before the experiment began.

Once the participant could demonstrate an understanding of the study to the research assistant, a short practice block consisting of 12 trials followed. Formations used in the practice block resembled different basketball formations to avoid familiarity effects once the real training began. After the practice block, 8 training blocks occurred followed by a final generalization block to test a subject's strategies to novel members of the acquired categories. At the end of the experiment, participants filled out a debriefing questionnaire which asked them to describe the strategies they used to categorize each

group of formations. Each session lasted around 2.5 H, and participants were compensated course credit for their participation.

Learning Criterion

To remain consistent with Chapter I, we used the fixed-number of consecutive responses method (FCCR) in order to determine when a participant had sufficiently acquired the formations in each category (Luu et al., 2007). To re-iterate this method: a subject fulfilled the learning criterion when they could make four correct responses in a row for each stimulus.

EEG Recording and Post-Processing

EEG was recorded and processed for ERP analysis using the method described in Chapter I. Channel montages describing the location of each ERP are shown in Fig. 21.

Results

Behavioral

Behavioral measures for the two visually distinct categories were averaged together to represent the rule-based condition in our experiment. On average, subjects were significantly better at categorizing the similarity-based category (95%) than the rule-based categories (90%), $t(43) = 5.45, p < .001$. In addition, subjects had significantly faster reaction times when responding to the similarity-based category (708 ms) when compared to the rule-based categories (823 ms), $t(43) = -8.97, p < .001$.

Event-Related Potentials

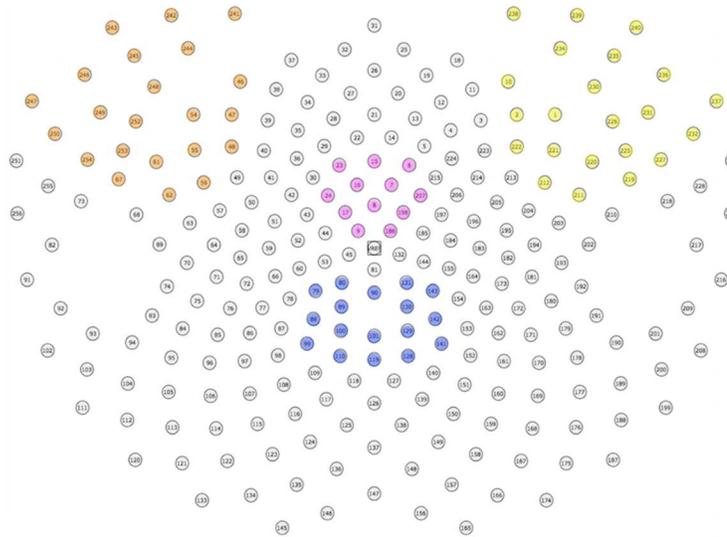


Fig. 21. Electrode montages used for the LIAN, MFN, and P3b ERP components. Orange and Yellow: Electrodes used for the LIAN analysis. Pink: Electrode cluster used to categorize the MFN. Blue: Electrodes used to analyze the P3b.

MFN

For the MFN analysis, a cluster of electrodes that best represent the medial frontal distribution of the component were chosen (pink electrodes, Fig. 21). To quantify the MFN, an adaptive mean amplitude corresponding to 20 ms before and 20 ms after the maximum negative peak amplitude in a window extending from approximately 180 – 300 ms after stimulus onset was computed for the MFN electrode cluster. The MFN was referenced to the preceding positive peak (P200) around 150 – 200 ms after stimulus onset. This method was applied for the post-learning trials for all three formation categories. Trials in the visually distinct category (6-1) were averaged together to form a single ERP for the similarity-based condition. Trials in the visually similar categories (3-4 and 4-3) were averaged together to form a single ERP for the rule-based condition. A paired-samples t-test was run to evaluate differences in MFN amplitude for the similarity-based and rule-based categories. The test revealed a marginally significant effect, such

that the MFN was largest for the similarity-based category ($M = -2.31 \mu\text{V}$) than the rule-based category ($M = -2.07 \mu\text{V}$), $t(43) = -1.98$, $p = .054$ (Fig. 22).

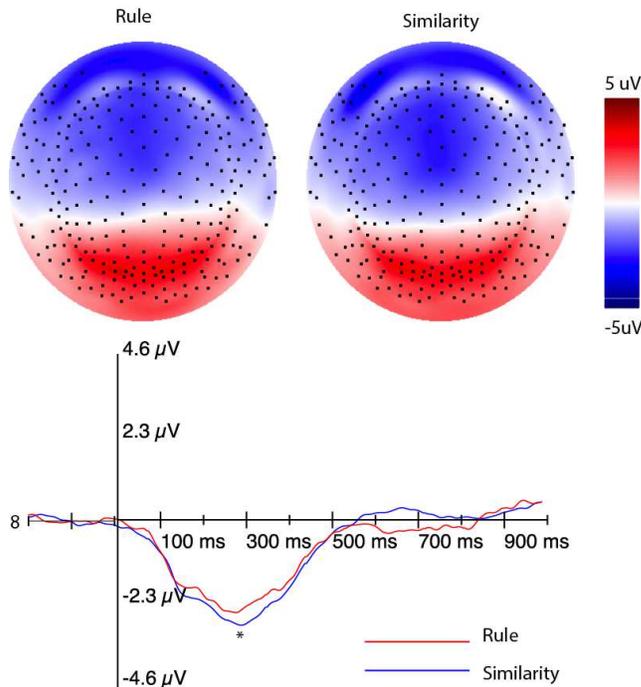


Fig. 22. Voltage map and representative waveform of the MFN. Top: A voltage map displays the voltage across the scalp for the rule and similarity-based conditions at the peak of the MFN (asterisk in bottom waveform image). A stronger negative voltage is seen over the medial frontal areas for the similarity-based condition. Bottom: Representative waveform showing the shape of the MFN for both conditions. The amplitude of the MFN is higher (more negative) for the similarity-based condition.

P3b

For the P3b analysis, a set of channels corresponding to the posterior-parietal distribution of the component were used (blue electrodes, Fig. 21). To quantify the component, an adaptive mean amplitude corresponding to 22 ms before and after the peak amplitude window extending from approximately 450 – 950 ms after stimulus onset was computed for the group of electrodes. This method was applied for the post-learning trials for all three formation categories. A single ERP for the rule-based and similarity-based categories was computed similar to the method described for the MFN. A paired samples t-test revealed that the amplitude of the P3b for the similarity-based category ($6.02 \mu\text{V}$) was significantly larger than the rule-based category ($5.34 \mu\text{V}$), $t(43) = 4.17$, $p < .001$. This effect is displayed in Fig. 23.

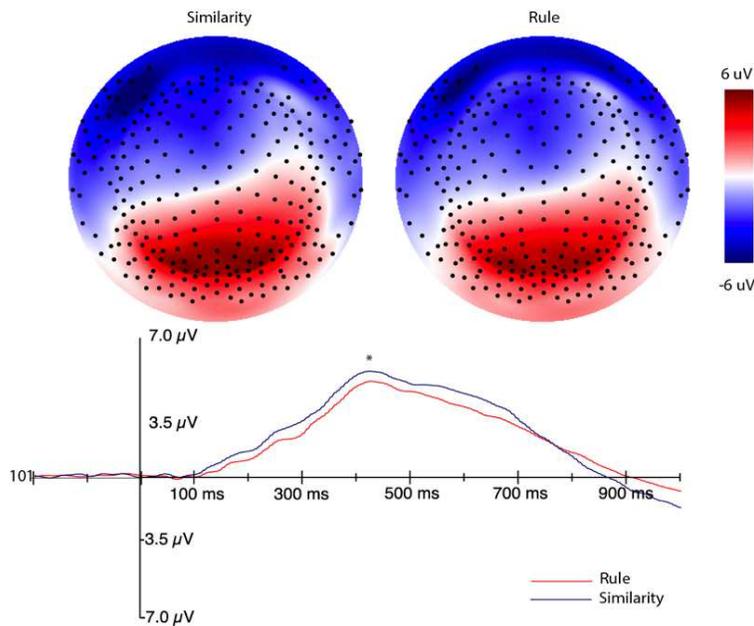


Fig. 23. Voltage map and representative waveform of the P3b. Top: A voltage map displays the voltage across the scalp for the rule and similarity-based conditions at the peak of the P3b (asterisk in bottom waveform image). A stronger positive voltage is seen over the posterior parietal areas for the similarity-based condition. Bottom: Representative waveform showing the shape of the P3b for both conditions. The amplitude of the P3b is higher (more positive) for the similarity-based condition.

LIAN

The LIAN was quantified by utilizing a set of channels in the left and right frontoparietal regions (orange and yellow electrodes in Fig. 21, respectively). An adaptive mean amplitude corresponding to 22 ms before and after the peak negative amplitude in a window extending from 450 - 950 ms (the same window as the P3b) was used to quantify the component. This method was applied for all post-learning trials for all three categories in each subject. Similar to the P3b and MFN, a single ERP was computed for the rule-based and similarity-based categories for both the left and right LIAN. A paired-samples t-test showed that the amplitude of the left LIAN was largest for the similarity-based category ($-7.06 \mu\text{V}$) compared to the amplitude of the rule-based category ($-5.54 \mu\text{V}$), $t(43) = -2.98$, $p = .004$ (Fig. 24). However, no significant amplitude difference for the right LIAN was found between the rule-based ($-3.55 \mu\text{V}$) and similarity-based category ($-2.92 \mu\text{V}$), $t(43) = 1.23$, $p = .23$ (Fig. 24).

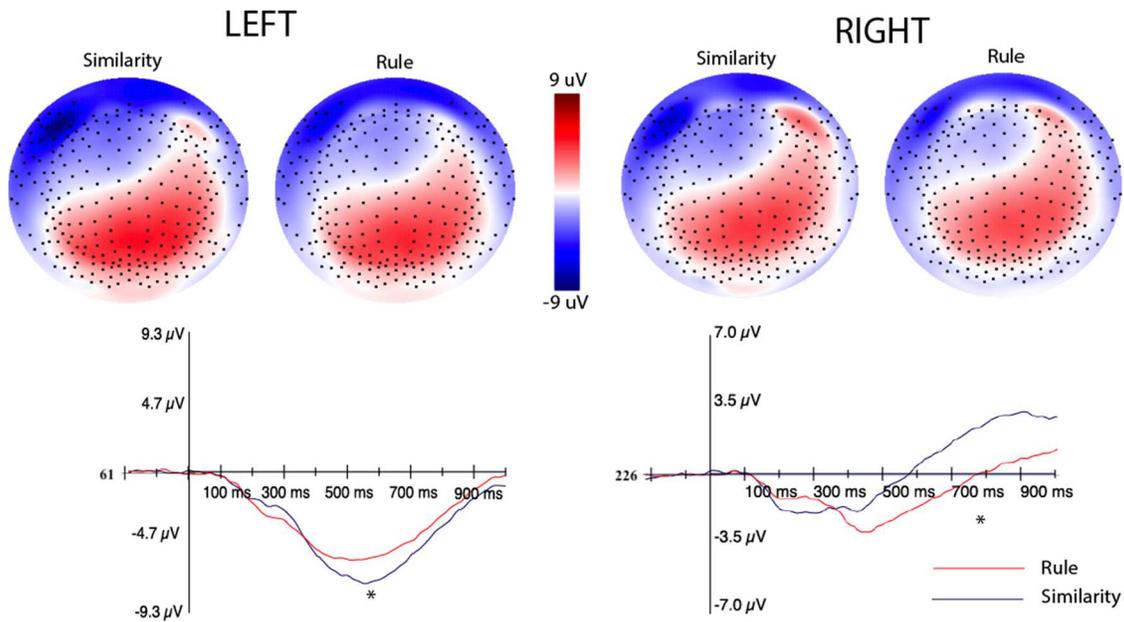


Fig. 24. Voltage maps and representative waveforms of the left and right LIAN. Top: Voltage maps display the voltage across the scalp for the rule and similarity-based conditions at the peak of the LIAN on the left and right sides (asterisks in bottom waveform images). A stronger negative voltage is seen over the left frontal areas for the similarity-based condition and a stronger negative voltage is seen over the right frontal areas for the rule-based condition. Bottom: Representative waveforms showing the shape of the LIAN for both conditions in the left and right hemispheres. The amplitude of the left LIAN is higher (more negative) for the similarity-based condition, whereas the right LIAN is higher (more negative) for the rule-based condition.

Machine Learning

Data Preparation

For every subject, post-learning trials were chunked into individual segments extending 200 ms before and 1,000 ms after stimulus onset for each category. Segments containing ocular or movement artifacts were rejected from analysis. Each segment was baseline corrected using a 200 ms pre-stimulus baseline before averaging the segments together to form one averaged waveform for each category of stimuli. Waveforms for the two rule-based categories were averaged together to be compared against the similarity-based category before re-referencing to an average reference. The waveforms were then broken down into their individual samples, which at a sampling rate of 250 samples/second resulted in 300 total samples per waveform (each sample representing 4 ms of recording).

In order to reduce the number of predicting elements in this analysis, we averaged together the raw voltages of electrodes within 10 regions: left frontal, right frontal, medial prefrontal, medial frontal, posterior parietal, left temporoparietal, right temporoparietal, left occipital, right occipital, and medial occipital (Fig. 25). This process was done for each individual sample for both categories. We then averaged together every 5 consecutive samples together, resulting in 60 timepoints for each waveform with every timepoint representing 20 ms of data. However, because the first 10 timepoints were used in the baseline correction, we did not include these in the analysis. In the end, this gave us 2 matrices (one for rule-based and one for similarity-based) for each subject with dimensions 50 (timepoints) x 10 (electrode groups).

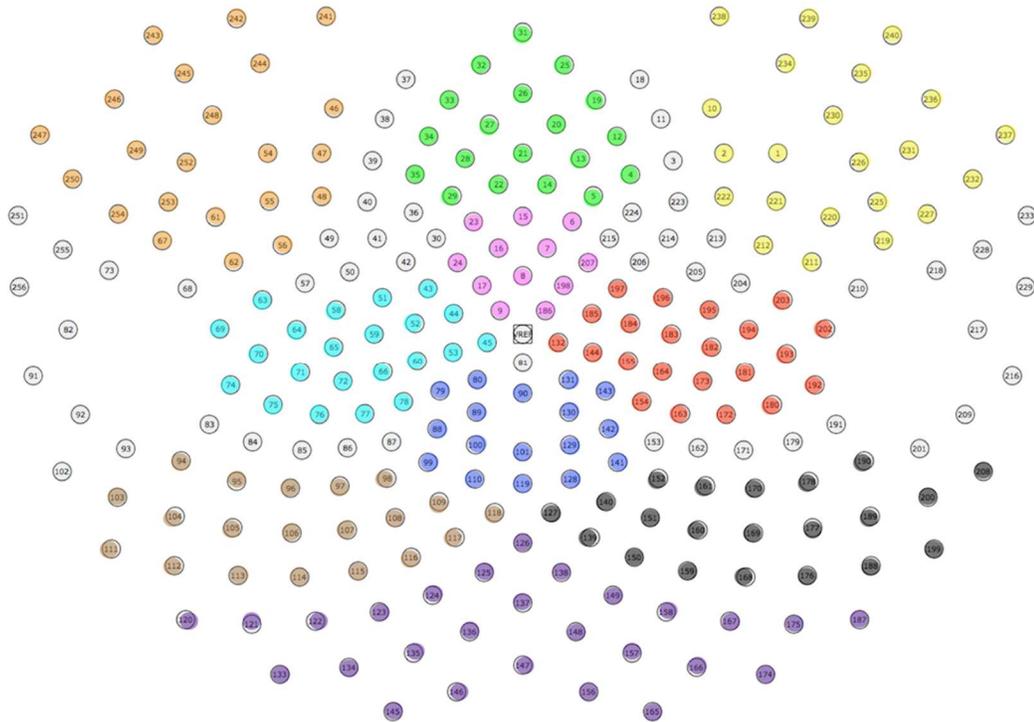


Fig. 25. Electrode montages used to define regions during machine-learning analysis. Orange = left frontal, yellow = right frontal, green = medial prefrontal, pink = medial frontal, blue = posterior parietal, cyan = left temporoparietal, red = right temporoparietal, brown = left occipital, purple = medial occipital, black = right occipital.

Whole-Brain SVM Classification

For each individual timepoint, a Support Vector Machine (SVM) was trained on the voltages of all 10 electrode groups for both categories before performing binary classification (rule vs. similarity) on a set of test data. Leave-one-subject-out cross validation was used, such that 43 out of the 44 subjects were used to train the classifier, and the subject that was left out of training was used as the test subject. This type of training and test format was performed iteratively until all subjects were used as a test subject. For each iteration, the classifier made a prediction of whether the pattern of voltages for the test subject's data reflected a rule-based or similarity-based category. Hypothetically, the classifier could label both pieces of test data as belonging to the same category even though the test data always included exactly one rule-based and one similarity-based array of voltages. To force the classifier into using each category label only once per subject, we evaluated the classifier evidence for the two categories in both predictions. The classifier prediction with the highest amount of evidence for a category was always given the prediction label for that category, and the other prediction was always given the opposing label regardless of the amount of evidence for both categories.

The classification accuracy of all iterations were averaged together to get a more accurate classification accuracy for each timepoint. A one-sample t-test was performed for each timepoint against a hypothesized mean of 50% accuracy (chance). The cross-validated classification accuracy for each timepoint is plotted chronologically in Fig. 26, and timepoints that had a classification accuracy significantly above chance at a significance level of below .05 are denoted by a blue diamond along the X axis. From this figure, we can clearly see that the earliest timepoint at which the classifier was able to

reliable differentiate between the two categories is between 260 and 320 ms, which coincides with the onset and peak of the MFN. Classification accuracy dips below significance between 320 and 420 ms, but returns to significantly above chance between 440 and 700 ms, corresponding to the peak and onset of the LIAN and P3b.

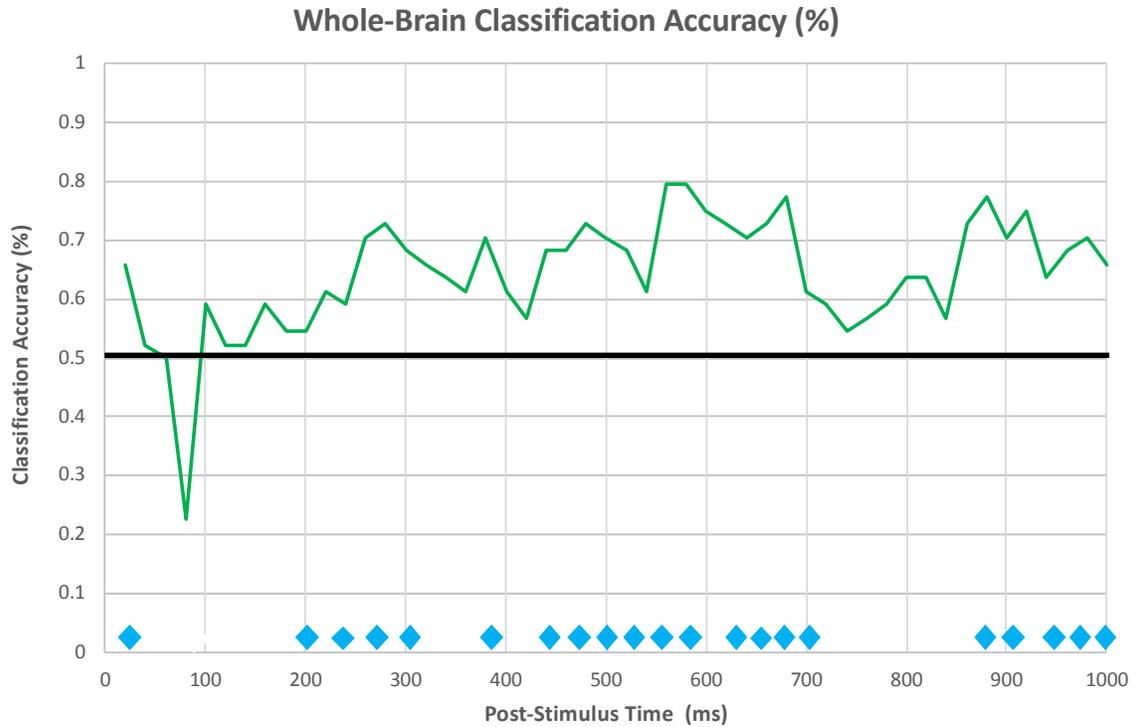


Fig. 26. Whole-brain classification accuracy over time on an experimental trial. Blue diamonds along the X-axis represent timepoints where classification accuracy is significantly above chance (i.e. $p < .05$). The earliest string of above-chance classification accuracies is observable between 200 and 300 ms after stimulus onset, followed by another group between 430 – 700 ms. A late string of reliable classification occurs around 890 – 1,000 ms.

Region-Based SVM Classification

To determine if any one particular region was driving the classification accuracy at each timepoint, the same SVM classification was run again using only the voltages in each region individually. A separate classification accuracy plot (over time) was created for all 10 regions. As an additional measure, we ran a Pearson’s correlation between the classification accuracy of the SVM and the behavioral performance on the categorization

task of each subject. This analysis allowed us to determine if reliable pattern differences between the two systems lead to better or worse performance on the task for each region.

Results from this analysis indicated that the medial prefrontal, left frontal, and posterior parietal regions show the earliest reliable (and strongest) classification accuracy amongst all regions, with a maximum classification accuracy of 82% (Fig. 27). Within these regions, reliable differentiation between the two categories occurs around 250 ms, and remains stable until around 740 ms. However, classification accuracy peaked earlier in the posterior parietal region compared to the medial prefrontal and left frontal regions, even though we can differentiate between the two categories with reliable accuracy using any of these 3 regions within the entire 500 ms window. Interestingly, the classification accuracy of the medial prefrontal region did not predict behavioral outcome at virtually any timepoint, whereas the left frontal region, which is the location of the left LIAN component, was positively correlated with behavior throughout its classification peak. Classification accuracy of the posterior parietal region was surprisingly negatively correlated with behavior between 440 and 520 ms; the initial positive deflection of the P3b. Of all the regions, the right frontal area (the location of the right LIAN) was responsible for the very latest classification accuracy peak, occurring between 800 and 1,000 ms. Classification accuracy in this region did not significantly correlate with behavior within this window. The three occipital areas along with the two parietal areas failed to demonstrate a consistent windows of reliable classification accuracy.



Fig. 27. Region-based classification accuracy over time and correlated with behavioral performance. Top: Classification accuracy for the left-frontal electrode montage. Classification accuracy peaks between 400 and 700 ms. During this timeslot, classification is positively correlated with performance. Middle: Classification accuracy for the medial frontal electrode montage. Accuracy peaks between 600 and 750 ms after stimulus onset and does not correlate with behavior in any way. Bottom: Classification accuracy for the posterior parietal electrode montage. Accuracy peaks the earliest in this region, occurring between 220 – 500 ms. Interestingly, classification accuracy is negatively correlated with behavioral performance within this window.

Discussion

The main goal of this experiment was to determine the earliest timepoint at which the brain differentiates between two memory systems during a categorization task. The results showed that, once participants acquired the task, clear differences in the Left LIAN, MFN, and P3b components were seen between the two conditions. However, the direction of each effect appeared to be contrary to what was predicted. Overall, the amplitude of each ERP that reflected a difference between the rule-based and similarity-based categories was largest for the similarity-based category. However, amplitude for the right LIAN was larger for the rule-based category, although this effect did not reach statistical significance. For the machine learning analysis, classification accuracy peaked earliest in the posterior parietal region (the location of the P3b), but reliable classification could be performed using additional electrode clusters including the left prefrontal and medial prefrontal areas. Electrode clusters over the lateral occipital areas, which were preferentially recruited when subject performed a similarity-based categorization in Chapter III, failed to produce accurate classification between categories.

ERP Analysis

MFN

Originally we hypothesized that the amplitude of the MFN would be largest for the rule-based category compared to the similarity-based category. This assumption was based on past research that has established rule-based categorization primarily engages the working memory system and regions associated with declarative recall. However, the MFN amplitude in this experiment was larger for the similarity-based category, albeit the significance of this effect was only marginal. The moderate difference in amplitude

between our categories support recent findings that suggest rule-based *and* perceptual similarity-based categorization need executive functions in order to select the memory system that is optimal for a task (Miles, Matsuki, & Minda, 2014). For stimuli that would benefit most from perceptual similarity, this requirement of effortful control would be very brief -- commencing well before an action is committed (Miles et al., 2014). The latency of the MFN (180 - 300 ms) corresponds to the initial orienting of attention in a visuomotor association task, and thus we propose that the MFN in our task is indexing the controlled attention required to select the memory system best suited for categorizing the presented stimulus and does not depend on the optimal system needed to perform a task.

P3b

The amplitude of the P3b in our experiment was significantly larger for visually distinct stimuli when compared to stimuli that required the application of an explicit rule for categorization. Our hypothesis for this component was that the amplitude should be largest for the rule-based category, but this would only be the case if participants were exclusively relying on perceptual similarity to categorize members of the visually distinct category. This hypothesis was formulated based on results from Chapter III, which showed a robust activation of posterior visual cortex for the visually distinct category. However, subtle differences between the fMRI and current EEG versions of the task could result in subjects utilizing different systems to categorize formations in the visually distinct group.

The fMRI version of our task included 36 exposures per stimulus, whereas the EEG version had 80 to account for the higher presence of noise in our recordings. It

seems probable that having double the number of repetitions raises the chances of participants developing explicit rules or using declarative recall for categorizing formations in the visually distinct category through more trial and error. Indeed, when we analyze the strategies subjects used in the EEG version, 89% of participants reported using an explicit counting rule (e.g. “I counted 6 people on the line of scrimmage”) or declarative recall (e.g. “I memorized each formation individually”) for categorizing the visually distinct formations, while only 11% reported using a perceptual similarity strategy (e.g. “There appeared to be a lot of people on the line of scrimmage for formations in this category, such that I did not need to count any players”). Contrasted with the strategies used by subjects in Chapter III, 92% of participants reported using a perceptual similarity strategy. Based on these differences, we can interpret the P3b in our experiment as indexing posterior parietal engagement during declarative recall, similar to the theory described in Chapter II.

As for the amplitude difference between our categories, no study to date has directly compared the amount posterior corticolimbic engagement during declarative recall of individual stimuli and the application of an explicit rule, although results from several different studies discussed earlier suggest both of these methods rely on the same memory system. The larger amplitude observed in the visually distinct category could be indicating that the recall of specific stimuli engages the posterior parietal region more than the recall of an explicit rule. A more controlled experiment that directly contrasts these two strategies with more spatial resolution is required.

LIAN

The amplitude of the left LIAN was largest for the similarity-based condition whereas the right LIAN was largest for the rule-based condition, although the latter effect did not reach statistical significance. The left/right conditional flip makes the interpretation of this component fairly difficult. At this time we are unsure if both components are interpretable on their own, or if the LIAN is a hemisphere-specific component and the effect observed on the contralateral side is a byproduct of volume conduction. Given the lack of attention the LIAN has received in the literature, our current interpretation is mostly speculative.

Luu et al. (2007) found that the amplitude of the right LIAN *decreased* as subjects acquired the ability to perform spatial analyses in a visuomotor association task, but the amplitude of the component remained unchanged when the targets in the task were digits that evoked the phonological loop. They also found that the amplitude of the left LIAN *increased* as subjects acquired digit targets in their task, whereas the amplitude remained unchanged as they acquired the ability to perform spatial analyses. Motivated by the findings of their experiment, we drew an initial assumption that the amplitude of both the left and right LIAN should be largest for the rule-based condition in our experiment. As similarly discussed in our interpretation of the P3b, however, this would only be the case if subjects relied exclusively on perceptual similarity analyses to categorize formations in the visually distinct category – similar to the spatial analyses performed in Luu et al. (2007).

Given the vast majority of subjects in our experiment used rote memorization to categorize the similarity-based condition instead of the hypothesized perceptual similarity, we interpret our findings as a contrast between declarative recall of individual

stimuli (visually distinct category) and explicit rule application (visually similar categories). When viewed from this perspective, the location of the LIAN coincides with structures that are essential for both forms of analysis, such as the temporal lobe and inferior frontal gyrus (IFG) (Toni, Rammani, Josephs, Ashburner, & Passingham, 2001; Grol et al., 2006). In addition, the latency of the LIAN is in-line with the action selection and memory retrieval processes, indexed by how closely it mirrored subject reaction times. Based on the higher accuracy and shorter reaction times for the visually distinct category, our findings that the right LIAN was smaller for this category matches meta-analytic findings that show a right hemisphere-specific reduction in anterior temporal and IFG activity with the development of expertise in visuomotor tasks (Chein & Schneider, 2005). We could be seeing right hemisphere-specific reductions in the attentional resources needed to categorize the visually distinct group of formations simply because our subjects are consistently at a more advanced stage of learning for this condition compared to the rule-based condition. This reduction could be reflecting the shift towards the automaticity system (seen in our earlier EEG experiment), which is involved in more routinized or habitual memory retrieval. Our left LIAN results also become more interpretable through this lens. If our subjects are significantly more advanced at declaratively recalling the visually distinct formations, then we would expect the left LIAN to be larger for this condition based on the findings of Luu et al. (2007). The amplitude of the left LIAN linearly increased for digit targets in their visuomotor learning task which theoretically engage the same explicit form of memory as both conditions in our experiment. Thus, the left LIAN differences seen in our study could be reflecting

differences in expertise between our subject's ability to categorize the visually similar and visually distinct categories.

ML Analysis

Using machine learning we were able to successfully dissociate between our two conditions when utilizing raw voltages distributed across the entire scalp. The ability to predict the category of a currently viewed stimulus using the entire brain is not surprising given that the number of predictors (10 regions) greatly outnumbered the prediction outcomes (2 conditions). However, the novelty of our approach lies in the timepoint-by-timepoint classification that helps us understand the earliest point at which we can differentiate between our conditions as a subject views a stimulus. This differs from standard approaches in fMRI and other EEG classification attempts which either average over several seconds of activity or ignore the time-domain altogether. In our study, the earliest reliable dissociation point was around 220 ms after stimulus onset, which coincides with the initial onset of the MFN ERP component. We interpret this early classification timepoint as reflecting the initial controlled attention required to select a memory system based on the stimulus being presented.

To understand which individual regions were driving the classification accuracy at each time-point, we ran a second machine learning analysis on only the voltages of single groups of electrodes in 20 ms intervals. Our results from this analysis showed that the medial prefrontal, left frontal, and posterior parietal regions collectively contributed to the earliest reliable classification point. Results using MVPA in fMRI studies have consistently demonstrated that individual rules can be reliably decoded in frontal and parietal regions (Woolgar, Thompson, Bor, & Duncan, 2011; Reverberi, G6rger, &

Haynes, 2012; Nelissen, Strokes, Nobre, & Rushworth, 2013). Our results in the current study expand on these findings by specifying that the pattern representations of these concepts coincides with the initial orientation of attention. Through sufficient trial and error learning, the context under which an action is learned in a visuomotor task becomes tied to each individual stimulus in the task (Donchin & Coles, 1988). Since we only analyzed trials after our subjects had been sufficiently trained on the task, we can assume that the initial conscious registration of a stimulus prompted a conditioned re-establishment of the explicit rules (the learning context) that would dictate their subsequent action selection. This theory could explain why the first pattern dissociation between our two categories happens around the earliest time that a person can explicitly orient attention.

Although the spatial resolution of our EEG machine learning analysis is limited by default, the results from the current analyses are further supported by the strong classification accuracy of the IFG and superior parietal region discussed in Chapter III. We must caution that these spatial shortcomings may make it difficult to effectively detect patterns at the individual-stimulus level as is common in fMRI. Our inability to reliably dissociate between our categories in visual and other sensory areas leads us to believe that each electrode records data from too many sources to have the fine resolution necessary to detect individual item representations. Thus, we limit our interpretation of the EEG machine learning findings as dissociating between more general concepts such as rules and action selection.

Limitations

One limitation of the current study is the lack of continuity in strategies used by subjects in Chapter III. Ideally, subjects would have consistently utilized the same strategies between both studies to ensure the systems engaged between the two task versions are the same. It is possible that subjects in the current study used the hypothesized perceptual similarity analysis at some point in training, but with no way to label trials within this time period we may be forced into a future study which establishes control over the strategies used in the experiment.

Other limitations in our study are generally common when using EEG. Since we only analyzed stimulus-locked ERP components, we are undoubtedly attenuating the amplitude of our components by not controlling for reaction time variances between our conditions. Future analyses will need to control for RT effects, such as in joint-time-frequency (JTF) or constructing ERPs around trials with similar RTs to better understand amplitude differences across our conditions. Volume conduction may also play a role in the uncertainty of our LIAN results, as at the time of this writing we cannot ensure that the laterality effects are independent of one another or a simple byproduct of EEG dynamics. We can investigate this concern by using independent components analysis (ICA) in future analyses of this dataset.

Conclusions

The goal of the present study was to demonstrate the feasibility of dissociating between our two conditions using EEG and determine the earliest timepoint on a trial that we can reliably do so. Overall, the general location and known processes associated with the dissociating ERPs in this study do not map well onto the fMRI results discussed in Chapter III. However, some enthusiasm is restored when looking at pattern

representations of our conditions using region-based machine learning. Whereas our fMRI results showed us *where* in the brain we can find reliable patterns for dissociating between two memory systems, the results from this chapter establish *when* these patterns become reliable during a categorization trial. Results from this study serve as a stepping stone toward combining neuroimaging modalities to offset the weaknesses of each method and better describe the basic memory processes taking place in the human brain.

CHAPTER V

GENERAL DISCUSSION

Review of Empirical Results from Dissertation

The main goal of this dissertation was to understand the degree to which distinct learning and memory systems may be recruited within a single task. In Chapter II, we conducted an EEG study where we used ERPs to correlate changes in cortical electrophysiology with marked performance improvements during category learning. The results from this study indicated that frontal and posterior regions were both essential for the development of expertise in the task. A possible interpretation of these findings is that multiple memory systems (rather than a single system) which may have different working memory requirements were engaged in the task. To empirically test this hypothesis, we designed an fMRI experiment (Chapter III) that was optimized for being able to identify distinct memory systems within a single task. Consistent with our predictions, we found that two different memory systems, the rule-based and perceptual similarity systems, were alternated between depending on the category structure presented. Finally, we tested whether the same dissociations may be observed when using EEG in Chapter IV, allowing us to better evaluate the timing of system engagement within a trial. We found evidence for distinct systems that differentiate very early during a categorization trial, around 250 ms after viewing a probe. However, we also found evidence that subjects shift away from relying on the two systems found in fMRI with extended training. This suggests that multiple memory systems may interact to optimize performance during the early learning stage before transitioning to the automaticity system.

Earlier we discussed a general two-stage model of learning; one that describes an anterior-posterior shift in brain activity as a person develops expertise. However, the results in our first EEG study detracted from classic models of the distinct brain mechanisms associated with each stage of learning, with anterior regions becoming increasingly important beyond the development of expertise. In retrospect, these models were primarily constructed around how expertise develops as people memorize individual exemplars, which is a strategy that is often viewed to be of lesser importance in category learning (Ashby & Waldron, 2000). The lack of supporting literature surrounding the phenomenon lead us to ponder the possibility that participants in our task were utilizing other methods of categorization beyond exemplar memorization, such as methods that rely on frontal control regions well after learning occurs.

A large number of studies have outlined the behavioral and neural processes associated with different methods of categorization (see Chapter I for a review). The overwhelming consensus amongst these studies is that the different categorization systems serve the purpose of making learning as efficient as possible under different learning conditions. These strategies rely on distinct memory systems. A common feature of category learning studies is that they use tasks that are designed to recruit memory systems one at a time. Yet, they consistently pose an un-tested assumption that real-world learning relies on an innate ability to switch between memory systems. Through empirical testing throughout this dissertation, we established firm evidence that people can switch between memory systems to optimize performance in a single task. We demonstrated that subjects recruit the perceptual similarity, rule-based, and exemplar memorization systems depending on the structure of the stimulus being observed as well

as the amount of exposure a subject has had with that stimulus. We also determined the time course by which the brain shows dissociable neural signatures signifying the selection of these different memory systems. The memory systems described in our experiments have a differential reliance on controlled attention mediated by the frontal cortex, and can potentially help us explain the mechanisms associated with the stages of learning in our first EEG experiment. In the next section, we will describe these systems in greater detail, and highlight how the structure of the experiments in this dissertation lead to the recruitment of each system individually.

Characteristics of Multiple Memory Systems Used in Category Learning

Exemplar Memorization

Our hypotheses regarding the initial experiment looking at expertise development were based solely around distinct processes that happen during learning to remember individual exemplars. Exemplar memorization requires a conscious effort to evaluate a stimulus and search through individual items in memory to place an exemplar into the right category (Ashby & Ell, 2001; Smith & Minda, 2001). Over time, the amount of controlled attention required to recall individual exemplars decreases – marked by a noticeable reduction in frontal control regions and an increase in medial temporal lobe engagement (Shiffrin & Schneider, 1977). The arbitrary category structure used in our first EEG study *should* have forced subjects into remembering individual formations, as arbitrary/poor category structure is one instance where exemplar memorization is preferred over other methods (Lei & Zhansheng, 2003). Our assessments in this chapter established that subjects were performing under an automatic mode of operation during the extended training days, yet the frontal areas became increasingly important to their

success in the task. Despite our attempts to randomize the formations and the categories they belong to, we cannot discount the potential for subjects to develop their own set of explicit rules for making categorization decisions. When viewed from this perspective, the linear increases in both frontal and posterior parietal engagement throughout learning support the idea that subjects relied on rule application during this task, which in turn motivated the careful restructuring of our categories during the fMRI experiment.

Rule-Based Categorization

Rule-based categorization is a memory system based-on explicit memory (Ashby et al., 1998). The development and application of explicit rules relies heavily on the working memory system – this typically holds true throughout the learning process (Ashby & Ell, 2001). A conscious effort is needed to analyze the stimulus being presented, recall the rule associated with the stimulus, and apply this rule to make a decision about category membership. This method of categorization is particularly useful when between-category similarity is high, such as a subtle location difference in one player on the line of scrimmage as seen in the visually similar categories used in our more structured category learning experiments. Indeed, our fMRI results supported our assumption that subjects rely on rule-based categorization to dissociate between categories with visually similar structures, indexed by the importance of frontal control regions correlated with successfully categorizing members of these groups. Looking back at Chapter II, had our subjects been relying on a rule to place exemplars into categories, then ERPs that index controlled attention (i.e. the MFN) should have increased as they did. However, our attempts at confirming this assumption were complicated by several

unsupported hypotheses when converting the experiment into a format that is compatible with EEG.

Interactions Between Exemplar Memorization and Rule-Based Categorization

As discussed earlier, a stark difference between the experiments utilizing fMRI and EEG was the number of repetitions used in each. Due to the nature of building ERPs, EEG often requires a considerably higher number of trials to measure conditional effects compared to fMRI. Although our assumptions were not supported using EEG, we know from each subject's scores on the generalization block that they had discovered the explicit counting rule. This technique was further confirmed by their description of the strategies they used on the debriefing questionnaire. Yet, the ERPs for the rule-based condition during training more closely matched what we would expect if they had memorized each exemplar individually in the two visually similar categories. Specifically, the amplitude of the MFN was smallest for the rule-separable condition compared to the visually distinct category. To help explain this result, we must consider the possibility for subjects to switch their strategy (and underlying memory system) from one that is conducive to maximizing performance during initial training towards another strategy that maximizes efficiency and automation in later learning stages.

At the time of this dissertation, a series of literature that specifically focuses on strategy changes throughout category learning does not exist. The category learning literature describes the various memory systems as being independent, without providing a clear narrative for how learning occurs within each system or if transitioning between them signifies different stages of learning. On the other hand, the expertise literature commonly describes the learning stages within the context of declarative memory, with

few references to how the mechanisms within each stage relates to a distinct memory system (Ashby et al., 1998; Ashby & Waldron, 1999; Medin, 1975; Medin & Schaffer, 1978). Logan's (1988) instance theory is perhaps the closest attempt at reconciling these two bodies of literature. Instance theory describes the development of automaticity as a shift from calculated algorithmic processes towards the retrieval of more specific memories from past actions, but does not specify how these processes are supported by dissociable memory systems. In our own attempt to resolve these differences, we postulate that the process of changing strategies as one becomes proficient in explicit category learning is synonymous with the distinct mechanisms associated with the stages of expertise development. Both changes serve the function of reducing effort to allow attentional resources to be available for other processes, and share the commonality that separate neural systems are associated with the measurable changes in behavior (Gabriel et al., 2002; Chein and Schnieder, 2005).

Relevant to the experiments in this paper, we propose that rule-based and exemplar categorization can be viewed as intermediate strategies when learning to classify exemplars. The end-goal of both systems is to work towards a routinized method of categorization, supported by a posterior automaticity system. In the intermediate stages of learning, these systems can develop a reliable means for defining the relevant features that dictate category membership. An example of this would be our fMRI results, where the experiment was shorter and did not allow enough repetitions for subjects to move toward automaticity. With extended exposure to the same stimuli, the rule-based and exemplar memorization systems give-way to other systems that automate the attention process, such as the automaticity system centered on posterior corticolimbic structures.

An example would be our follow-up results using EEG, where several hundred exposures to the same stimuli prompted subjects to memorize each formation individually. The conditions under which rule-based categorization would not be beneficial as an intermediate strategy include instances where categories are small (only a few exemplars), or when exemplars include exceptions to a categorization rule (Minda & Smith, 2001; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994). For situations where neither of these strategies fit the conditions of the task, categorization strategies that do not rely on explicit memory may be more desirable.

Perceptual Similarity Categorization

Our category learning tasks utilized a visually distinct category of football formations as a separate condition to compare against the two visually similar categories discussed in the previous section. We hypothesized that subjects would rely on the perceptual similarity memory system when categorizing exemplars of the visually distinct category. The perceptual similarity system specializes in rapid identification of stimuli that are perceptually similar to one another, and has been shown to be independent of several forms of explicit memory – including exemplar memorization (Schacter, Cooper, Tharan, & Rubens, 1991; Wagner, Gabrieli, & Verfaellie, 1997). These findings stem from studying perceptual similarity within the repetition priming framework, where quick categorization decisions are made for an individual stimulus even when a subject cannot explicitly recall if they had seen that particular stimulus on prior trials. The clear dissociation between perceptual similarity and multiple explicit memory systems has led category learning researchers to classify this system as a form of implicit, non-verbalizable memory (Tulving & Schacter, 1990). Yet, the conditions under which the

perceptual similarity system is engaged spans a variety of different task contexts including some where successful strategies are easily verbalizable.

Classically, the perceptual similarity system has been associated with prototype learning tasks (Keri et al., 2002). Recall that in these types of tasks subjects are exposed to exemplars with varying degrees of distortion from a central category prototype (i.e. a representative example of the category) (Posner & Keele, 1968; Posner & Keele, 1970). Often, the subject never sees the actual prototype itself, and instead creates a mental representation of the prototype by abstracting relevant features from perceptually similar members of the same category. Taking a closer look at the visually distinct category in our tasks, we do not believe the structure of this category was conducive to prototype abstraction and does not fit the implicit model described in the literature.

In the visually distinct category, the relevant players on the line of scrimmage were distributed such that a central prototype would not be the most efficient method for accurately categorizing each exemplar. Matching each exemplar to a manifested prototype would sacrifice speed over them simply knowing that a large sum of players on the line of scrimmage determined the identification of formations in this category. In addition, our subjects' strategies for identifying members of this category consisted of verbalizable rules in both the fMRI and EEG versions of the task. The behavioral results in our experiments clearly detract from the canonical conditions under which the perceptual similarity system should be engaged during category learning. However, the robust engagement of sensory cortices our fMRI experiment leave us convinced that this system was essential to subjects' success.

Perhaps a more accurate way to describe the categorization strategy used for the visually distinct category stems from context theory (Medin, 1975; Medin & Schaffer, 1978). Under this theory, subjects are assumed to rely exclusively on stored exemplar information. Each new or old probe acts as a cue to retrieve information associated with other stimuli that are visually similar to that probe item. This diverges from prototype theory, such that it hypothesizes subjects do not make references to abstracted prototypes when categorizing a probe. Instead, it is the visually similar information shared amongst the probe and stored exemplars that is cross referenced and ultimately leads to a categorization decision (Medin & Schaffer, 1978; Nosofsky, 1984). In our own experiments, this would be the large group of players on the line of scrimmage that are the essential element dictating category membership. It is more probable that subjects compared this group of players within each exemplar to other exemplars of the visually distinct category they had recently come into contact with.

An open question still remains whether subjects were relying on perceptual similarity or declarative recall when categorizing the visually distinct category in our final EEG experiment. One explanation is similar to the rule-based condition discussed earlier: subjects relied on perceptual similarity for a short period of time before transitioning to declarative recall associated with the automaticity system. This proposal is in-line with Logan's instance theory, but would require labeling perceptual similarity as an algorithmic rule that eventually transitions to exemplar memorization to better fit the model. Given the poor spatial resolution of EEG combined with the trial count differences between the EEG and fMRI tasks, our interpretation is only speculation.

Category Learning Strategies as a Function of Expertise

The theories of categorization outlined thus far commonly discuss these memory systems individually. However, a novel finding from our experiments is that categorization systems can develop alongside one another in a single task, alternating from trial-to-trial to meet task demands. The development of expertise within each system happens independently, but they share the same end-goal of automating the attention process with extended training.

Palmeri (1997) made one of the first attempts at describing the time it takes subjects to reach automaticity using perceptual similarity versus rule-based categorization. In one experiment, Palmeri had subjects categorize objects with high within-category similarity, whereas in a separate experiment had subjects categorize objects with high between-category similarity which required the discovery of a rule. The results from these experiments demonstrated that subjects utilizing perceptual similarity reached automaticity notably faster than those relying on rules (Palmeri, 1997). This led to the development of a new theory termed Exemplar-Based Random Walk (EBRW), which is a synthesis between Logan's (1988) Instance Theory and an alteration of Medin & Schaffer's (1978) Context Theory (Nosofsky, 1984; Nosofsky, 1986). EBRW proposes that, when a probe is presented, exemplars stored in memory race to be retrieved with a speed that is proportional to their similarity to the probe. Each one of the retrieved exemplars drives a random walk until sufficient evidence is presented. Once enough evidence has been retrieved, a subject makes a response. In other words, decisions during category learning are driven by a race between stored exemplars *and* a calculated rule-based/algorithmic process (Nosofsky et al., 1994; Palmeri & Nosofsky, 1995).

Computational models of EBRW allow for reaction times to be sped up by increasing within-category similarity and increasing the number of exposures to an exemplar (Palmeri, 1997). This would result in a shorter training period before subjects reach automaticity when categorizing visually similar exemplars. The model also accounts for a longer training period when subjects are forced to rely more on the random walks or the evidence-gathering aspect of the process when categories have low within-category similarity, which can be the case in rule-based categorization. EBRW, when interpreted on a purely conceptual level, helps explain how implicit and explicit forms of categorization are a simple function of expertise development. The different strategies are called upon depending on the structure of a category being presented, and share the common function of serving as an intermediate strategy before transitioning to an automatic mode of operation. However, a major flaw of EBRW is that it infers the race between exemplars and the random walks are the process of a single, *unitary* memory system that performs these operations in series. We propose that this theory be altered to accept these processes as the work of distinct memory systems. It is clear that future work is needed to develop new theories for how these distinct systems develop under learning conditions that may require more than one type of system to optimize performance.

Practical and Theoretical Implications

Understanding the conditions under which different learning and memory systems are engaged has many theoretical and practical implications. Regarding theory, when we accepted the possibility that our category structures may have contributed to the unexpected findings in Chapter II, we were able to conduct a new series of experiments that showed the successful switching between memory systems to accomplish a common

goal. This increased our knowledge of how memory systems are utilized in the real-world by establishing the first piece of direct evidence of something that only been posed as a theory predating this dissertation. We took these results a step further by focusing on describing the timings by which a memory system is selected. In the process we borrowed techniques from another neuroimaging method (fMRI), but applied them in a novel way which helps us understand when, on a millisecond timescale, pattern representations in the brain become reliable enough to perform classification. We expect that this new method of machine learning analysis will be adopted and refined by researchers spanning well beyond the domain of learning and memory.

From a practical standpoint, our experiments motivate a careful look at how learning may be hindered in the real world by failing to structure training plans around the memory systems optimized for a task. Relatedly, and although learning can be slowed by many different factors, it is likely that a failure to recruit the appropriate system for categorization can lead to poor performance. It seems essential to work toward optimizing training paradigms and simultaneously helping people recruit the appropriate memory systems to make learning as efficient as possible. This starts with having a firm understanding of the memory systems underlying several forms of learning.

In addition, our results have practical implications for the clinical realm, mainly in demented diseases such as Alzheimer's (AD) and Parkinson's (PD). Evidence from demented patients with AD and PD has demonstrated that memory loss as a result of each disease can be attributed to the degradation of 1-2 memory systems, with only a small minority showing global memory loss in the early and middle disease stages (Heindel, Salmon, Shults, Wallcke, & Butters, 1989; Kuzis et al., 1999). Having a better

understanding of the individually affected memory systems in such diseases can lead to more targeted interventions to rehabilitate memory loss. For example brain stimulation can be used to help facilitate the engagement of the degrading memory system, as well as facilitate switching between intact systems. Within this line of research, attempts at restoring motor function in PD using brain stimulation have already shown promising effects, but the efforts to improve memory are lacking (Fregni, Simon, Wu, & Pascual-Leone, 2005).

Future Directions

The results from this dissertation suggest multiple memory systems work in an alternating fashion to maximize success during category learning. Although in our experiments we only focused on the presence of each system, future work is needed to fully understand the development of expertise within each of these systems. It would be beneficial to understand the point at which subjects stop relying on algorithmic operations, and switch to exemplar memorization as seen in our final EEG study. Having a way to ensure that our individual ERPs are coming from subjects that are at the same learning stage will allow us to better separate the cortical signatures of the distinct memory systems. Gaining this control will also aid in mapping our ERP results onto those measured with fMRI. Currently, our understanding is that the systems used in both studies are not the same. Gaining control over which strategies (and underlying memory system) our subjects are relying on in both experiments will allow us to reconcile results from both imaging modalities.

Questions still remain regarding if or how the rule-based, perceptual similarity, and exemplar memorization systems interact during categorization. We now have a firm

understanding that they can be flexibly engaged within a task. In addition, our EEG results indicate that memory systems are selected very early within a trial – around 250 ms after viewing a stimulus. However, we are limited in the information we can extract from the current results regarding the selection process.

Finally, we can continue to ask questions about how best to integrate our results into the practical applications discussed earlier. In the context of brain stimulation, understanding the neural signatures of different memory systems will enable us to establish confidence in the areas we target during these interventions. Future work can evaluate the extent to which stimulation over the areas associated with the memory systems discussed in this paper leads to enhanced learning. We may also use the temporal information described in our final EEG experiment to inform the optimal timing to deliver an intervention during a single trial. Many tasks can be accomplished using several strategies, and helping people recruit the system that is optimal can help overcome learning deficits. Through our future efforts, we will continue to expand our knowledge of how learning within each system coupled with the ability to recruit the appropriate system contributes to learning.

REFERENCES CITED

- Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., et al. (2000). Complementary category learning systems identified using event-related functional mri. *Journal of Cognitive Neuroscience*, *12*(6), 977-987.
- Assad, W. F., Rainer, G., & Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron*, *21*(6), 1399-1407.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.
- Anderson, J. R. (1983). *The architecture of cognitive skill*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1993). *Rules of mind*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Arsalidou, M., & Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382-2393.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442-481.
- Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*(3), 363-378.
- Ashby, F. G., & Waldron, E. M. (2000). The neuropsychological bases of category learning. *Current Directions in Psychological Science*, *9*(1), 10-14.
- Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Science*, *5*(5), 204-210.
- Ashby, F. G., & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. In L. Jimenez (Ed.), *Attention and implicit learning* (pp. 109-141). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu Rev Psychol*, *56*, 149-178.
- Barcelo, F., Munoz-Cespedes, J. M., Pozo, M. A., & Rubia, F. J. (2000). Attentional set shifting modulates the target p3b response in Wisconsin card sorting task. *Neuropsychologia*, *38*, 1342-1355.

- Baudena, P., Halgren, E., Heit, G., & Clarke, J. M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalography and Clinical Neurophysiology*, *94*, 251-264.
- Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., et al. (2004). Localizing p300 generators in visual target and distractor processing: A combined event-related potential and functional magnetic resonance imaging study. *The Journal of Neuroscience*, *24*, 9353-9360.
- Bledowski, C., Kadosh, K. C., Wibral, M., Rahm, B., Bittner, R. A., Hoechstetter, K., et al. (2006). Mental chronometry of working memory retrieval: A combined functional magnetic resonance imaging and event-related potentials approach. *The Journal of Neuroscience*, *26*, 821-829.
- Bode, S., & Haynes, J. D. (2009). Decoding sequential stages of task preparation in the human brain. *NeuroImage*, *45*(2), 606-613.
- Brankack, J., Seidenbecher, T., & Muller-Gartner, H. W. (1996). Task-relevant late positive component in rats: Is it related to hippocampal theta rhythm? *Hippocampus*, *6*, 475-482.
- Brovelli, A., Laksiri, N., Nazarian, B., Meunier, M., & Bossaoud, D. (2008). Understanding the neural computations of arbitrary visuomotor learning through fmri and associative learning theory. *Cerebral Cortex*, *18*, 1485-1495.
- Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain*, *111*, 23-45.
- Bruner, J., Goodnow, J. J., & Austin, G. A. (1967). *A study of thinking*. New York, NY: Science Editions.
- Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., et al. (2002). Dorsal anterior cingulate cortex: A role in reward-based decision making. *Proceedings of the National Academy of Sciences*, *99*, 523-528.
- Chein, J. M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, *25*(3), 607-623.
- Dale, A. M. (1999). Optimal experimental design for event-related mri. *Human Brain Mapping*, *8*, 109-114.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, *9*, 179-194.

- DeCoteau, W. E., Thorn, C., Gibson, D. J., Courtemanche, R., Mitra, P., Kubota, Y., & Graybiel, A. M. (2007). Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task. *PNAS*, *104*(13), 5644-5649.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, *18*, 193-222.
- Donchin, E., & Coles, M. G. H. (1988). Is the p300 component a manifestation of context updating? *Behavioral and brain Sciences*, *11*, 357-374.
- Ebmeier, K. P., Steele, J. D., MacKenzie, D. M., O'Carroll, R. E., Kydd, R. R., Glabus, M. F., et al. (1995). Cognitive brain potentials and regional cerebral blood flow equivalents during two- and three-sound auditory "oddball tasks". *Electroencephalography and Clinical Neurophysiology*, *95*, 434-443.
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci*, *1*(1), 41-50.
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *127*, 107-140.
- Ferree, T., Eriksen, K. J., & Tucker, D. M. (2000). Region head tissue conductivity estimation for improved eeg analysis. *IEEE Trans Biomed Eng*, *47*, 1584-1592.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*, 341-355.
- Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Oxford, England: Brooks/Cole.
- Frank, M., Loughry, B., & O'Reilly, R. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cogn Affect Behav Neurosci*, *1*, 137-160.
- Friedman, D., Simpson, G. V., & Hamberger, M. (1993). Age-related changes in scalp topography to novel and target stimuli. *Psychophysiology*, *30*, 383-396.
- Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, *443*(7107), 85-88.
- Fregni, F., Simon, D. K., Wu, A., & Pascual-Leone, A. (2005). Non-invasive brain stimulation for Parkinson's disease: A systematic review and meta-analysis of the literature. *Journal of Neurosurgery and Psychiatry*, *76*(12), 1614-1623.

- Gabriel, M., Burhans, L., Talk, A., & Scalf, P. (2002). Cingulate cortex. In V.S. Ramachandran (Ed.), *Encyclopedia of the Human Brain* (pp. 775-791). Amsterdam, Netherlands: Elsevier Science.
- Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review Psychology*, *49*(1), 87–115.
- Grave de Paralta Menendez, R., Murray, M. M., Michel, C. M., Martuzzi, R., & Gonzalez Andino, S. L. (2004). Electrical neuroimaging based on biophysical constraints. *NeuroImage*, *21*, 527-539.
- Groll, M. J., de Lange, F. P., Verstraten, F. A. J., Passingham, R. E., & Toni, I. (2006). Cerebral changes during performance of overlearned arbitrary visuomotor associations. *Journal of Neuroscience*, *26*, 117-125.
- Gruber, A. J., Dayan, P., Gutkin, B. S., & Solla, S. A. (2006). Dopamine modulation in the basal ganglia locks the gate to working memory. *Journal of Computational Neuroscience*, *20*, 153.
- Halgren, E., Baudena, P., Heit, G., Clarke, J. M., Marinkovic, K., & Chauvel, P. (1994). Spatio-temporal stages in face and word processing. 2. Depth-recorded potentials in the human frontal and Rolandic cortices. *J Physiol*, *88*, 51-80.
- Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liegeois, C., Chauvel, P., et al. (1995). Intracerebral potential to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalography and Clinical Neurophysiology*, *94*, 191-220.
- Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Marinkovic, K., Devaux, B., et al. (1995). Intracerebral potential to rare target and distractor auditory and visual stimuli. II. Medial, lateral, and posterior temporal lobe. *Electroencephalography and Clinical Neurophysiology*, *94*, 229-250.
- Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Curr Biol*, *17*(4), 323-328.
- Haynes, J. D. (2015). A primer on pattern-based approaches to fmri: Principles, pitfalls, and perspectives. *Neuron*, *87*, 257-270.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-2430.

- Heindel, W. C., Salmon, D. P., Shults, C. W., Wallcke, P. A., & Butters, N. (1989). Neuropsychological evidence for multiple implicit memory systems: A comparison of Alzheimer's, Huntington's, and Parkinson's disease patients. *Journal of Neuroscience*, *9*(2), 282-287.
- Hintzman, D. L. (1984). Minderva 2: A simulation model of memory. *Behavioral Research Methods, Instruments, & Computers*, *16*(2), 96-101.
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*, 435-448.
- Kahana, M. J., Seelig, D., & Madsen, J. R. (2001). Theta returns. *Current Opinion in Neurobiology*, *11*, 739-744.
- Katayama, J., & Polich, J. (1998). Stimulus context determines p3a and p3b. *Psychophysiology*, *35*, 23-33.
- Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, *110*(2), 316-339.
- Keri, S., Janka, Z., Benedek, G., Aszalós, P., Szatmary, B., Szirtes, G., et al. (2002). Categories, prototypes and memory systems in Alzheimer's disease. *Trends in Cognitive Sciences*, *6*(3), 132-136.
- Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature*, *383*, 256-259.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*(5140), 1747-1749.
- Kramer, A. F., & Strayer, D. L. (1988). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. *Biological Psychology*, *26*, 231-267.
- Kuzis, G., Sabe, L., Tiberti, C., Merello, M., Leiguarda, R., & Starkstein, S. E. (1999). Explicit and implicit learning in patients with Alzheimer disease and Parkinson disease with dementia. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *12*(4), 265-269.
- Laberge, D. (1981). Acquisition of automatic processing in perceptual and associative learning. In J. Long, & A. Baddeley (Eds.). *Attention and performance ix*. Hillsdale, NJ, England: Lawrence Erlbaum Associates.

- Lei, M., & Zhansheng, C. (2003). Rule-based categorization strategy and example-based categorization strategy in categorization. *Acta Psychologica Sinica*, 35(1), 29-36.
- Lewis, S. J., Dove, A., Robbins, T. W., Barker, R. A., & Owen, A. M. (2004). Striatal contributions to working memory: A functional magnetic resonance imaging study in humans. *Eur J Neurosci*, 19(3), 755-760.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.
- Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., et al. (1999). Wisconsin card sorting test performance following head-injury: Dorsolateral fronto-striatal circuit activity predicts perseveration. *Journal of Experimental Neuropsychology*, 21, 2-16.
- Luft, C. D., Takese, E., & Bhattacharya, J. (2014). Processing graded feedback: Electrophysiological correlates of learning from small and large errors. *Journal of Cognitive Neuroscience*, 26, 1180-1193.
- Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science*, 14, 47-53.
- Luu, P., Tucker, D. M., & Stripling, R. (2007). Neural mechanisms for learning action in context. *Brain Research*, 1179, 89-105.
- Luu, P., Shane, M., Pratt, N. L., & Tucker, D. M. (2009). Corticolimbic mechanisms in the control of trial and error learning. *Brain Research*, 1247, 100-113.
- Luu, P., Jiang, Z., Poulsen, C., Mattson, C., Smith, A., & Tucker, D. M. (2011). Learning and the development of contexts for action. *Frontiers in Human Neuroscience*, 5(December), 159.
- Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, 32(4), 582-591.
- Mandelbaum, E. (2013). Numerical architecture. *Topics in Cognitive Science*, 5, 367-386.
- Medin, D. L. (1975). A theory of context in discrimination learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9). New York, NY: Academic Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207-238.

- Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 333-352.
- Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*, 113-138.
- Miles, S. J., Matsuki, K., & Minda, J. P. (2014). Continuous executive function disruption interferes with application of an information integration categorization strategy. *Atten Percept Psychophys*, *76*(5), 1318-1334.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*, *24*, 167-202.
- Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775-799.
- Morgan, K. K., Luu, P., & Tucker, D. M., (2016). Changes in p3b latency and amplitude reflect expertise acquisition in a football visuomotor learning task. *PLoS ONE*, *11*, e0154021.
- Morgan, K. K., Tucker, D. M., & Luu, P. (2018). Understanding the neural mechanisms of memory in rapid recognition of football formations. In R.A. Carlstedt & M. Balconi (Eds.), *Handbook of sport neuroscience and psychophysiology* (Ch. 8). Abingdon, UK: Routledge.
- Naatanen, R. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*, 201-287.
- Nelissen, N., Strokes, M., Nobre, A. C., & Rushworth, M. F. (2013). Frontal and parietal cortical interactions with distributed visual representations during selection attention and action selection. *Journal of Neuroscience*, *33*(42), 16443-16458.
- Newman, J. P., Patterson, M. C., Howland, E. W., & Nichols, S. L. (1990). Passive avoidance in psychopaths: The effects of reward. *Pers Individ Diff*, *11*, 1101-1114.
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi voxel pattern analysis of fmri data. *Trends in Cognitive Neuroscience*, *10*(9), 424-430.

- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104-114.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39-57.
- Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 282-304.
- Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A.F. Healy, & S.M. Kosslyn (Eds.), *Essays in honor of William K. Estes, vol. 1: From learning theory to connectionist theory; Vol. 2: From learning processes to cognitive processes* (pp. 149-167). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1994). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 211-233.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53-79.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266-300.
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*(4), 247-255.
- Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 548-568.
- Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 324-354.
- Pineda, J. A., Westerfield, M., Kronenberg, B. M., & Kubrin, J. (1997). Human and monkey p3-like responses in a mixed modality paradigm: Effects of context and context-dependent noradrenergic influences. *International Journal of Psychophysiology*, *27*, 223-240.
- Poldrack, R. A., Clark, J., Paré-Blagojev, E. J., Shohamy, D., et al. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546-550.

- Polich, J. (2004). Clinical application of the p300 event-related brain potential. *Phys Med Rehabil Clin N Am*, 15, 133-163.
- Polich, J. (2007). Updating p300: An integrative theory of p3a and p3b. *Clinical Neurophysiology*, 118, 2128-2148.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3), 353-363.
- Posner, M. I., & Keele, S. W. (197). Retention of abstract ideas. *Journal of Experimental Psychology*, 83(2), 304-308.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). New York, NY: Cambridge University Press.
- Rao, S. M., Bobholz, J. A., Hammeke, T. A., Tosen, A. C., et al. (1997). Functional mri evidence for subcortical participation in conceptual reasoning skills. *Neuroreport*, 8, 1987-1993.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory*, 5(6), 420-428.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional mri. *Proceedings of national academy of Sciences*, 95(2), 747-750.
- Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkins's disease. *Behavioral Neuroscience*, 113(2), 235-242.
- Reverberi, C., G6rger, K., & Haynes, J. D. (2012). Compositionality of rule representations in human prefrontal cortex. *Cerebral Cortex*, 22(6), 1237-1246.
- Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, 23, 752-263.
- Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of Wisconsin card sorting test in detecting and localizing frontal lobe lesions. *Journal of Consulting & Clinical Psychology*, 48, 605-614.

- Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience*, *12*, 142-162.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192-233.
- Rosch, E. (1978). *Principles of categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schacter, D. L., Cooper, L. A., Tharan, M., & Rubens, A. B. (1991). Preserved priming of novel objects in patients with memory disorders. *Journal of Cognitive Neuroscience*, *3*(2), 117-130.
- Schneider, G. E. (1969). Two visual systems. *Science*, *163*, 895-902.
- Shin, J. (2011). The interrelationship between movement and cognition: Theta rhythm and the p330 event-related potential. *Hippocampus*, *21*, 744-752.
- Smith E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory and Cognition*, *22*(4), 377-386.
- Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*(2-3), 167-196.
- Smith, J. D., & Kemler Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, *113*(1), 137-159.
- Smith, M. E., Halgren, E., Sokolik, M., Baudena, P., Musolino, A., Liegeois-Chauvel, C., et al. (1990). The intracranial topography of the p3 event-related potential elicited during auditory oddball. *Electroencephalography and Clinical Neurophysiology*, *76*, 235-248.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*(2), 127-190.
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*, 387-401.
- Tesche, C., & Karhu, J. (2000). Theta oscillations index human hippocampal activation during a working memory task. *Proceedings of the National Academy of Sciences*, *97*(2), 919-924.

- Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annu Rev Psychol*, *63*, 483-509.
- Toni, I., & Passingham, R. E. (1999). Prefrontal-basal ganglia pathways are involved in the learning of arbitrary visuomotor association: A pet study. *Exp Brain Res*, *127*, 19-32.
- Toni, I., Rammani, N., Josephs, O., Ashburner, J., & Passingham, R. E. (2001). Learning arbitrary visuomotor associations: Temporal dynamics of brain activity. *NeuroImage*, *14*, 1048-1057.
- Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, *87*(3), 154-163.
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301-306.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.
- Wagner, A. D., Gabrieli, J. D. E., & Verfaellie, M. (1997). Dissociations between familiarity processes in explicit recognition and implicit perceptual memory. *Journal of Experimental Psychology: Learning, Memory, and Condition*, *23*(2), 305-323.
- Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence from multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168-176.
- White, I. M., & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res*, *126*(3), 315-335.
- Wise, S. P., & Murray, E. A. Arbitrary associations between antecedents and actions. *Trends in Neuroscience*, *23*, 271-276.
- Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage*, *56*(2), 744-752.
- Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P.M. Matthews, & S.M. Smith (Eds.), *Functional mri: An introduction to methods* (pp. 251-270). New York, NY: Oxford University Press.
- Zeithamova, D., & Maddox, W. T. (2006). Dual task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387-398.

Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: Evidence from brain imaging and behavior. *Journal of Neuroscience*, 28(49), 13194–13201.