2019

# Beta oscillations underlie top-down, feedback control while gamma oscillations reflect bottom-up, feedforward influences

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

# BETA OSCILLATIONS UNDERLIE TOP-DOWN, FEEDBACK CONTROL WHILE GAMMA OSCILLATIONS REFLECT BOTTOM-UP, FEEDFORWARD INFLUENCES

by

**ROMAN F. LOONIS**

B.A.Sc., McGill University, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader _____

Douglas Rosene, Ph.D.
Professor of Anatomy and Neurobiology.

Second Reader _____

Earl K . Miller, Ph.D.
Picower Professor of Neuroscience

*La neurologie cherche à comprendre l'homme lui-même.*
*The problem of neurology is to understand man himself.*

*-- Wilder Penfield 1891-1976*

## DEDICATION

The absurd goal of this work and, I think, of much academic achievement was to help us as humans better understand who we are and, thus, live a more rational and "good" life. This work is dedicated to all those who seek such understanding. However, this search always comes with some personal cost. And these burdens are made light by those we love. Therefore, I would like to dedicate this work to those whom have been my foundation, Charles, Isabelle, Manon, and Anne-Sophie. Thank you to all my friends whose support has been an eternal source of happiness and stability.

## ACKNOWLEDGMENTS

All of the work presented here would have not been possible without the help of all of my fellow lab members. From animal trainings to neural recordings to data analysis, there was not one part that I could have done alone. So, thank you so much to Evan Antzoulatos, Scott Brincat, Jake Donoghue, Simon Kornblith, Morteza Moazami, Mikael Lundqvist, Vicky Puig, Jonas Rose, Jefferson Roy, Matt Silver, and Alik Widge. Brenna Gray your constant help for years will always be remembered. I would like to thank my co-authors, Andre M. Bastos and Andreas Wutz, for our fruitful and stimulating collaboration. Thank you to my BU advisors, Jarrett Rushmore and Douglas Rosene, in the Department of Anatomy and Neurobiology for having been a source of encouragement, and for being helpful guides throughout my graduate years. Most importantly, thank you to Earl Miller, my PI and MIT advisor, for the mentorship and flexibility to pursue those questions which intrigued me the most. You have invaluably taught me how to be a better thinker and clearer writer.

**BETA OSCILLATIONS UNDERLIE TOP-DOWN, FEEDBACK CONTROL WHILE GAMMA OSCILLATIONS REFLECT BOTTOM-UP, FEEDFORWARD INFLUENCES**

**ROMAN F. LOONIS**

Boston University School of Medicine, 2019

Ph.D. degree requirements completed in 2017

Dual M.D./Ph.D. degrees expected in 2017

Major Professor: Douglas Rosene Ph.D., Professor of Anatomy and Neurobiology

ABSTRACT

Prefrontal cortex (PFC) is critical to behavioral flexibility and, hence, the top-down control over bottom-up sensory information. The mechanisms underlying this capacity have been hypothesized to involve the propagation of alpha/beta (8-30 Hz) oscillations via feedback connections to sensory regions. In contrast, gamma (30-160 Hz) oscillations are thought to arise as a function of bottom-up, feedforward stimulation. To test the hypothesis that such oscillatory phenomena embody such functional roles, we assessed the performance of nine monkeys on tasks of learning, categorization, and working memory concurrent with recording of local field potentials (LFPs) from PFC. The first set of tasks consisted of two classes of learning: one, explicit and, another, implicit. Explicit learning is a conscious process that demands top-down control, and in these tasks alpha/beta oscillations tracked learning. In contrast, implicit learning is an unconscious process that is automatic (i.e. bottom up), and in this task alpha/beta

oscillations did not track learning. We next looked at dot-pattern categorization. In this task, category exemplars were generated by jittering the dot locations of a prototype. By chance, some of these exemplars were similar to the prototype (low distortion), and others were not (high distortion). Behaviorally, the monkeys performed well on both distortion levels. However, alpha/beta band oscillations carried more category information at high distortions, while gamma-band category information was greatest on low distortions. Overall, the greater the need for top-down control (i.e. high distortion), the greater the beta, and the lesser the need (i.e. low distortion), the greater the gamma. Finally, laminar electrodes were used to record from animals trained on working memory tasks. Each laminar probe was lowered so that its set of contacts sampled all cortical layers. During these tasks, gamma oscillations peaked in superficial layers, while alpha/beta peaked in deep layers. Moreover, these deep-layer alpha/beta oscillations entrained superficial alpha/beta, and modulated the amplitude of superficial-layer gamma oscillations. These laminar distinctions are consistent with anatomy: feedback neurons originate in deep layers and feedforward neurons in superficial layers. In summary, alpha/beta oscillations reflect top-down control and feedback connectivity, while gamma oscillations reflect bottom-up processes and feedforward connectivity.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

ACC ..... Anterior Cingulate Cortex

ANOVA ..... Analysis of Variance

CM ..... Category-Match

CS ..... Category-Saccade

CSD ..... Current Source Density

DLPFC ..... Dorsolateral Prefrontal Cortex

DMPFC ..... Dorsomedial Prefrontal Cortex

DVA ..... Degree of Visual Angle

EEG ..... Electroencephalogram

ERN ..... Error-Related Negativity

ERP ..... Event-Related Potential

FEF ..... Frontal Eye Fields

GC ..... Grangier Causality

HPC ..... Hippocampus

LFP ..... Local Field Potential

MRI ..... Magnetic Resonance Imaging

MSE ..... Mean Squared Error

MUA ..... Multi-Unit Activity

OM ..... Object-Match

PAC ..... Phase Amplitude Coupling

PEV ..... Percent Explained Variance

PFC .....Prefrontal Cortex

PING .....Pyramidal Interneuron Network Gamma

PLV ..... Phase Locking Value

PMD .....Dorsal Premotor Cortex

PPC .....Pairwise Phase Consistency

SEF..... Supplementary Eye Fields

SEM .....Standard Error of the Mean

SMA .....Supplementary Motor Area

STD .....Standard Deviation

STR .....Striatum

TMS .....Transcranial Magnetic Stimulation

VLPFC .....Ventrolateral Prefrontal Cortex

WM ..... Working memory

# CHAPTER 1: INTRODUCTION AND BACKGROUND

## Neural Oscillations

Neural oscillations are a ubiquitous phenomenon in the brain. They exist in all regions spanning from occipital to frontal lobes, and they vary with cortical layer, development, and cognitive state (learning, memory, attention, and categorization). A number of neuro-psychiatric disorders, including, but not limited to, schizophrenia (Uhlhaas and Singer, 2010), Parkinson's (Schnitzler and Gross, 2005), and Alzheimer's (Jeong, 2004), are characterized by abnormal patterns of neural oscillatory activity. Understanding the functional role and underlying circuitry producing neural oscillations, therefore, is both a question central to the understanding of human cognition and of neuropsychiatric disease. The goal of this thesis is to examine how different oscillations within the nonhuman primate brain are shared across a number of different tasks, serve distinct roles, and reflect the properties of the underlying microcircuit (i.e. the connectivity of different neurons across cortical layers).

Neural oscillations are visible within the local field potentials (LFP) recorded from brain tissue. These continuous voltage recordings, known as LFPs, reflect the mean of the extracellular ionic currents from between 100μm to 6mm away, depending on the reference scheme (Kajikawa and Schroeder, 2011). Oscillatory activity found within these LFPs are often transient, sinusoidal-like rhythms with distinct frequencies that are thought to reflect the synchronization of post-synaptic currents from local populations of neurons (Nunez and Srinivasan, 2006).

Not only are they the product of synchronous neural activity, but neural oscillations exert their own influence on neural spiking (Fries, 2015). Rapid changes in extracellular voltage, as is observed during periods of oscillatory activity, can decrease the spiking threshold of neurons and promote excitable brain states. Oscillatory activity that is shared or simultaneous across different brain regions, therefore, can coordinate excitable brain states and favor spike transmission and, hence, information transfers between them (Fries, 2015). These rhythmic changes also impact whether local neurons engage in long-term potentiation or depression. Both of these processes thought to be central to learning are cellular mechanisms that increase or decrease synaptic efficacy, and, hence, how sensitive any post-synaptic neuron is to pre-synaptic inputs (Huerta and Lisman, 1995). This sensitivity can be reflected both in terms of the magnitude and the speed of post-synaptic potential changes in response to pre-synaptic inputs. Moreover, neural oscillations, given their periodicity, may also serve as an internal clock, organizing the activity of groups of neurons with respect to its own phase. For example, spike timings of place cells within the hippocampal theta rhythm can predict the path an animal is about to take (O'Keefe and Reece, 1993).

Neural oscillations have been observed across a wide frequency range (0.1 – 200 Hz), and different Greek letters are used to subdivide this frequency space into different rhythms. To this day, however, there is considerable debate about the appropriate sub-division of these frequency bands, and their underlying mechanisms are largely unclear.

In general, delta rhythms refer to 1-4 Hz signals, theta 4-7 Hz, alpha 8-12 Hz, beta 13-30 Hz, and gamma 30-140 Hz. After 200 Hz, observed signals are thought to reflect spectral bleed-through from impulse-like spiking and, as such, are often labeled as multi-unit activity. In this introduction, we will restrict our discussion to the alpha, beta, theta, and gamma bands. While these bands are spectrally defined, we will also attempt to classify them according to their presumed functional roles and comment on their mechanism.

## Different Oscillatory Frequencies: Functions and Mechanisms

*Alpha Oscillations* (8-12 Hz) are rhythms which were first recorded in leads overlying occipital cortex when subjects closed their eyes (Berger, 1929), and their presence has been correlated with a number of inhibitory, feedback, top-down processes (Klimesch et al., 2006). In an attentional task, covert attention to one visual hemifield lead to a drop in alpha power in the relevant, contralateral hemisphere, while alpha power increased in the irrelevant, ipsilateral hemisphere (Jensen and Mazaheri, 2010). Using TMS to probe cortical excitability, others have found that in states of high alpha, TMS-induced phosphenes were perceptibly smaller (Thut and Miniussi, 2009). Similarly, in a dual-rule task, when a behaviorally dominant rule needed to be suppressed, there was an increase in alpha band synchrony (Buschman et al., 2012). These alpha oscillations are not only associated with an inhibitory process, but they are also thought to reflect feedback from higher to lower order cortical areas. For example, alpha oscillations were evoked in V1 (lower order region) when V4 (higher order region) was stimulated; this did not occur in V4, when V1 was stimulated (van Kerkoerle et al., 2014). These alpha oscillations are

believed to exert feedback via pulsed inhibition, in which particular phases of alpha inhibit spiking and reduce gamma power (Klimesch et al., 2006; Jensen and Mazaheri, 2010). Mechanistically, alpha is believed to arise from both circuits intrinsic to cortex (either by special pacemaker cells, or recurrently organized pyramidal neurons modulated by GABAergic interneurons), and those connecting thalamus and cortex.

While the alpha oscillations reflect feedback control, so do *beta oscillations* (13-30 Hz). But their role has been more related to the maintenance of a default state (whether it is a current motor state or a cognitive rule). Historically, beta oscillations have been most studied in the context of the motor system and in their relationship to Parkinson's disease. Beta band activity in motor cortex is increased with the successful suppression of a motor response. In the case of Parkinson's disease, excess beta leads to excessive motor suppression and, hence, to Parkinson's hallmark symptom: bradykinesia (Swann et al., 2015). Beta oscillations, however, are not an exclusively motor signal. Increased beta band activity correlates with anticipations (Engel and Fries, 2010), changes in bistable perception (Engel and Fries, 2010; Hipp et al., 2011), and increased top-down control (Buschman and Miller, 2007). Likewise, beta oscillations increase with category learning (i.e. a new rule), organize category-selective spiking across distal brain regions (prefrontal cortex, striatum, and parietal cortex), and encode behaviorally dominant rules (Antzoulatos and Miller, 2011; Antzoulatos and Miller, 2016; Buschman et al., 2012). Beta oscillations, like alpha, are an important feedback signal, and predominate within infragranular layers of visual cortex. These deep cortical layers have been found to be the

source of feedback projections toward lower order cortical regions (Bastos et al, 2015; Michalareas et al., 2016). Modelling studies have subdivided the beta band into a low frequency beta (beta1) and a high frequency beta (beta2). Beta2 is believed to arise from interactions of intrinsically bursting pyramidal cells (due to a muscarinic receptor suppressed M-current), and beta1 is thought to arise in a period of lower cortical excitability through interactions with superficial gamma networks (Kopell et al., 2011). This dissertation will not attempt to dissociate physiologically these different beta bands.

Unfortunately, dissociating alpha and beta oscillations consistently across subjects has been difficult due to the intrinsic variance of oscillatory phenomena, and their largely overlapping putative roles. Centered within a range of 7-14 Hz, peak alpha frequencies can extend up to 16 Hz (given an inter-subject variability of 2.8 Hz, and a mean of 10.4 Hz) (Haegens et al., 2014). This high cut off overlaps with the beta band (13-30 Hz), which itself has a comparable variability (Haegens et al., 2014; Espenhahn et al., 2017). Making the situation more complicated, most electrophysiological studies have found that these oscillatory phenomena are transient and are, hence, non-stationary. Due to the short-lived nature of these oscillatory phenomena, the frequency resolution is limited. For example, when applying the discrete fast Fourier transform, a mathematical method to decompose any time series data into frequency content, on 1kHz sampled signal with oscillatory events lasting between 100ms to 1s, the frequency resolution ranges from 0.5 to 5 Hz (Nyquist Frequency / (Number of Time Bins), i.e. 500 Hz / 100 samples). This is a limited resolution if one is attempting to separate a 12 Hz signal from 13 Hz one. Given

their spectral overlap, their shared sources within infragranular layers, and their common role as feedback signals, alpha and beta oscillations are often lumped together. While this is true in this thesis, these two phenomena are different, for they can be observed simultaneously within a single subject and, in at least one nonhuman primate study, both of these bands have been functionally dissociated (Buschman et al., 2007). However, for the sake of simplicity we will refer to these lower frequency oscillations as alpha/beta for the rest of this thesis.

In contrast to both alpha and beta oscillations, those functional roles associated with *theta oscillations* (3-7 Hz) are somewhat broader: learning, memory, and neuronal coordination. These rhythms have been most studied within the hippocampus (Colgin, 2013), where they organize spiking activity to encode an animal's path through a maze (past and present) (Pfeiffer and Foster, 2013). In brain slices, electrical stimulation at the peaks and troughs of the ongoing theta have differential effects on long term potentiation and long term depression (Huerta et al., 1995; Hyman et al. 2003). Theta has also been shown to facilitate long-distance communication between both cortical and subcortical brains regions, such as between V4-FEF (Liebe et al., 2012), STR-HPC (DeCoteau et al., 2007), FEF-ACC (Babapoor-Farrokhan et al., 2017), PFC-HPC (Benchenane et al. 2010), and PFC-STR-HPC (Herweg et al., 2016). This long-distance coordination between theta oscillations has noticeable impacts on behavior. For instance, when theta oscillations within V4 and FEF synchronized, working memory performance improved (Liebe et al., 2012). Moreover, much like gamma, theta has been demonstrated to reflect feedforward

processes (Bastos et al., 2015). To date, theta within the hippocampus is believed to result from medial septal activation of weakly coupled, hippocampal theta oscillators (Colgin, 2013). The origin of cortical theta oscillations, if different, remains largely unexplored.

And finally, *gamma oscillations* (30 – 140 Hz) represent a class of higher frequency oscillations that are increasingly subdivided into low- (30-50), mid- (50-90), and high- (90-140 Hz) gamma bands. These different sub-bands are known to coexist, yet their different mechanisms and putative roles remain unclear. As a result, for the purposes here, gamma oscillations will be characterized as a whole.

Gamma band oscillations are often local, transient, and correlated with neuronal spiking (Buzsáki et al., 2012). They are more dominant in superficial layers of visual cortex, and (Buffalo et al., 2011; Godlove et al., 2014), they are often coupled in time with lower frequency oscillations (theta, alpha, and beta). For example, in visual cortex, deep alpha or beta oscillations are known to modulate the amplitude of gamma oscillations (Spaak et al., 2012). Physiologically, gamma oscillations arise from feedforward processes, for gamma is elicited in higher cortical regions when lower cortical regions are stimulated and not vice versa (Michalareas et al., 2015; Bastos et al., 2015; van Kerkoerle et al., 2014). In terms of its behavioral correlates, gamma band synchrony in hippocampus correlates with better recognition performance (Jutras et al., 2009) and, in visual regions, increases with attention (Fries et al., 2001). Similarly, gamma oscillations increase with perceptual binding (Gray et al., 1989) and improve working memory performance

(Pesaran et al., 2002; Fries et al., 2001). Moreover, in visual pop-out tasks, where salient visual features drive behavior and, hence, are bottom-up driven, gamma oscillations are increased (Buschman et al., 2007).

Conceptually speaking, gamma oscillations are thought to organize neuronal spiking into functional ensembles. Lisman and others hypothesize that gamma-organized neuronal ensembles represent "letters", while the lower frequency oscillations organize them into "words" (Lisman and Jensen, 2013). Alternatively, Fries and colleagues proposed that when different populations of neurons synchronize within the gamma band, spike transmission is facilitated. This idea is known as the communication through coherence hypothesis (Fries, 2015). The mechanisms underlying gamma oscillations, unlike those of the other oscillations are more worked out. They are thought to arise, like many oscillations, from pyramidal and GABAergic interneuron interactions (PING model: Pyramidal-Interneuron-Network-Gamma). In particular, Parvalbumin neurons, which powerfully inhibit pyramidal neuron activity by targeting their soma, are necessary for these gamma rhythms (Kim et al., 2016; Sohal et al., 2009).

## Top-down control vs. bottom-up influence

Overall, there is increasing evidence that despite differences in their putative functions, behavioral correlates, and exact mechanisms, neural oscillations can be grossly characterized into two classes: those reflecting bottom-up processes (gamma), and those reflecting top-down control (alpha and beta) (Buschman et al., 2007; Buschman et al.,

2012; Antzoulatos and Miller, 2014; Antzoulatos and Miller, 2016). Bottom-up and top-down signals were first introduced in the context of attention and visual perception (Desimone and Duncan, 1995). In this setting, different stimuli within the sensory world were understood to compete for neural representation, and that this competition was mediated by both top-down and bottom-up processes (biased competition model). Bottom-up processes are considered to be largely automatic processes that arise as a function of the intrinsic properties of neurons and neural circuits, and that are not actively dependent on any current set of cognitive demands (Desimone and Duncan, 1995). In contrast, top-down processes bias those neural representations that are relevant for behavior by modulating bottom-up neural responses. In simplest terms, bottom-up processes reflect exogenous inputs and are related to gamma, while top-down control reflects endogenous ones and are related to alpha/beta.

In addition to their correlates with top-down and bottom-up processes, neural oscillations have also been found to correlate with different levels of feedback and feedforward connectivity across brain regions. As mentioned above, alpha/beta oscillations reflect feedback processes, and theta/gamma reflect feedforward processes. However, feedback and feedforward connections are also organized by cortical layer, and this laminar organization is most salient for neurons that connect distant brain regions (Felleman and Essen, 1991). For these neurons, feedback connections arise preferentially from infragranular layers, while feedforward connections arise preferentially from within supragranular layers (Felleman and Essen, 1991; Markov et al., 2014). Again, oscillations

within the LFPs match this anatomical organization of feedback and feedforward neurons. Alpha/beta oscillations are greatest in infragranular regions, while gamma oscillations are strongest in supragranular regions (Buffalo et al., 2011; Xing et al., 2012).

Overall, alpha/beta oscillations reflect top-down control and feedback connectivity, while theta/gamma oscillations reflect bottom-up influences and feedforward connectivity. In this thesis, we sought to further corroborate these roles for gamma and alpha/beta within prefrontal cortex. Prefrontal cortex has long been an important source of top-down control (Miller and Cohen, 2001). For instance, lesions in prefrontal cortex lead to behavioral inflexibility and, hence, a failure to implement top-down control over contextually- and environmentally-driven habits (i.e. bottom-up signals). Anatomically, prefrontal cortex sits high up on the cortical hierarchy, suggesting much of its top-down control must arise from its feedback connectivity on lower cortical regions (Felleman and Essen, 1991). We sought to confirm whether feedback/forward and top-down/bottom-up signals in this cortical region (PFC) varied predictably with task demands and cortical layer. We made the following hypotheses:

- More top-down control should result in an increase in alpha/beta
- More bottom-up influences should result in more gamma
- Alpha/beta should peak in infragranular layers in PFC
- Gamma should peak in supragranular layers in PFC

## Predictions

Much of cognition can be summarized in the following way: learn, remember what you have learned (memory), and generalize what you have learned (categorization). In the following studies, we recorded LFPs from the prefrontal cortex of a cohort of monkeys to probe the oscillatory correlates of top-down/bottom-up and feedback/forward processes spanning the following cognitive tasks: learning, working memory, and categorization. We tested the hypothesis that alpha/beta oscillations reflect top-down, feedback signals, while theta/gamma oscillations reflect bottom-up, feedforward signals.

Learning is thought to reflect at least two different processes, one more bottom-up (implicit), and another more top-down (explicit). Implicit learning is generally conceived as an unconscious learning process that is automatic, non-hippocampal dependent, and tied to metabolic changes in lower order cortices (Cleeremans et al, 1998; Ashby and Maddox, 2005; Reber et al., 2003). In contrast, explicit learning is thought to reflect top-down, hypothesis-driven learning that is conscious, and hippocampal-dependent (Milner et al., 1968; Ashby and Maddox, 2005). If the above hypothesis is correct, we predict an increase in the relative prevalence of alpha/beta band oscillations in explicit compared to implicit learning.

Categorization is the capacity to organize the sensory world into a set of classes (i.e. categories). Those rules governing the membership of objects to these categories can be made at fundamentally different levels of abstraction. At the low end, for example,

different chairs or tables can each be constitutive of their own category; however, at the high end, both chairs and tables can be constitutive of the furniture category. The different levels of abstractness should differentially engage top-down control with more abstract categories having an increase in alpha/beta band oscillations relative to less abstract categories, and less abstract categories having an increase in gamma oscillations relative to more abstract categories.

Working memory is characterized by the maintenance of behaviorally-relevant information during a delay, and has been correlated with persistent modulation of prefrontal neuronal activity. We trained three monkeys on different versions of a working memory task, and we investigated the laminar organization of alpha/beta and gamma oscillations. While previous reports found that there are frequency asymmetries across the layers in visual cortex, we sought to confirm those same asymmetries in those prefrontal regions from which we recorded. We predicted that within those frontal regions sampled alpha/beta oscillations would be more prevalent in deep layers, and gamma oscillations superficially.

# CHAPTER 2: A META-ANALYSIS SUGGESTS DIFFERENT MECHANISMS UNDERLIE EXPLICIT AND IMPLICIT LEARNING

Roman F. Loonis[1,2], Scott L. Brincat[1], Evan G. Antzoulatos[1,3], and Earl K. Miller[1*]

[1]The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

[2]Department of Anatomy and Neurobiology, Boston University

[3]Center for Neuroscience, Department of Neurobiology, Physiology and Behavior, University of California Davis

*To whom correspondence should be addressed

## Abstract

Learning can be explicit (hippocampus-dependent and conscious) or implicit (hippocampus-independent and unconscious). A meta-analysis of three pairs of non-human primates performing three different learning tasks (Object Match, Category Match, and Category-Saccade associations) revealed signatures of explicit and implicit learning. Errors were used to improve performance in the explicit (Match) tasks, but not the implicit (Saccade) task. Error-related negativity, an evoked potential indicating processing of negative feedback, was greater in both Match tasks. The Match (explicit) tasks vs Saccade (implicit) task also showed different patterns of local field potential synchrony within the prefrontal cortex (PFC) and other brain areas (hippocampus,

dorsomedial PFC, and striatum) following correct vs incorrect choices. All tasks showed an increase in alpha/beta (10-30 Hz) synchrony after correct choices. However, while delta/theta (3-7 Hz) synchrony increased after correct choices in the Saccade (implicit) task, it did so only after incorrect choices in the Match (explicit) tasks. Alpha/beta synchrony increased with, and decreased after, learning in the explicit (Match) tasks, but not the Saccade (implicit) task. Instead the Saccade (implicit) task showed a decrease in theta synchrony with learning. In sum, our results suggest that explicit vs implicit learning not only engage different brain systems, but they may also engage different neural mechanisms that rely on different patterns of oscillatory synchrony.

## Introduction

Learning was once believed to be a unitary process. As it turned out, however, patient HM and other amnesia patients have preserved skill learning despite an inability to retain and recall new facts and episodes (Scoville and Milner, 1957; Milner, 1962; Milner et al. 1968; Cohen and Squire, 1980). This led to the notion that there are at least two major forms of learning: one, hippocampal-dependent and episodic in content (explicit learning), and another, non-hippocampal and largely unconscious (implicit learning).

While it is clear that explicit and implicit learning engage distinct brain systems, differences in their neural mechanisms have been less clear. For the most part, studies of the neural correlates of both types of learning report similar findings. On the neuron level, tuning sharpens, signal-to-noise ratio improves and their activity becomes a better

predictor of task events (Antzoulatos and Miller, 2011; Asaad and Miller, 1998; Brincat and Miller, 2015; Chen and Wise, 1995; Sakai and Miyashita, 1991; Pasupathy and Miller, 2005; Williams and Eskandar, 2006; Wirth et al., 2003; Wirth et al., 2009). On the network level, learning enhances oscillatory activity, improves synchrony between neurons, and even sculpts unique oscillatory ensembles (Antzoulatos and Miller, 2014; Brincat and Miller, 2015; Buschman et al., 2012; Hargreaves et al., 2012; Jutras et al., 2009; Jutras et al. 2013). Animal studies are generally agnostic as to whether this plasticity is related to explicit or implicit learning. Assignment to one or the other is typically made by whether the brain area in question has been associated with explicit learning (e.g., the hippocampus) or implicit learning (e.g., the basal ganglia) and whether learning is fast (explicit) or slow (implicit). There is no clear neural signature differentiating the two.

This is due, in part, to practical considerations. A typical experiment trains animals to learn one task. That is difficult enough. Training animals to learn two or more tasks is prohibitively time-consuming. It occurred to us, however, that we had data from three experiments that differed in their formal demands in two ways: one, in the content of what was learned (paired associations between objects vs category membership), and, two, in how that learning was "read out" (via a match decision or visuomotor association). Fortuitously, there was enough overlap in the tasks for us to isolate these different factors. We found different patterns of post-choice synchrony that varied with the read-out, not with the content. Examination of the animals' behavior and neural

activity supported the conclusion that these different synchrony patterns were signatures of explicit and implicit learning.

## Results

### *Tasks*

Six monkeys (three different pairs) performed three different learning tasks (Fig. 2.1A-C). During each session of the Object Match task (OM; Fig. 2.1A), animals learned through trial-and-error four novel associations between pairs of objects (see Methods). They saw two objects in succession: first a sample and then a test. If the test object was the pre-assigned paired associate of the sample, they were rewarded, with juice, for making a saccade to a subsequent, randomly positioned target. If they responded to the wrong test object, they received negative feedback, i.e. there was no reward and a red screen flashed on.

The other two tasks required animals to categorize dot patterns that were distortions of prototype patterns (Fig. 2.1B,C). These prototypes were jittered according to a set of statistical rules to produce a large number of exemplars for each category. For each recording session, two novel categories were generated, and the animals had to learn through trial-and-error which exemplars belonged to which categories. The animals were first presented with a sample exemplar from one of the categories, which was then followed by a short delay. In the Category Match task (CM), two test exemplars

appeared side-by-side after this delay – one on the right side of the screen, and the other on the left (Fig. 2.1B). One of the exemplars was from the same category as the sample (a category match); the other was from the other category. The left vs right location of the matching exemplar was random. The monkeys free-viewed the test exemplars and were rewarded for maintaining fixation on the correct one. If incorrect, the animals received negative feedback: there was no reward and the chosen stimulus turned red. In the Category Saccade task (CS), at the end of the delay, two green dots appeared on the right and left side of the screen. Each of the categories was arbitrarily associated with a saccade to the right or left dot. (Fig. 2.1C). The monkeys learned by trial-and-error which saccade was associated with which category. As before, an incorrect response was followed by negative feedback (no reward and a presentation of the sample exemplar at the correct location).

In order to facilitate learning in the category tasks, each session was organized into a set of blocks. In the first block, the animals were presented with only two exemplars from each category. To move on from one block to the next, the animals had to perform at or above 70% correct. With every subsequent block, a greater number of novel exemplars from each category was used. In each block, there were a total of $2^{block}$ exemplars (Antzoulatos and Miller, 2011). Thus, each time they reached criterion, the animals were challenged with more novel exemplars, facilitating their gradual acquisition of the categories.

For all three tasks, the animals were well trained on the formal demands of the tasks, but had to learn new stimuli for each recording session. In each task, animals started near, or at, chance and gradually reached a good level of performance (> 75% correct) within a single recording session (average 2-3 hours). Mean performance during Category-Saccade and Category-Match learning was no different (CM, 79.3%; CS, 82.1%; $p = 0.1637$), while performance during Object-Match learning was somewhat lower (69.1%, vs. CS $p = 1 \times 10^{-7}$; vs. CM, $p = 2 \times 10^{-24}$, two-sided t-test). Recordings were obtained from sites distributed evenly across dlPFC and vlPFC in the banks and gyri of the caudal third of the principal sulcus (Fig. 2.1D-E). Additional recordings were obtained in the Object Match task from the hippocampus (HPC), in the Category Saccade task from the anterior caudate (STR, striatum), and in the Category Match task from dorsomedial Prefrontal Cortex in the vicinity of the supplementary eye fields (dmPFC).

For the sake of analysis, we divided learning into stages. In the Object-Match task, the animals gradually acquired the paired associations. Thus, we evenly divided the session into Early (first third of trials), Middle (middle third) and Late (final third) learning stages. Because of the blocked structure of the category learning tasks, defining the learning stages was less straightforward. In order to do so, we focused on the acquisition of category information to determine learning stages, as we had in our prior work (Antzoulatos and Miller, 2015). Category knowledge was assessed by performance to the novel exemplars. Early in learning, the monkeys had not yet acquired any category information. When novel exemplars were introduced at the start of a block of trials,

performance to novel exemplars was substantially diminished, if not at chance (i.e., they guessed). By contrast, late in learning, they had acquired the categories and performance to novel exemplars was at a high level (75% correct) and stable (i.e. the animal's performance to novel stimuli reached an asymptote). In order to characterize this asymptote, we tested sequentially whether or not novel performance in the first *n* trials differed from novel performance on trials *n+1* to the end of the day. When the null hypothesis could no longer be rejected, we considered those trials as late learning. To reject this null hypothesis, we used a two-sided t-test and set a threshold of $p < .05$. In the Category-Saccade task, this plateau of novel exemplar performance occurred in block 5, and, in the Category-Match task, it occurred in block 2.

As a result, in the Category-Saccade task, we identified an Early stage of learning (prior to acquisition of category information, i.e., chance performance to novel exemplars) as the first 80 trials (the first two blocks of trials, on average). The Late learning stage occurred in blocks 5 to 8 (on average, trials 210-550), when performance to novel exemplars was high and stable (and thus the categories were learned). We could also identify a Middle stage (around trials 81-209 and blocks 3-4, on average) in which performance to novel exemplars did not drop to chance, but was below criterion and improved over the block of trials (and thus categories were being acquired). Finally, in the Category-Match task, learning occurred more rapidly. Within the first 100 trials, and by block 2, the monkeys' average performance to novel exemplars reached an asymptote. There was no clear middle stage. Therefore, in the Category-Match task, the first 100

trials were classified as Early learning, and the following 100 trials were classified as Late.

The idea in comparing these three tasks was that they overlapped in different ways. Two of them shared the requirement to make a match decision (but to different types of stimuli - a paired associate object vs dot category exemplars). The other two shared similar stimuli (dot category exemplars) but differed in response (match decision vs performing an associated visuomotor response).

*Positive feedback was emphasized during the visuomotor learning task*

An examination of the animals' behavior suggested that match tasks relied on explicit learning while the visuomotor (saccade) task relied on implicit learning. Prior studies have shown that implicit learning relies more on positive than negative feedback, while explicit learning utilizes both comparably. For example, skill learning in amnesia patients is better when positive feedback is emphasized (Evans et al., 2000; Squires et al., 1997; Roberts at al., 2016; Maxwell et al., 2001; Poolton et al., 2005). The different use of feedback information for learning seemed to divide our match and saccade tasks.

We found that Category-Saccade learning improved more after correct choices (and positive feedback) than after incorrect choices (and negative feedback). In fact, negative feedback in this task appeared to be disruptive; performance worsened immediately after an incorrect trial and reaction times increased. Figure 2.1A shows performance on

Category-Saccade trials that were immediately preceded by either a correct response (and positive feedback) or an incorrect response (and negative feedback). Performance on the Category-Saccade task was significantly better if the preceding trial was correct (blue bars) than if the previous trial was incorrect (red bars) (+23.97% after correct choices, $p < 1 \times 10^{-4}$, for all stages, bootstrap). This performance advantage after correct trials held throughout learning, even as overall performance improved. Further, reaction times on those trials following an incorrect trial increased by 27.9ms (Fig. 2.3, $p < 1 \times 10^{-4}$). This increase in reaction times was not driven exclusively by a rise in the number of errors, for even correctly performed trials after an error had much higher reaction times (23.9ms, $p < 1 \times 10^{-4}$).

By contrast, during the two Match tasks (OM and CM), there was barely a difference in performance (+1.86% in Object-Match task, $p < .005$, Fig. 2.2B; +3.75%, in Category-Match, $p < 0.025$, bootstrap, Fig. 2.2C) and in reaction times (Fig. 2.3: OM: 5.1ms, $p < 1 \times 10^{-4}$; CM: 8.8ms, $p < 1 \times 10^{-4}$) on trials following correct trials vs incorrect trials. Only early in learning did both Match tasks differ marginally in their performance improvement after correct trials (Early, +2.86% CM > OM, $p = 0.0524$; Late, +2.32%, CM > OM, $p = .1048$). Moreover, this difference in performance was minor relative to those differences between both Match tasks and the Saccade task (CS vs. OM, +29.24% (early), +21.45% (late), $p < 1 \times 10^{-4}$; CS vs CM, +26.31% (early),+19.13% (late), $p < 1 \times 10^{-4}$). In contrast, the differences in reaction times following an error were not statistically different between Match tasks ($p = 0.171566$), and were significantly smaller

than in the Saccade task (vs. OM. -22.8ms, vs. CM**,** -17.7ms, $p < 1 \times 10^{-4}$). We will see next that these task differences were also mirrored in differences in an evoked potential called the error-related negativity.

*Error-related negativity was stronger for match than saccade learning*

Error-related negativity (ERN) is an event-related potential observed after committing errors during learning. It has been correlated with error awareness and the use of errors to improve learning (Frank et al., 2005; Gehring and Willoughby, 2002; Scheffers and Coles, 2000; Walsh and Anderson, 2012; Wessel et al., 2011; Wessel, 2012). The behavioral analysis described above suggested that Category-Saccade learning was less reliant on errors than the two match tasks. This was paralleled in a weaker error-related negativity (ERN) in the Category-Saccade task relative to the match tasks.

Figure 2.4 shows the event-related potentials (ERPs) following positive (blue) or negative (red) feedback, averaged across all electrodes, for the Object-to-Match (left column), Category-to-Match (middle column), and Category-to-Saccade (right column) tasks. In humans, the ERN typically peaks between 80-300ms after an error. This time period is shaded grey in Figure 2.4A-F. As predicted, during Category-Saccade learning, there was no prominent ERN in the expected time window in either the PFC (top right) or STR (bottom right) (Fig. 2.4C,F). In contrast, in that same time window, during both Object-Match and Category-Match learning, there was a prominent negative potential, an ERN, following errors (Fig. 2.4A,D). This sharp negative potential on error trials (red line) was

most clear when compared to its absence following a correct response (blue line). The ERNs in both Match tasks were seen in PFC, dmPFC, and HPC (Fig. 2.4A,B,D,E). In fact, the ERN in these Match tasks correlated with behavioral performance. This was not the case in the Saccade task.

In order to account for any differences associated with different neural latencies across the tasks, we recomputed the ERN by z-scoring the raw voltage differences to a within trial mean and STD, and aligning each of the tasks to their maximal difference. We quantified these differences over a 50-ms window centered on the peak negativity for each of the 3 tasks (Fig. 2.4H). There was no significant difference between the match tasks ($p = 0.911$, bootstrap). By contrast, the ERN in the match tasks was significantly greater than in the saccade task ($p < .001$). This supported the conclusion that errors were of greater use in the match tasks and, thus, that they depended on explicit learning. The lack of ERN during the saccade task supports its reliance on implicit learning. Next, we'll show that feedback-period patterns of oscillatory synchrony also differed between the match vs saccade tasks.

*Neural synchrony differentiates learning styles*

In a previous report using the Object-Match task, Brincat and Miller (2015) found differences in LFP-LFP synchrony between and within the PFC and HPC during the feedback period. After correct responses, there was long-latency, long-duration synchrony, mainly at 10-30 Hz (the alpha-2/beta band). By contrast, after incorrect

responses, there was short-latency, short-duration synchrony at 3-7 Hz (i.e., delta/theta band). This was interpreted as reflecting the network interactions that guide learning by signaling success or failure. The differences in how animals responded to success and failure between the Match tasks vs Category-Saccade task in this report raised the question of whether feedback-related network interactions also differed: they did.

Figure 5 plots differences in synchrony (PPC) during the feedback period between correct and incorrect trials. Because both the Category-Match and Category-Saccade tasks involved eye movement responses, we removed the potentials related to the saccade away from the response target. Both of the Match tasks showed the same pattern: an increase in alpha-2/beta synchrony after a correct response and an increase in delta/theta synchrony after an incorrect response. During the Object-Match task, there was an increase in alpha-2/beta synchrony on correct trials (red colors) both within the PFC (Fig. 2.5A) and between the PFC and HPC (Fig. 2.5B). On incorrect trials, there was an increase in delta/theta synchrony (blue colors), especially within the PFC (Fig. 2.5A). Note that the increase in alpha-2/beta on correct trials tended to be long in duration and peaked after a long latency (750-1000ms), whereas the increase in theta/delta synchrony on incorrect trials peaked earlier (400-500ms). This is consistent with our prior report (Brincat and Miller, 2015). Similar results were obtained from the Category-Match task (Fig. 2.5C, D), especially within the PFC (Fig. 2.5C).

By contrast, the Category Saccade task produced a different pattern of results. Like the

Match tasks, there was a modest increase in alpha-2/beta after correct responses (light red colors). However, correct responses in the Category-Saccade produced a large increase in theta/delta (darker red colors), unlike the Match tasks in which delta/theta synchrony only increased after *incorrect* responses. This was true both within the PFC (Fig. 2.5E) and especially between the PFC and STR (Fig. 2.5F).

We quantified these differences by averaging over the 1.25 seconds after feedback was delivered (time zero in Fig. 2.5A-F). The results are shown in Fig. 2.5G. Positive values indicate greater synchrony after a correct response; negative values indicate greater synchrony after an incorrect response. In the alpha-2/beta band, all three tasks showed a significant increase in synchrony after correct trials only, within the PFC (OM, +0.012; CM, +0.0186; CS, +0.0358; $p < 2 \times 10^{-4}$, bootstrap) and, albeit weaker, between PFC and other areas (OM, +0.0012, $p = 0.002$; CM, +0.0112, $p < 2 \times 10^{-4}$; CS, +0.0268, $p < 2 \times$ 10-4). In the delta/theta band, there was a marked difference between the Match tasks and the Category-Saccade task. The Match tasks showed a significant increase in delta/theta synchrony after incorrect responses, especially within the PFC (OM, -0.0517; CM, -0.0129, $p < 2 \times 10^{-4}$) but also between PFC and HPC, and PFC and dmPFC ("PFC-Other", OM, -0.018, $p < 2 \times 10^{-4}$; CM, -0.0054, $p = 0.0458$). By contrast, in the Category-Saccade task there was an increase in delta/theta synchrony after correct responses within the PFC (CS, +0.067, $p < 2 \times 10^{-4}$) and, more prominently, between PFC and STR (CS, +0.1057, $p < 2 \times 10^{-4}$).

To compare the relative differences in synchrony between tasks and frequency bands, we used correct trials to calculate a ratio of the alpha-2/beta synchrony relative to delta/theta synchrony. A value greater than 1 indicates that synchrony in the alpha-2/beta band was greater than that in the delta/theta band. Values less than one indicate stronger delta/theta synchrony relative to alpha-2/beta. This is plotted in Figure 2.6H. Both Match tasks had beta-theta ratios significantly above 1 (OM, ratio = 1.0641, $p = 0.0034$; CM, ration = 1.2425, $p < 2 \times 10^{-4}$, bootstrap). This shows that alpha-2/beta dominated over delta/theta synchrony on correct trials during the Match tasks. By contrast, in the Category-Saccade task, the ratio was significantly less than 1 (ratio= 0.2759, $p < 2 \times 10^{-4}$), indicating the dominance of delta/theta over alpha-2/beta on correct trials. Taking the absolute values of the ratio differences from 1 showed that the increase in delta/theta synchrony following correct Category-Saccade responses was greater than the increase in alpha-2/beta synchrony following correct trials in the Object-Match (+0.66, $p < 2 \times 10^{-4}$) or Category-Match (+0.4816, $p < 2 \times 10^{-4}$) tasks. These results suggest explicit and implicit learning may be differentiated by distinct patterns of network synchrony in response to feedback.

*Neural synchrony changes with learning and learning style*

We next examined whether synchrony during the feedback period changed with learning and, if it did, whether it did so differently for the putative explicit vs implicit learning tasks. To investigate these changes, we focused on correct trials, because, for one, in the Category-Match task, animals learned rapidly and we only had a small number of

incorrect trials. Two, longer duration synchrony (~1-2s) tended to be seen after correct trials in all three tasks (albeit at different frequencies), and these longer duration events helped reduce the noise in our synchrony estimates. And three, whether the task was implicit or explicit, performance levels following correct trials was comparable (Fig. 2.2). To compute a frequency profile of the changes with learning, we computed PPC values using the traditional frequency bands: theta (3-7 Hz), alpha-1 (8-9 Hz), alpha-2 (10-12 Hz), beta-1 (13-17 Hz), and beta-2 (18-30 Hz). We used the entire 1.7 second interval following delivery of feedback and compared the average values from early vs late in learning (as defined previously).

Figure 2.6 shows the change in PPC from early to late in learning. Fig. 2.6A shows these changes within PFC alone and Fig. 2.6B shows them between the PFC and other areas (PFC-dmPFC, PFC-STR, and PFC-HPC). Positive values indicate an increase with learning, and negative values, a decrease. The largest effect was seen during Category-Saccade learning. There was a decrease in theta synchrony within PFC (Fig. 2.6A) and between PFC and STR (Fig. 2.6B) (PFC: -0.0472; PFC-STR: -0.0472, $p < .005$, bootstrap). By contrast, there was little or no change in theta synchrony with learning during either Object-Match or Category-Match tasks. Instead, Match learning showed a moderate, but significant, increase in synchrony in the higher frequencies, especially in the alpha-2/beta-1 band, within PFC (OM: alpha-2, +0.0097; beta-1, +0.0126, $p<.005$; CM: alpha-2, +0.0113, $p <.005$). There was also a modest, but significant, drop in beta-2 band synchrony across brain regions during the Match tasks, but not the Saccade task

(OM: PFC-HPC, beta-2, -0.0023, p<.005; CM: PFC-dmPFC, beta-2, -0.0046, p <.005). There were other increases and decreases in alpha-1, and beta-2 with learning in the Match tasks, but they were modest and not consistent across tasks (OM: PFC-HPC, alpha-1, +0.0012, p<.005; beta-2, -0.0023, p<.005; CM: PFC, beta-1, -0.0013, p<.005). Moreover, there were no shared changes in synchrony in any frequency band between any of the Match tasks and the Saccade task (CS: PFC, theta, -0.0472; alpha-1, -0.0213; alpha-2, -0.0172; beta-1; -0.009, p < .005; PFC-STR, theta -0.0513, alpha-1, -0.0183, p <.005; beta-1, +0.0026).

*During match learning, alpha-2/beta-1 synchrony increased then decreased*

Above, we showed a moderate increase in alpha-2/beta-1 synchrony with learning during the Match tasks. A closer examination revealed something more complex. Alpha-2/beta-1 first increased with learning, but then late in learning, after the animals reached criterion, it decreased. Because the animals' learning rate varied from task to task, from session to session, and even within a session itself, we could not simply relate PPC to an average learning curve. To better assess how alpha-2/beta-1 synchrony changed with performance, we computed PPC on correct trials over bins of 20 non-overlapping trials. We plotted average PPC values as a function of the animals' level of behavioral performance over the trials bracketed by this same 20 correct-trial window. To maximize statistical power, we averaged across all electrodes both within and outside the PFC for each session.

Figure 2.7 shows the outcome of this analysis. Figure 2.7A,B shows the results for the Match tasks in the 10-17 Hz (alpha-2/beta-1) band where we observed an increase with learning (Fig. 2.6). This revealed that 10-17 Hz synchrony increased as task performance improved. That is, until the animals had largely learned the tasks. For both the Object-Match (Fig. 2.7A) and Category-Match (Fig. 2.7B) tasks, alpha-2/beta-1 synchrony increased until the animals reached around 80% correct performance and after that, synchrony decreased. This drop-off largely erased prior learning-related increases. A piecewise linear model, made up of two linear models, LM1 and LM2, confirmed these changes with performance (Fig. 2.7A,B). Linear Model 1 estimated the changes in synchrony with performance increases from 50 to 80%, and Linear Model 2 estimated the changes in synchrony with performance increases from 80 to 100%. The coefficients for synchrony changes with performance were significant and opposite in sign around the 80% performance mark (OM: LM1, $\beta = 0.028449$, $p = 0.0068933$; LM2, $\beta = -0.079505$, $p = 0.0013742$; CM: LM1, $\beta = +0.11223$, $p = 4.7148e\text{-}9$; LM2, $\beta = -0.09765$, $p = 0.0087861$, two-sided t-test). We applied this same analysis in the Category-Saccade task both to PFC pairs alone and PFC-STR pairs included (where we saw a decrease with learning, see Fig. 2.6A), and yet we still did not find any increase with performance (PFC alone: LM1, $\beta = -0.04015$, $p = 0.61026$; PFC-STR included: LM1, $\beta = -0.13447$, $p = 0.067499$). Instead, the observed drops in alpha-2/beta synchrony were restricted to after the learning criterion was reached (PFC alone: LM2, $\beta = -0.20717$, $p = 0.0054737$; PFC-STR: LM2, LM2, $\beta = -0.039018$, $p = 0.55914$) (Fig. 2.7C).

*During saccade learning, theta synchrony drops continuously*

In the Category-Saccade task, we found that theta synchrony dropped with learning (Fig. 2.8A). Because alpha-2/beta changed differently before and after learning, we sought to identify whether theta synchrony changed at the same rate before and after performance levels reached 80%. This was the case. In Figure 2.8A, we found that theta synchrony dropped both early (LM1, $\beta$ = -0.28891, p = 0.027735, two-sided t-test) and late (LM1, $\beta$ = -0.27005, p = 0.015615). In other words, we did not see any difference in the change in theta synchrony before and after learning, like we did for alpha-2/beta.

Eye movements tend to be made in the theta range (3-7 Hz), and a concern was that these movements (and their correlates) could contribute to the observed changes in theta synchrony. However, we found little evidence that this was the case. We first sought to control for any timing differences in saccades made during the feedback period, both on correct and incorrect trials. To do so, we aligned all of the data on a trial-by-trial basis to the first saccade the animal made away from the target that was chosen. After this re-alignment, we recomputed the PPC and we found that, despite this realignment, theta synchrony remained significantly higher on correct trials (-200 to 0ms prior to the saccade away, +0.1012, $p < 2 \times 10^{-4}$, bootstrap, Fig. 2.8B). In fact, it appeared that the rise in the theta synchrony preceded this eye movement away from the target and was more closely time locked to the delivery of the feedback. Specifically, we found that theta synchrony on correct trials peaked 81-119ms before the saccade and approximately 231-269ms after the feedback (95%, CI). Alternatively, using the eye tracking data, we

assessed both saccade velocity and the average number of saccades made over the feedback period. We found that, while there was an increase in the number of saccades on incorrect trials (+4 saccades, $p < 2 \times 10^{-4}$), there were neither any changes with learning in the number of detectable saccades nor in the average saccade velocity over the entire feedback period analyzed (0-1.7s) (Fig. 2.8C,D). Finally, it was possible that the theta synchrony present in the Category-Saccade task may be tied to eye movements that do not occur in the match tasks. Again, this was not the case. We found that over the 1.7s feedback period there was a median number of 9 saccades in the Category-Match task, strikingly similar to the median number of 8 saccades found in the Category-Saccade task (Fig. 2.8E).

## Discussion

We found evidence that two tasks involving match decisions engaged explicit learning, whereas a task involving a visuomotor (saccade) association engaged implicit learning. Further, we demonstrated that these putative explicit and implicit learning tasks had different patterns of neural synchrony following correct vs incorrect behavioral choices. During the explicit (Match) tasks, there was an increase in alpha-2/beta synchrony following a correct choice and an increase in delta/theta synchrony following an incorrect choice. The implicit (Saccade) task showed a different pattern. Like the explicit (Match) tasks, alpha-2/beta synchrony increased after correct choices. But unlike the Match tasks, which showed increases in delta/theta synchrony after *incorrect* choices, the implicit (Saccade) task showed increased delta/theta synchrony after *correct* choices. The two

types of tasks also showed differences in how synchrony changed with learning. Alpha-2/beta-1 synchrony increased during explicit learning (both Match tasks) until the animals reached a high level of performance, then it dropped off. By contrast, during the implicit (Saccade) task, alpha-2/beta-1 did not increase with learning and theta synchrony decreased.

The evidence that the tasks engaged different learning systems came from how errors were treated. During the implicit (Saccade) task, performance was better immediately following a correct than following an incorrect trial. By contrast, during the explicit (Match) tasks, performance was equally good following correct and incorrect trials. This is consistent with observations that amnesia patients (who rely on implicit learning) acquire new skills more rapidly, and retain them longer, with errorless learning (Squires et al., 1997; Evans et al., 2000; Maxwell et al., 2001; Poolton et al., 2005; Roberts et al. 2016). By contrast, explicit learning utilizes feedback about both correct and incorrect responses to improve behavior. The greater error-related negativity (ERN) in the explicit (Match) tasks than in the implicit (Saccade) task further supports this conclusion. In humans, the ERN is correlated with error awareness, conflict monitoring, and the use of errors to improve learning, all hallmarks of explicit learning (Frank et al., 2005; Scheffers and Coles, 2000; Walsh and Anderson, 2012; Wessel et al. 2011; Wessel et al. 2012). For example, explicit learners of a sensorimotor sequence task exhibited an enhanced ERN relative to implicit learners **(**Russeler et al. 2003). Our results are also consistent

with other reports that use of positive vs negative feedback differentiates implicit vs. explicit learning tasks (Morrison et al., 2015; Smith et al. 2013).

Category learning, and more specifically the dot-category learning employed here, has been found to rely on either implicit or explicit learning systems, depending on the task structure and instructions (Ashby and O'Brien, 2005; Ashby and Maddox, 2011; Carpenter et al., 2016; Reber et al. 1998; Milton et al., 2011; Seger and Miller, 2010). If the dot learning was accompanied by motor instructions (such as point to center of dot pattern), or the task was an A-/not-A category distinction, implicit memory was used (Squire and Knowlton, 1995; Zeithmova et al., 2008). If instead, participants were told that there were different patterns, explicit memory was used (Aizenstein et al.,2000; Reber et al.,2003). Likewise, two of our tasks (Category-Match and Category-Saccade) required categorization of dot patterns, but had different behavioral requirements that seemed to engage explicit vs implicit learning, the latter based on a motor decision. Our results are also in line with other observations that working memory tasks that seem formally equivalent (e.g., "remember objects") can have different neural correlates depending on whether those memories are reported by actively choosing a match from alternatives or recognizing their match (Warden and Miller, 2011).

During explicit learning, alpha-2/beta band synchrony increased on correct choices, increased over the course of learning, and decreased after learning. Alpha-2/beta band synchrony has been tied to cognitive functions such as attention, top-down control, and

feedback processing. This seems consistent with its role in explicit memory formation. Moreover, in a previous report, this feedback-period, alpha-2/beta band synchrony had been found to first arise from the hippocampus (Brincat and Miller, 2015). Together, all of this information suggests that the alpha-2/beta band synchrony found in the feedback period may reflect the activity of specialized neural circuits originating from the hippocampus responsible for explicit learning.

Theta synchrony, on the other hand, has been linked with learning, memory, and conflict monitoring (Colgin, 2013). Theta oscillations and theta stimulation have been known to facilitate both long term potentiation (LTP) and long term depression (LTP) in brain slices (Huerta et al., 1995; Hyman et al. 2003). Theta synchrony has never been reported in non-human primates in response to positive feedback, nor has its decrease with learning. Our observations of theta synchrony within prefrontal cortex and between prefrontal cortex and striatum, in addition to the hippocampus, suggest that theta synchrony is a widespread plasticity signal. Implicit learning, therefore, may depend on global changes in LTD and LTP, rather than the activation of specific hippocampal-based networks. Alternatively, theta oscillations may act as a mechanism organizing neural activity between brain areas. Previous studies have suggested that low frequency synchronizations facilitate long-distance communication. For instance, theta synchrony has been reported to coordinate activity between regions, such as V4-FEF (Liebe et al., 2012), STR-HPC (DeCouteau et al., 2007), FEF-ACC (Babapoor-Farrokhan et al., 2017), PFC-HPC (Benchenane et al. 2010), PFC-STR-HPC (Herweg et al., 2016), and LIP-

TEO-V4-Pulvinar (Wang et al., 2012). In particular, one study found that as animals learned a procedural task, theta oscillations within STR and HPC became anti-phasic (DeCouteau et al., 2007). The presence of theta synchrony, hence, between PFC and STR may facilitate the functional connectivity between PFC and STR over that of PFC and HPC.

In sum, our results suggest that explicit vs implicit learning not only engages different brain systems, it may also engage different neural mechanisms that rely on different patterns of oscillatory synchrony.

## Methods

### *Animals*

All experiments were performed in adult (~8–10 years old) rhesus macaques (*Macaca mulatta*), ranging from 5 to 13 kg. All procedures followed the guidelines of the MIT Animal Care and Use Committee and the US National Institutes of Health. In total, 4 females and 2 males were trained in this study. In the Category-Saccade task, one of the animals had been previously trained on a conditional association task. In the Category-Match task, one of the animals was being actively treated with cyclosporine daily. All animals spent approximately 1-2 years of training on their respective tasks.

*Tasks*

**Object-Match Task:** The details of this task have been presented previously (Brincat and Miller, 2015). In each session, six novel objects were chosen from an image database (Hemera Photo-Objects). Four were randomly designated as cue objects and the remaining two as associate objects. In turn, each cue object was randomly paired with an associate object. The monkeys' task throughout each session was to learn, through trial-and-error, which associate was paired with each cue. To initiate a trial, each monkey fixated on a central white dot for 0.5s. After this fixation period, a cue object (foveal, 3° DVA wide) was presented for 0.5s, followed by a blank delay of 0.75s. Two associate objects were then presented in a randomly-ordered series. Each object presentation lasted 0.5s, and was then followed by another a brief delay of 0.6s. To indicate that an object was a match, the monkey had to saccade to a subsequently presented visual target, a white dot presented 7.5° to the left or right of fixation. And if it did so, the animal received juice and a new trial began within 3s. If incorrect, instead of juice, a red "error screen" flashed on for 1.5 s, and the animal had to wait 6s for the subsequent trial. The location (left versus right) of the response target after each associate was randomized and unrelated to task performance.

**Category-Saccade Task:** The details of this task have also been presented previously (Antzoulatos and Miller, 2011 and 2014). In this task, animals had to learn to classify a number of category exemplars generated from two different prototypes into two categories. Each category was directly tied to a specific saccadic target (right or left). To

start a trial, animals first fixated within 1.5-2° DVA of a red, central target (0.4° DVA in diameter) for 0.7s. After this fixation period, a randomly chosen category exemplar (6° by 6° DVA) from either category was presented for 0.6s. Trials from both categories were randomly interleaved throughout the session. One second after the end of the exemplar period, two saccade targets (a green dot, 0.6° DVA in diameter) appeared on the left and right of the center of fixation (5° DVA from the center). In order to indicate a response, the animals had to make a single, direct saccade within 1s of the saccade target presentation and maintain fixation on it for 0.2s. If the animal chose correctly, it was rewarded with drops of juice. If the animal did not, it was punished with a 5-s timeout, during which the cue was presented again, at the location of the corresponding target.

**Category-Match Task:** In each session, animals had to classify a number of category exemplars generated from two novel different prototypes into two categories. Each category, however, was neither tied to any particular saccade nor saccade location. Instead at the test period, the animal had the opportunity to freely investigate two exemplars, one of which matched the category of the sample exemplar, and then had to choose by fixating on this match. To initiate each trial, each animal had to fixate within 2.5° DVA of a centrally located, red dot (0.2° DVA in diameter) for 0.5s. After this fixation, an exemplar of one of the two categories was presented at the center of the screen (7° by 7° DVA) for 1s. If the animal continued to fixate through this sample period and a subsequent delay of 0.85s (with an additional jitter of max. 0.4s), then the central fixation dot disappeared and two new exemplars were presented on the left and

right side of the screen (9° DVA from the center of the screen). Once the test exemplars appeared, the animal had the opportunity to freely view both of the exemplars presented and make the correct choice. To indicate this choice, the animal had to fixate on one of the two peripherally presented exemplars for 0.7s. If it made the correct choice, the white dots of the chosen exemplar turned green and the animal received juice. If the animal did not make the correct choice, the chosen exemplar turned red and no juice was given. Depending on the animal, the length of timeout incurred on error trials varied from 5-16s.

### *Neurophysiology and Hardware*

**Category-Saccade Task & Object-Match Task:** In both the *Category-Saccade* and *Object-Match* task stimulus presentation and reward delivery were controlled by Cortex (NIMH, Laboratory of Neuropsychology) and presented on a 100 Hz CRT monitor. Eye movements and pupil size were monitored and recorded using an infrared eye tracking system (Eyelink I & Eyelink II, SR Research @ 500 Hz). In these tasks, up to 16 electrodes were lowered in PFC, HPC, or STR acutely. All recordings from PFC and STR, and most from HPC, were performed with epoxy-coated tungsten microelectrodes (FHC). On some HPC recordings, 24-channel linear probes with 300-um spacing between adjacent platinum iridium contacts were used (U-probes, Plexon). For targeting, the animals' implanted chambers were co-registered with structural MRI images. For all of the PFC, STR, and some HPC recordings, these electrodes were lowered daily though the dura using custom-built, screw micro-drives. The exact location on the grid and orientation of the grid were varied to limit cortical damage and maximize coverage of the

intended regions. For the linear probes, electrodes were lowered through a 25-gauge transdural cannula using a motorized drive system (NAN-S4, NAN instruments). The electrodes would be lowered until spiking was detected, and then electrodes were allowed to sit for about an hour to limit apparent neural drift. Neural activity was amplified, filtered, digitized and stored using an integrated multichannel recording system (Multichannel Acquisition Processor, Plexon). The signal from each electrode was amplified by a high input–impedance, unitary gain headstage (HST/8o50-G1, Plexon), referenced to ground, filtered from 0.7–300 Hz, and amplified 1000-fold. LFPs were recorded continuously at 1 kHz. Only electrodes with cells present on them were included for these analyses, and after trial cutting, for all of the synchrony analyses, evoked potentials were subtracted out from each individual trial.

**Category-Match Task:** In the *Category-Match* task, stimulus presentation and reward delivery were controlled by custom software written in Matlab using PsychToolbox. All stimuli were presented on a LCD screen at 144 Hz (ViewSonic VG2401mh 24" Gaming Monitor). Eye movements and pupil size were monitored using EyeLink II at 500 Hz sampling. Four 8x8 channel Blackrock Cereport arrays with 1mm long electrodes were implanted in dorsomedial prefrontal cortex (dmPFC), dorosolateral prefrontal cortex (dlPFC), and ventrolateral prefrontal cortex (vlPFC). Each electrode was separated by 400 µm. vlPFC, dlPFC, and dmPFC were all defined by anatomical landmarks following the craniotomy. The vlPFC array was placed 1 mm ventral to the principal sulcus and was centered at 9-12 mm anterior to the genu of the arcuate sulcus. In contrast, the dlPFC

array was positioned slightly more rostral, 12-15 mm anterior to the genu of the arcuate and 1 mm dorsal to the principal sulcus. Finally, we placed the dmPFC (dorsomedial prefrontal cortex) array in the vicinity of where others have reported to identify the supplementary eye fields. The medial edge of the array was placed 5mm from the midline, and 5mm anterior to the genu of the arcuate sulcus. Signals were recorded through a headstage (Blackrock Cereplex M and Cereplex E), sampled at 30 kHz, band-passed between 0.3 Hz and 7.5 kHz ($1^{st}$ order Butterworth high-pass and $3^{rd}$ order Butterworth low-pass), and digitized at a 16-bit, 250 nV/bit. All LFPs were recorded with a low-pass 250 Hz Butterworth filter, referenced to ground, sampled at 1 kHz, and AC-coupled. In Monkey G, an error in the design of the Cereplex E head-stage made the system susceptible to ground loops and to DC-drifts in the signal. This required us to apply a low-pass, 0.5 Hz FIR filter in both directions on the whole dataset to avoid any phase distortions. All arrays had units present on at least 5, if not typically a large proportion of channels. All channels were included in this analysis, and for all synchrony analyses the evoked potentials averaged across trials were subtracted from each individual trial.

### *Prototype and exemplar generation*

In both the *category-match* and *category-saccade* tasks, the visual stimuli were composed of 7 randomly located dots on a black background. To construct the categories, we followed previously published procedures (Posner et al., 1967; Vogels et al., 2002, Antzoulatos and Miller, 2011). Every day, two novel prototypes were created at random.

These prototypes (as would be the exemplars) were generated as 7 arbitrarily positioned, 7-pixel dots on a grid of 140 by 140 pixels. In order to control for difficulty and ease, these arbitrarily constructed prototypes had to obey a number of rules: (1) They had no dot centers that fell within 14 pixels of one another. (2) The average dot position of the prototype was at the center of the grid. (3) No dots from each exemplar fell within a 10-dot margin on the edge. And, (4) the minimum Euclidean distance between all pairs of dots between each prototype was no greater than 200 pixels. Each of these 140 x 140 pixel exemplars a subtended 6-7 degrees of visual angle.

In order to generate the exemplars, the prototype dot patterns were jittered according to a procedure first established by Posner and colleagues (Posner et al., 1967). To determine this jitter, we first defined 5 concentric annular regions. These annuli were centered around each dot, and spaced apart radially by 7 pixels. Region 1 refers to the annulus immediately surrounding the dot center, 1 dot-diameter away, and region 5 refers to the annulus 5 dot-diameters away from this dot. Next, each dot from each prototype was shifted away from its prototypical location by at least 1 region; no exemplar was identical to the prototype. Whether any particular dot was moved to regions 2 to 5 depended on the distortion level desired. Each exemplar had to be unique, different from any other exemplar, and, to ensure such, no more than 2 dots from each exemplar could be less than 10 pixels away from any other exemplar's dots.

Posner et al. defined 9 distinct levels of distortion, based on the probability of a dot to shift to each of these 5 concentric regions. Two of these 9 distortions were used in this task. At distortion level 1, 88% of dots were shifted to region 1, 10% to region 2, 1.5% to region 3, 0.4% to region 4, and 0.2% to region 5. At distortion 2, 75% of dots were shifted to region 1, 15% to region 2, 5% to region 3, 3% to region 4, and 2% to region 5. In the Category-Saccade task, the generated exemplars were largely at distortion level 2. In the Category-Match task, which appeared more difficult for animals to acquire, distortion level 1 exemplars were used for both animals. As a side note, in order to rule out that any of the reported effects were a result of the level distortion, we repeated 3 sessions in one of the two monkeys at distortion level 2. The results were similar; the monkey showed an enhancement of synchrony in the beta band on correct trials and theta band on incorrect trials.

Overall, the use of these visual stimuli in both tasks provided for us a number of advantages: (1) These categories were not imbued with any overt meaning to the subject, for they held no apparent relationship to objects seen in daily life. (2) The exemplars which could, in fact, look distinctively different from one another were always perceptually related and averaged out to the original prototype. And, (3) these categories could not be distinguished by any simple rule.

*Block design*

To facilitate learning, in both the *category-match* and *category-saccade task*, each learning session was organized into blocks. The blocks were defined by a progressively growing pool of available exemplars from each which any could be used for a given trial and any task period (sample or test exemplars). In any given block, with the exception of block 1 in the Category-Saccade task, there were a total of $2^{block}$ number of exemplars for each category. In the Category-Saccade task, in the first block, there was a single exemplar per category. The pool of available exemplars grew by accretion, "new" exemplars were added to a bank of "familiar" ones, so that the total available exemplar was equal to $2^{block}$. The terms novel and familiar are not an indication for how familiar any exemplar was to an animal, but simply a reflection of when it became available in the pool of potentially usable exemplars. As the blocks progressed, the chances for only seeing novel exemplars increased substantially, and performance on these novel exemplars suggested successful categorization. In fact, block transition was not possible without successful categorization, and the overlap of available exemplars between blocks favored a smoother learning process.

In order to pass from one block to another, each animal had to achieve a particular behavioral criterion. The criteria diverged somewhat between the two category tasks. In the *Category-Saccade* task, a block transition occurred when the animal had correctly responded to 80% of the previous 20 trials. In the *Category-Match* task, both animals had a tremendous capacity for being biased in either choosing a particular location and/or a

particular category. In order to limit these behavioral biases, each animal had to successfully complete 70% of the previous 10 trials for each potential condition (Category A – on left, Category A – on right, Category B – on left, and Category B – on right). Because of these behavioral criteria in both tasks, not all available exemplars were presented in each block. In the Category-Match task, an additional restraint was imposed on the pool of available exemplars presented in block 1. Because both animals struggled to pass block one, in which two exemplars from each category were presented, the exemplars from each category had to have a Euclidean distance of less than 20 pixels apart. This constraint reduced the difficulty of the first block, promoted rapid block passage, and ultimately favored category abstraction. Following block one, there was no limitation on the presented exemplars.

### *Bias Correction*

As stated above, the *Category-Match* was a more difficult task to maintain high levels of performance and, as a result, if left to each of their own devices, each of the animals engaged in suboptimal strategies and would fail to learn to categorize stimuli. To avoid these aberrant behaviors, we detected the animals' biases, and scaled the probability that any particular condition was shown to counteract these "easier," inefficient strategies. In order to assess bias in any one of the four conditions enumerated above, we compared performance in each of the four conditions to one another, and computed the Mann Whitney U test statistic (U) for each of these comparisons. $R_1$ represents the sum of ranks for condition 1 and the $n_1$ represents the sample size for sample 1. From this test statistic,

we obtained the area under the curve, subtracted 0.5 to obtain a bias measure, and remapped this bias measure to a value between [0-1] by dividing it by 0.5. We then used this measure to scale the probability that any particular condition would be seen. We only implemented this bias correction algorithm after 20 trials were performed in each of the blocks.

**Eqn. 2.1** $U = R_1 - \frac{n_1(n_1+1)}{2}$

**Eqn. 2.2** $AUC = \frac{U}{n_1 n_2}$

**Eqn. 2.3** $Bias = \frac{(AUC-0.5)}{0.5}$

*Learning Stages*

**Category-Saccade & Category-Match Tasks:** In a previous report, which analyzed this same Category-Saccade task, block 5 similarly marked the transition towards a late stage of learning (Antzoulatos and Miller, 2011). In that study, early learning (alternatively called stimulus-response association) was defined as the first two blocks of the task, where novel exemplar performance was at chance. By contrast, late learning (alternatively called category-performance) was defined as those blocks when novel performance on each category in a 16-trial window was above 75% correct. This behavioral criterion was reached on average by block 5. The middle stage of learning (alternatively called category learning) was defined, therefore, as those blocks which remained, i.e. those blocks between the early and late stages of learning, and hence blocks 3 and 4. Due to the congruence of the sliding t-test first presented here and those

previously published behavioral results, we separated learning into the same learning stages.

In the Category-Match task, novel performance had already plateaued by block 2 (on average, trial 89), and a division of learning stages by blocks was not appropriate. Because future analyses divided behavioral curves by 20 trial bins, we rounded the 89-trial criterion to the next multiple of 20, and defined early learning as the first 100 trials and late learning as the next 100. Moreover, because of the rapid learning particularly present in the *Category-Match* task, we restricted those analyses involving the learning stages in both the *Category-Saccade* and *Category-Match* tasks to those days in which learning was a bit more difficult and performance on all trials (not just novel) was less than or equal to 80% in the first 20 trials. In the *Category-Match* task, 39 days matched these criteria, evenly spread across both monkeys (23 days from Monkey P, 16 days from Monkey G). In the Category-Saccade task, 12 days matched these criteria (8 days from Monkey 1, 4 days from Monkey 2). Otherwise, all days from all monkeys in each task were pooled together. As we will see, learning in these analyses was assumed to be somewhat linear, and we expected to see physiological changes that correlated with average changes in performance in late vs. early trials. And while we did, we also showed (when binning trials from the entire session by performance) that none of these results depend on a particular learning stage classification nor an assumption that learning need be linear.

**Object-Match Task:** Because the Object-Match task was not blocked, unlike the category tasks, the behavioral criteria for learning stages differed as well. As previously reported, in the Object-Match task, only those sessions for which each of the final cue-associate pairings were performed at 32 correct response over the final 50 trials ($p \cong .01$, binomial test) were included in this final analysis. 61 sessions met these criteria, and trials within them were simply divided into thirds. The first-third of the session was early learning, the second-third middle learning, and the last-third late learning.

*Behavioral analyses*

In all three tasks, we computed performance in the trial following a correct trial or an incorrect trial irrespective of the following category or cue-associate presented. A binomial distribution was fit to each of the post-correct and post-incorrect trial performances, and 95% confidence intervals were generated. During these same trials, we compared reaction times on trials following correct and incorrect trials. In both the *Category-Saccade* task and *Object-Match* task, we computed the differences in the time it took the animals to respond following target presentation between correct and incorrect trials. In the *Category-Match* task, because the task involved free viewing, we computed the differences in time between the presentation of the match exemplars and the chosen response after controlling for the number of saccades. To control for saccade number, we subtracted correct and incorrect trial times when the animal completed a single saccade, and the same for two saccades. All significance testing was done on a distribution generated through 10,000 bootstraps of correct-incorrect trial time differences over

sessions in each task. From these distributions, we computed 95% confidence intervals by finding the 2.5% and 97.5% quantiles of each distribution.

*Evoked potential analysis and error-related negativity*

In three tasks, we computed evoked potentials on correct and incorrect trials by averaging across trials on each electrode from each region. As in previous work, in both the *Object-Match* and *Category-Saccade* tasks, recordings from each of these acute electrodes were considered independent (Brincat and Miller 2014, Antzoulatos and Miller 2011). In these two sets of recordings, pairs of electrodes were placed >1.5 mm apart across a large swath of ventrolateral and dorsolateral prefrontal cortex. Evoked potentials from all of these sites were averaged and plotted. For these analyses and others, we performed statistical testing by bootstrapping the trials and pairs to ensure that the reported findings were not dependent on any independence assumption. In the *Category-Match* task, evoked potentials were computed across an entire array and averaged across sessions. The error-related negativity is an evoked potential that arises more prominently on incorrect trials and occurs in epochs spanning 80-300 ms following feedback. In order to compute the error-related negativity, the evoked potential from incorrect trials was subtracted from correct trials. We aligned all of the tasks to their maximal peak (i.e. where the error-related negativity is greatest), and then normalized the entire signal by the mean and standard deviation of the differences across the entire trial and feedback (-2s to 1.7s post-feedback). This normalization step allowed us to eliminate those voltage differences related to impedance differences that varied between the arrays and the acute

electrodes (as can be observed in the different voltage scales on Figure 2.3). To compare the error-related negativity across the tasks, we computed the mean value of the normalized voltage differences for a 50ms period of time centered around the maximum difference from each task. We computed 10,000 bootstraps and found the empirical value for each comparison. In order to account for multiple comparisons, we applied a Bonferroni correction and multiplied all p-values by 3.

*Time-frequency analysis*

Spectrogram and coherograms were computed by applying the continuous wavelet transform to the trial-by-trial data. In order to compute the transform efficiently, for each complex-valued Morlet wavelet of wave number 6, we multiplied its Fourier transform by the Fourier transform of the trial data, and took the inverse Fourier transform of this product (Torrence and Compo, 1998). The 61 daughter wavelets sampled the frequency space exponentially (base of 2), starting at a max frequency of 80 Hz, and ending 6 ($num_{octaves}$) exponential decreases later (here, $80/2^6 = 1.25$ Hz). For every exponential decrease, here, for every halving, we sampled 10 frequencies (*num_per_octave*). All of the sampled frequencies from this paper can be given by the following equation:

**Eqn. 2.4.** $$Sampled\ Frequencies = \frac{max\_frequency}{2^{(0:\frac{1}{num\_per\_octave}:num_{octaves})}}$$

In order to obtain the amplitude or power from the complex output of the wavelet transform, we took the complex modulus of wavelet coefficients. Alternatively, for phase

values, we took the inverse tangent of the ratio of the real and imaginary components of the wavelet coefficients.

*Synchrony analysis*

To estimate synchrony, we computed the pairwise phase consistency (PPC), an unbiased estimator of the magnitude squared resultant (Vinck et al., 2010). In other words, the pairwise phase consistency statistic is a measure equivalent on the population level to the square of the oft-used phase locking value (PLV) but which is uninfluenced in its average by trial number. In this study, differences in trial number were prominent, as there were both differences in the number of incorrect and correct trials for any given session, as well as differences in the number of correct trials in any given 20-trial window. In order to implement the PPC statistic, we used a simplification derived previously (Kornblith, Buschman, and Miller, 2016)

**Eqn. 2.5:**

$$\text{PPC} = \frac{2}{N(N-1)} \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} \left(\cos\theta_j \cos\theta_k + \sin\theta_j \sin\theta_k\right)$$

$$= \frac{1}{N(N-1)} \left[ \left| \sum_{j=1}^{N} \exp(i\theta_j) \right|^2 - N \right]$$

$$= 1 - \text{var}(\cos\theta) - \text{var}(\sin\theta)$$

where $\theta$ is the vector of angular differences between two channels at any given point in frequency and time. *N* is the number of trials. For every single channel-pair, we subtracted the phases (obtained from the wavelet-transformed data) of one channel from those of another, took the variance of the sine and cosine of these angle differences across trials, and subtracted both of these variances from 1. To compute changes in synchrony early and late in learning, we bootstrapped the channels or sessions 1,000 times and obtained an empirical distribution. We used this distribution to estimate p-values. Because we were looking for differences above or below 0, we applied both a two-way test as well as a Bonferroni correction. The Bonferroni correction was used, instead of a more statistically powerful multiple comparisons correction, because it is particularly effective at controlling the type I error when different tests are potentially correlated.

### *Subtraction of eye movement*

Because one of the animals in the Category-Match task had some stereotyped eye movements after correct responses, we removed from both eye movement tasks (in this case, both Category tasks) the saccade related to the eye movement away from the response target. In order to do so, we identified the time of the saccade away by taking the derivative of the eye trace and setting a voltage threshold (+0.02V for the Category-Saccade, +0.05V for the Category-Match). We determined each of these thresholds based on the threshold evoked by the saccade to the target. Once this threshold was reached, we cut the data prior to the threshold crossing (50 ms) and after (Category-Match, 450ms; Category-Saccade, 250 ms). These times were determined by the shape of the evoked

potentials in each of the tasks (i.e. when the mean saccade potential returned to 0V). Different time windows around the saccade were used, and none changed the results. After all of the saccadic periods were cut from each trial, we obtained a template for a saccade for each electrode by taking the mean across all of saccades. Once we obtained this template, we returned to the original data, and applied linear regression to obtain an optimal fit of the template to the saccade potential on each individual trial. Once we found the optimal fit of the template to the data, we subtracted the template from the original data. In order to avoid spurious edge effects, we re-filtered all of the data in both directions by two FIR equiripple filters. Filter 1 was a 106 order, low pass, equiripple FIR filter at 80 Hz (Fpass = 80 Hz, Fstop = 100, Apass = 1, Astop = 60). Filter 2 was a 1047 order, high pass, equiripple FIR filter at 1.5 Hz (Fstop = 0.7, Fpass = 1.5, Astop = 20, Apass = 1).

### *Linear regression*

To both better assess the synchrony changes across learning, and avoid particular assumptions regarding whether learning occurred linearly over time, we estimated synchrony using PPC across all electrode pairs over 20-trial window bins across each session. We computed the PPC only on correct trials, for we had previously found that most of the longer duration synchrony events occurred on correct trials. These 20-trial windows were non-overlapping, and hence independent. For every 20-trial bin, we had a single PPC value (averaged across all electrode pairs of interest) and an average performance value. We pooled all 20-trial PPC values across all sessions in each of the 3 tasks, and performed linear regression to assess whether PPC varied directly with

performance. Our design matrix was simply a column of 1s (our intercept) for each 20-trial bin and another column for performance. After solving the normal equation and calculating the residuals, we computed a t-value and a corresponding p-value. In order to confirm that the assumptions underlying linear regression were met, we plotted how synchrony varied performance from chance (50%) to high performance (100%). We found that over the range of different performance level the variances at each level were largely equivalent with the exception at 100%. To ensure that our results were resistant to this possible violation of heteroscedasticity, we performed both linear regression on a reduced performance range from 50% to 95% (where variances were equal), and robust regression using iteratively reweighted least squares with a bi-square weighting function and a tuning constant of 4.685. Robust regression, unlike ordinary least squares regression, minimizes the influence of data points that are outliers by penalizing their influence. Neither of these methods changed the results; and are unreported here.

### *Category saccade controls*

Because eye movements tend to be rhythmic and occur in the 3-4 Hz range, we wanted to ensure that both correct-incorrect synchrony differences and learning-related synchrony decreases were unrelated to changes in eye movements. To address these concerns, we did two things: One, we controlled for the latency differences in the saccade away from the target that arose between correct and incorrect trials. Two, we examined whether eye movements, like theta synchrony, changed with learning. To control for the latency differences in the eye movement away from the target, we first estimated when saccades

were made following the initial saccade to the target using analog outputs from the eye monitoring system. After computing our wavelets and removing the evoked potentials related to the feedback, we aligned each individual trial to this saccade away from the target. Despite this procedure controlling for eye movements, theta synchrony differences emerged immediately following feedback, and did not appear influenced by the realignment procedure.

In order to investigate eye movement changes with learning, we compared the mean saccade velocity and the number of saccades made in the feedback period (0-1.7s). To assess saccade velocity, we cut the raw eye signal into trials, computed the animal's position on the screen relative to center (by taking the square root of the sum of squares of the X and Y eye position signals), and took the 1$^{st}$ derivative of the resulting position signal. Saccade velocity was averaged across the 1.7s of the feedback period, and the number of saccades identified was equal to the number of time points that exceeded a voltage threshold of .025V (~178 deg/s). The feedback period is a difficult time to capture behavioral events such as saccades, because the animal may move out of eye tracking limits. As a result, the values computed here may be underestimates of the true saccade count. Moreover, while visual inspection of the eye traces at the thresholds given above suggests that many of these events are saccades, we cannot preclude that some of the saccadic events are, in fact, eye blinks. Note that their inclusion here is conservative, and these flaws of behavioral observation are shared across tasks. To compute significance, we bootstrapped the trials and sessions 10,000 times.

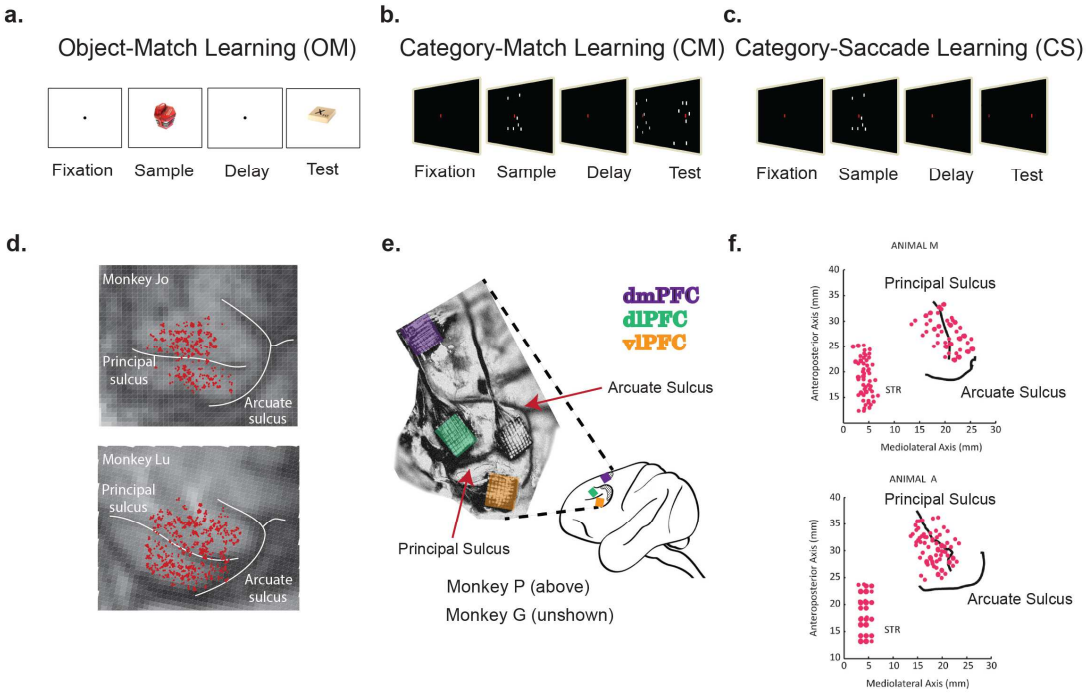


**Figure 2.1** ***A.*** In the Object-Match, task the animal was instructed to learn by trial- and error- whether a sample and test object were the correct pair. In each trial, the animal was first presented a sample object for 0.5s, and after a delay of 0.75s, a test object. The animal had to confirm whether the test object was the correct associate of the sample object (i.e. the correct match). In every session, each animal learned to associate 4 different sample objects with 2 test objects (Brincat and Miller, 2015). ***B.*** In the Category-Match task, each animal had to learn by trial-and-error two different, *de novo* dot-pattern categories. In each trial, the animal was first presented with a sample exemplar from one of two possible and, then after a variable delay (.85-1.25s), test exemplars one from each of the two possible categories. In order to select the test exemplar corresponding to the category of the sample, the animal had to fixate on it for .7s. In this task, the animals were matching the category of the sample to the category of the test, hence *Category-Match* learning. ***C.*** In the Category-Saccade task, each animal had to learn two different, *de novo* categories. In this task, the animal was presented with a sample exemplar for 0.6s, held fixation through a 1s delay, and had to indicate the category membership of this exemplar by making a saccade to the right or left target (Antzoulatos and Miller, 2011). ***D.*** In the Object-Match task, we recorded from 617 pairs within prefrontal cortex (PFC) and 941 pairs between PFC and hippocampus (HPC). Electrodes within PFC were spread equally across ventrolateral and dorsolateral prefrontal cortex (vlPFC and dlPFC). ***E.*** In the Category-Match task, we recorded from 64 electrode arrays in each vLPFC, dLPFC, and dorsomedial PFC (dmPFC). For PFC, we combined vlPFC and dlPFC electrode pairs, and recorded from 4032 pairs. For PFC-dmPFC, we recorded from 8192 pairs of electrodes. ***F.*** In the Category-Saccade task, we recorded from 240 PFC pairs (across both vl- and dlPFC), and 426 PFC-striatal (STR) electrode pairs.

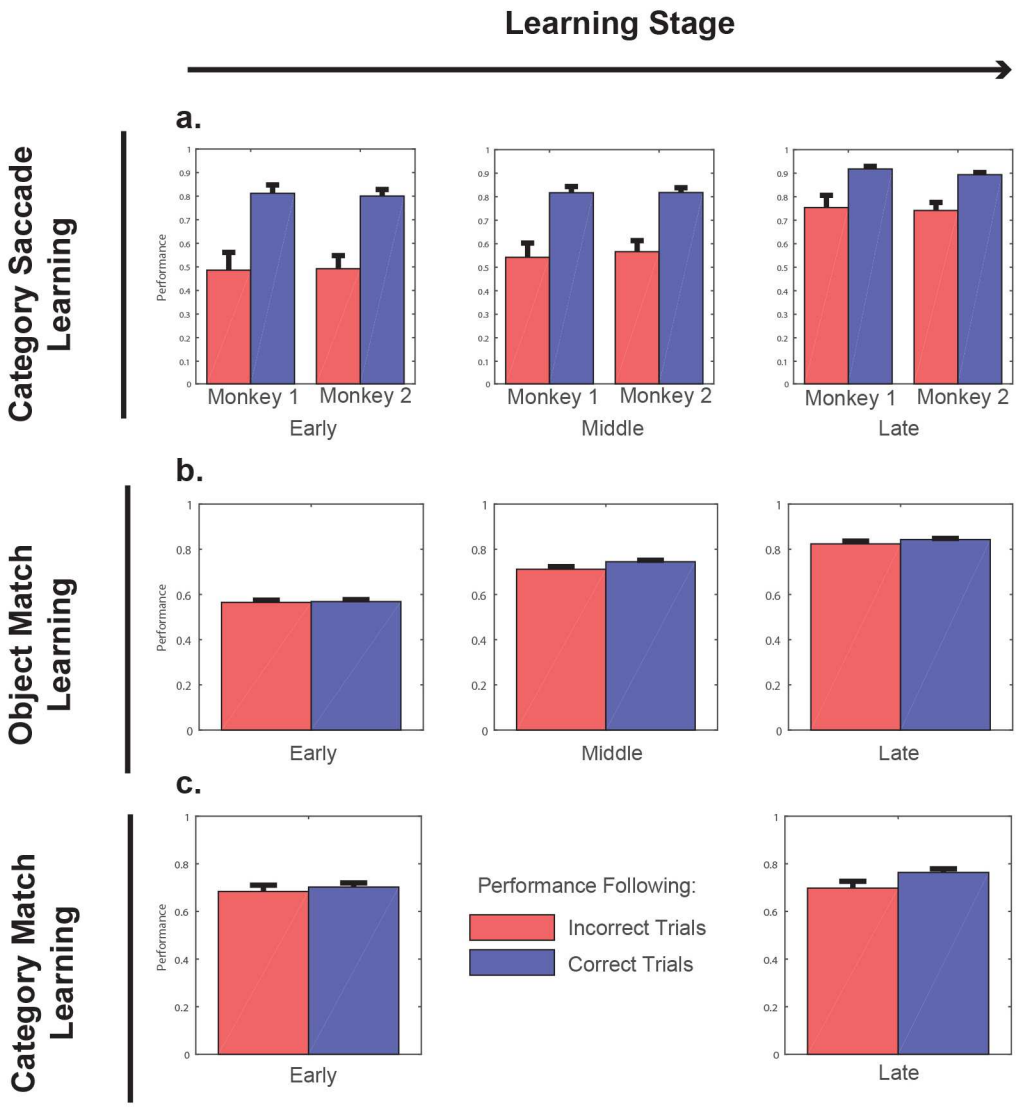


**Figure 2.2** In ***A***, each bar represents performance separated out by monkey on trials following either an incorrect response (red bar) or a correct response (blue bar) during Category-Saccade learning. All trials were pooled across days for each stage of learning (presented as separate columns). Error bars show 95% confidence interval generated from a binomial distribution. In ***B***, ***C*** we again plotted performance following a correct and incorrect trial for both the Object-Match and Category-Match tasks; however, both monkeys in each task were pooled.

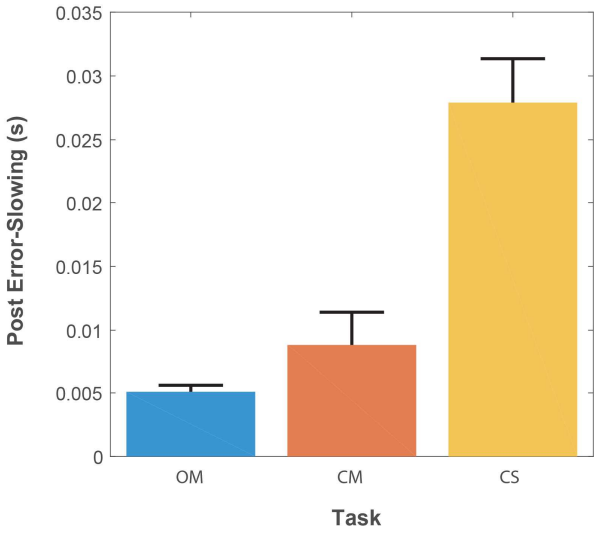


**Figure 2.3** Post-error slowing across the 3 different learning tasks: Object-Match (OM), Category-Match (CM), and Category-Saccade (CS). CM and OM post-error slowing were not significantly different ($p = 0.171566$). CS post-error slowing was significantly larger than both Match tasks ($p < 1 \times 10^{-4}$). Error bars represent the SEM.

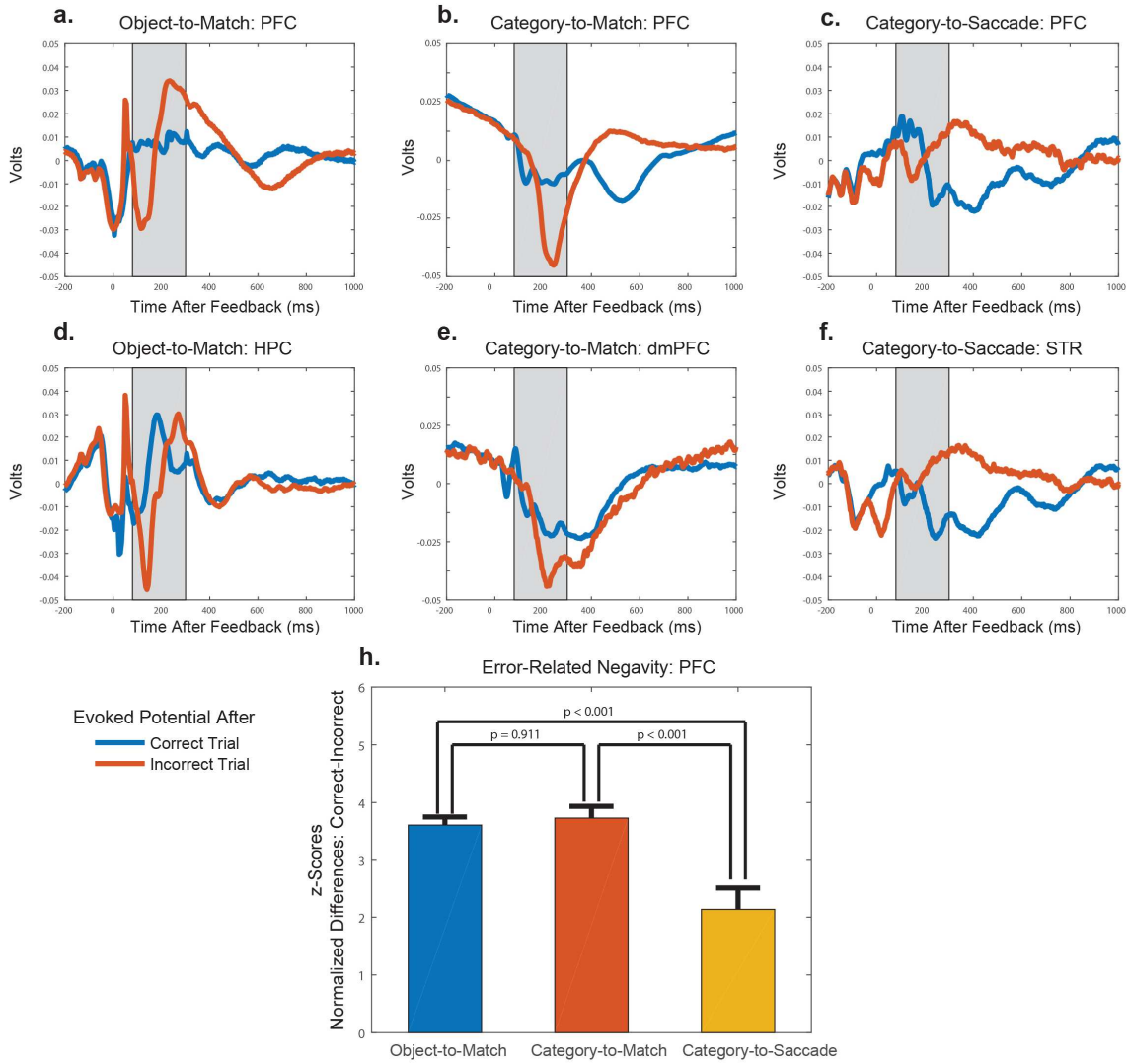


**Figure 2.4** ***A-C.*** Across all electrodes in prefrontal cortex, we computed the evoked potentials in each task separately. The blue line is the evoked potential averaged across correct trials, and the red line is the evoked averaged across the incorrect trials. The shaded grey region represents the time of an expected error-related negativity, between 80-300 ms after feedback. In the Object-Match and Category-Saccade tasks, we averaged across 242 and 64 electrodes distributed across vl- and dlPFC, respectively. In the Category-Match task, we averaged across each array within vl- and dlPFC (n = 97 arrays). ***D.*** Evoked potentials within the hippocampus (n= 162 electrodes), ***E.*** the supplementary eye fields (n = 30 arrays), and ***F.*** the striatum (n=65 electrodes) during the Object-Match, Category-Match, and Category-Saccade tasks respectively. ***G.*** The error-related negativity is plotted here for each task. The error-related negativity was computed by subtracting the evoked potentials on correct and incorrect trials, and aligned to the maximal differences across tasks. We compared the peak negativity by averaging around this peak (+/- 25 ms). The error bars represent +1 SEM.

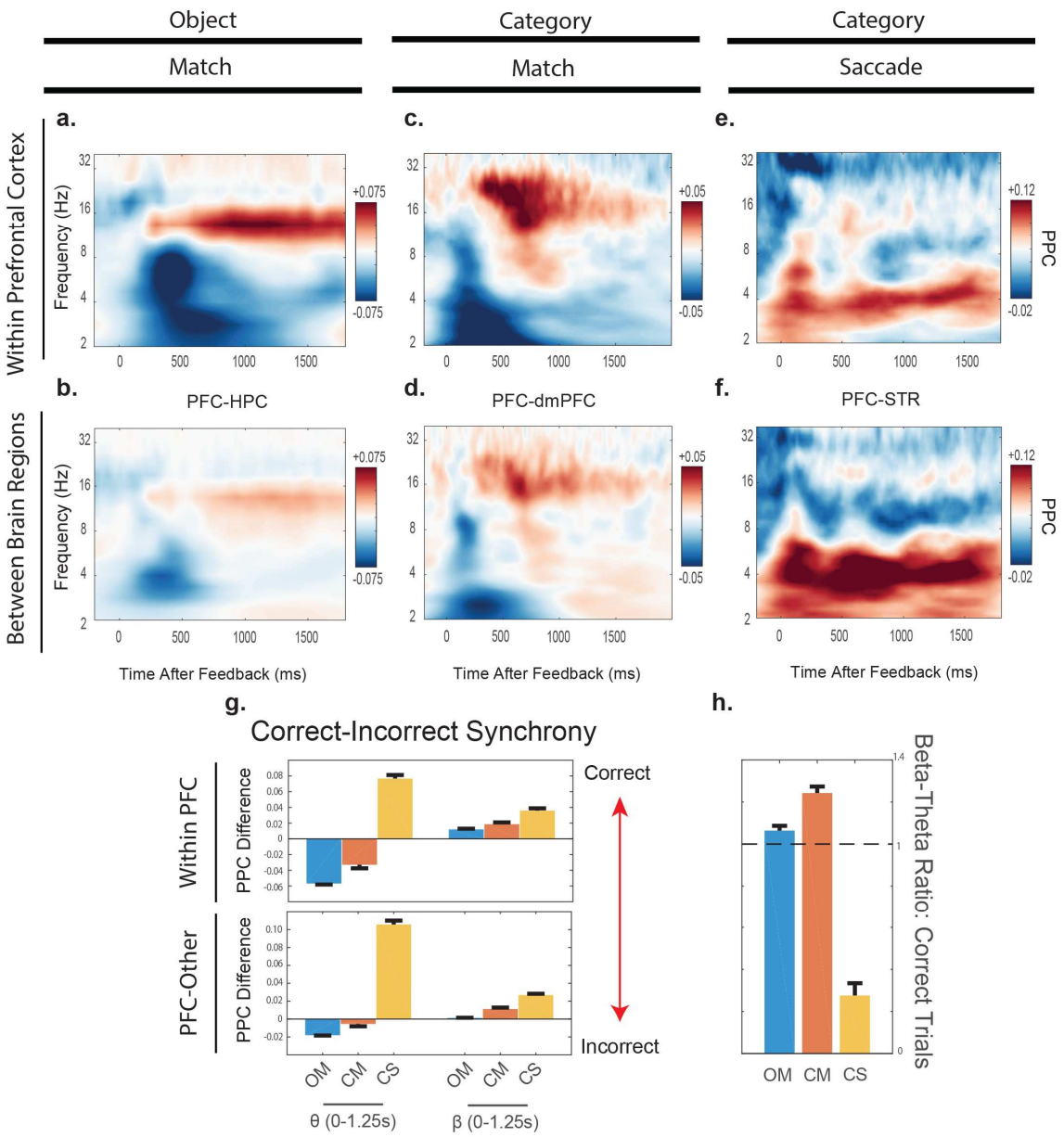


**Figure 2.5** In panels ***A-F.***, plotted is the difference in pairwise phase consistency between correct and incorrect trials. Red means a greater synchrony value on correct trials, and blue means a greater synchrony value on incorrect trials. Note the difference in color scales among the 3 panels. In ***G.***, we compared how the magnitude of these normalized synchrony values differed within the theta (3-7 Hz) and the alpha-2/beta bands (10-30 Hz) over time. In ***H***, we computed that the ratio of beta-theta synchrony on correct trials varied for each task. We found that in both Match tasks the beta-theta ratio was greater than 1. This was not true in the Category-Saccade task (ratio = 0.2759).

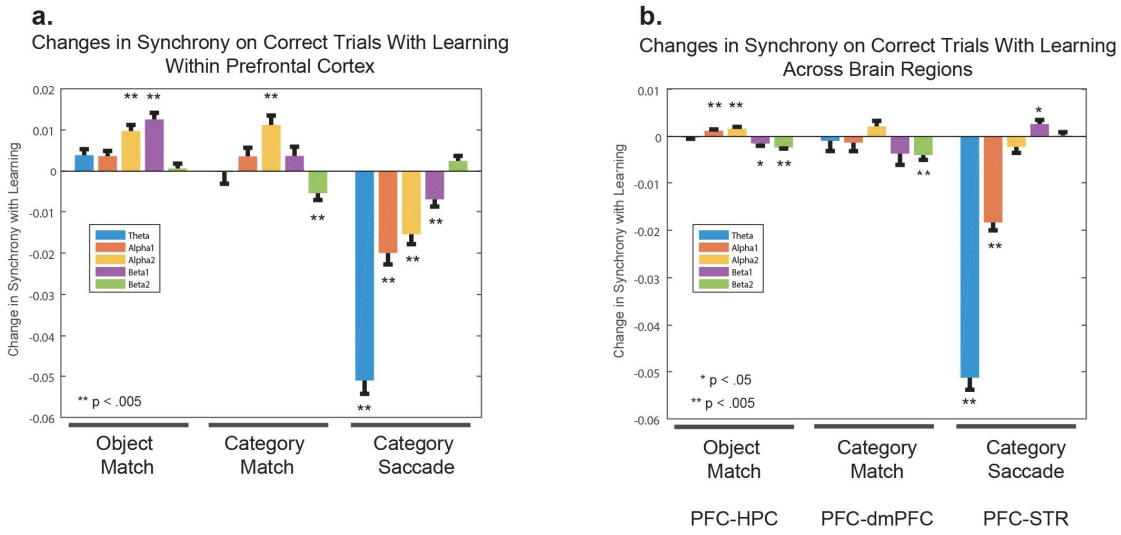


**Figure 2.6** We compared synchrony on correct trials early and late in learning. The bar plots represent a change from early to late learning, a positive value is associated with an increase in synchrony with learning and a negative value, a decrease. We computed synchrony at 5 different frequency bands: theta (3-7 Hz), alpha-1 (8-10 Hz), alpha-2 (10-12 Hz), beta-1 (13-17 Hz), and beta-2 (18 – 30 Hz). Error bars represent the SEM, and the stars represent significance. In ***A***., we compared synchrony increases within PFC. In ***B.*** we compared synchrony changes with learning across different brain regions (OM: PFC-HPC, CM: PFC-dmPFC, CS: PFC-STR).

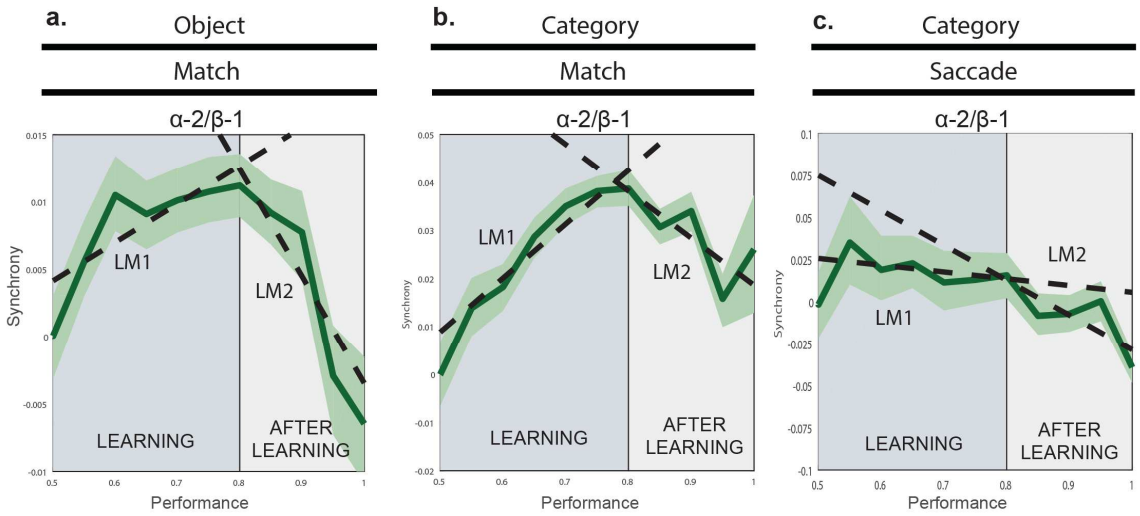


**Figure 2.7** Synchrony binned by non-overlapping, 20-trial window intervals. Error-bars represent the standard error of the mean. The dotted lines represent the fits of two different linear models, accounting for changes with learning. The mean synchrony at 50% performance level has been subtracted from each graph for visual purposes only. Linear Model 1 estimated the changes in synchrony with performance increases from 50% to 80%. Linear Model 2 estimated the changes in synchrony with performance increases from 80% to 100%. In **A.,** all electrodes from within and between PFC and HPC in the OM task were used to obtain a single synchrony value for each performance bin. In ***B.,*** we took all electrodes from within PFC and again obtained a single synchrony value for each 20-trial non-overlapping bin. ***C.*** We applied the same methods (as above) for all PFC electrodes within the Category-Saccade task. We repeated the same analysis with the PFC-STR electrodes (not shown), and still found no change before or after the criterion.

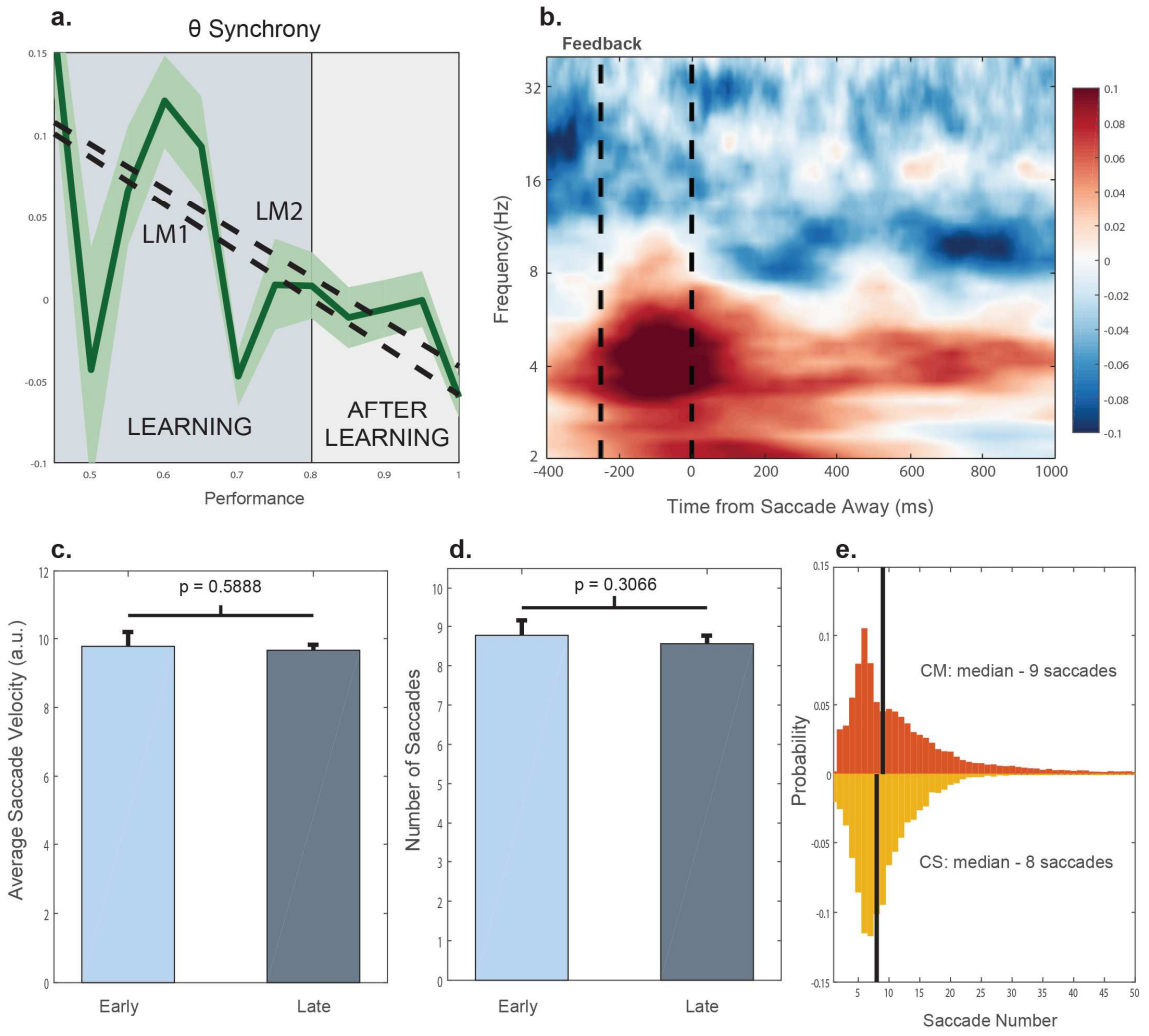


**Figure 2.8** ***A,*** Theta band synchrony binned by non-overlapping, 20-trial window intervals in the Category-Saccade task. LM1 and LM2 estimate changes with learning from 50-80% learning, and 80-100% learning respectively. ***B,*** Instead of aligning the LFPs to the feedback, all of the data was aligned to the saccade away from the target – which occurred on average 339ms after feedback. For the sake of visualization, this data has been normalized to the mean synchrony across the theta and beta bands. Here, theta synchrony increased on correct trials prior to the saccade away. ***C,*** Average saccade velocity early and late in learning (the mean of the first derivative of the eye position signal relative to the center of the screen). ***D,*** The number of saccades was taken as the number of threshold crossings of the eye signal during the entire feedback period (1.7s). ***E,*** Probability distributions of the number of saccades for both the Category-Match and Category-Saccade tasks.

## Bibliography

Aizenstein, H.J., MacDonald, A.W., Stenger, V.A., Nebes, R.D., Larson, J.K., Ursu, S., and Carter, C.S. (2000). Complementary category learning systems identified using event-related functional MRI. Journal of Cognitive Neuroscience *12*, 977–987.

Antzoulatos, E.G., and Miller, E.K. (2011). Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. Neuron *71*, 243–249.

Antzoulatos, E.G., and Miller, E.K. (2014). Increases in functional connectivity between prefrontal cortex and striatum during category learning. Neuron *83*, 216–225.

Asaad, W.F., Rainer, G., and Miller, E.K. (1998). Neural activity in the primate prefrontal cortex during associative learning. Neuron *21*, 1399–1407.

Ashby, F.G., and Maddox, W.T. (2011). Human category learning 2.0. Annals of the New York Academy of Sciences *1224*, 147–161.

Ashby, F.G., and O'Brien, J.B. (2005). Category learning and multiple memory systems. Trends in Cognitive Sciences *9*, 83–89.

Babapoor-Farrokhran, S., Vinck, M., Womelsdorf, T., and Everling, S. (2017). Theta and beta synchrony coordinate frontal eye fields and anterior cingulate cortex during sensorimotor mapping. Nature Communications *8*, 13967.

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P.L., Gioanni, Y., Battaglia, F.P., and Wiener, S.I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. Neuron *66*, 921–936.

Brincat, S.L., and Miller, E.K. (2015). Frequency-specific hippocampal-prefrontal interactions during associative learning. Nature Neuroscience *18*, 576–581.

Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., and Miller, E.K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron *76*, 838–846.

Carpenter, K.L., Wills, A.J., Benattayallah, A., and Milton, F. (2016). A comparison of the neural correlates that underlie rule-based and information-integration category learning. Human Brain Mapping *37*, 3557–3574.

Chen, L.L., and Wise, S.P. (1995). Supplementary eye field contrasted with the frontal eye field during acquisition of conditional oculomotor associations. Journal of Neurophysiology *73*, 1122–1134.

Clare, L., Wilson, B.A., Breen, K., and Hodges, J.R. (1999). Errorless learning of face-name associations in early Alzheimer's disease. Neurocase *5*, 37–46.

Cohen, N.J., and Squire, L.R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science *210*, 207–210.

Colgin, L.L. (2013). Mechanisms and functions of theta rhythms. Annual Review of Neuroscience *36*, 295–312.

DeCoteau, W.E., Thorn, C., Gibson, D.J., Courtemanche, R., Mitra, P., Kubota, Y., and Graybiel, A.M. (2007). Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task. PNAS *104*, 5644–5649.

Donaghey, C., McMillan, T., and O'Neill, B. (2010). Errorless learning is superior to trial and error when learning a practical skill in rehabilitation: a randomized controlled trial. Clinical Rehabilitation *24*, 195–201.

Evans, J.J., Wilson, B.A., Schuri, U., Andrade, J., Baddeley, A., Bruna, O., Canavan, T., Sala, S.D., Green, R., Laaksonen, R., et al. (2000). A comparison of "errorless" and "trial-and-error" learning methods for teaching individuals with acquired memory deficits. Neuropsychological Rehabilitation *10*, 67–101.

Frank, M.J., Woroch, B.S., and Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. Neuron *47*, 495–501.

Gehring, W.J., and Willoughby, A.R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. Science *295*, 2279–2282.

Gureckis, T.M., James, T.W., and Nosofsky, R.M. (2010). Re-evaluating dissociations between implicit and explicit category learning: an event-related fMRI Study. Journal of Cognitive Neuroscience *23*, 1697–1709.

Hargreaves, E.L., Mattfeld, A.T., Stark, C.E.L., and Suzuki, W.A. (2012). Conserved fMRI and LFP signals during new associative learning in the human and macaque monkey medial temporal lobe. Neuron *74*, 743–752.

Herweg, N.A., Apitz, T., Leicht, G., Mulert, C., Fuentemilla, L., and Bunzeck, N. (2016). Theta-alpha oscillations bind the hippocampus, prefrontal cortex, and striatum during recollection: evidence from simultaneous EEG–fMRI. Journal of Neuroscience *36*, 3579–3587.

Histed, M.H., Pasupathy, A., and Miller, E.K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron *63*, 244–253.

Huerta, P.T., and Lisman, J.E. (1995). Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro. Neuron *15*, 1053–1063.

Hyman, J.M., Wyble, B.P., Goyal, V., Rossi, C.A., and Hasselmo, M.E. (2003). Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to the peak of theta and long-term depression when delivered to the trough. Journal of Neuroscience *23*, 11725–11731.

Jutras, M.J., Fries, P., and Buffalo, E.A. (2009). Gamma-band synchronization in the macaque hippocampus and memory formation. Journal of Neuroscience *29*, 12521–12531.

Jutras, M.J., Fries, P., and Buffalo, E.A. (2013). Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. PNAS *110*, 13144–13149.

Knowlton, B.J., and Squire, L.R. (1993). The learning of categories: parallel brain systems for item memory and category knowledge. Science *262*, 1747–1749.

Kornblith, S., Buschman, T.J., and Miller, E.K. (2016). Stimulus load and oscillatory activity in higher cortex. Cerebral Cortex *26*, 3772–3784.

Lee, J.C., and Livesey, E.J. (2017). The effect of encoding conditions on learning in the prototype distortion task. Learning & Behavior *45*, 164–183.

Liebe, S., Hoerzer, G.M., Logothetis, N.K., and Rainer, G. (2012). Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. Nature Neuroscience *15*, 456–462.

Maxwell, J.P., Masters, R.S.W., Kerr, E., and Weedon, E. (2001). The implicit benefit of learning without errors. The Quarterly Journal of Experimental Psychology Section A *54*, 1049–1068.

Milner, B., Corkin, S., and Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. Neuropsychologia *6*, 215–234.

Milton, F., Bealing, P., Carpenter, K.L., Bennattayallah, A., and Wills, A.J. (2016). The Neural Correlates of Similarity- and Rule-based Generalization. Journal of Cognitive Neuroscience *29*, 150–166.

Milton, F., and Pothos, E.M. (2011). Category structure and the two learning systems of COVIS. European Journal of Neuroscience *34*, 1326–1336.

Morrison, R.G., Reber, P.J., Bharani, K.L., and Paller, K.A. (2015). Dissociation of category-learning systems via brain potentials. Frontiers in Human Neuroscience *9*, 389.

O'Connell, G., Myers, C.E., Hopkins, R.O., P, R., Gluck, M.A., and Wills, A.J. (2016). Amnesic patients show superior generalization in category learning. Neuropsychology *30*, 915–919.

Palmeri, T.J., and Mack, M.L. (2015). How experimental trial context affects perceptual categorization. Frontiers in Psychology *6*, 180.

Pasupathy, A., and Miller, E.K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature *433*, 873–876.

Poolton, J.M., Masters, R.S.W., and Maxwell, J.P. (2005). The relationship between initial errorless learning conditions and subsequent performance. Human Movement Science *24*, 362–378.

Posner, M.I., Goldsmith, R., and Welton, K.E., Jr. (1967). Perceived distance and the classification of distorted patterns. Journal of Experimental Psychology *73*, 28–38.

Reber, P.J., Gitelman, D.R., Parrish, T.B., and Mesulam, M.M. (2003). Dissociating

explicit and implicit category knowledge with fMRI. Journal of Cognitive Neuroscience *15*, 574–583.

Reber, P.J., Stark, C.E.L., and Squire, L.R. (1998). Contrasting cortical activity associated with category memory and recognition memory. Learning & Memory *5*, 420–428.

Roberts, J.L., Anderson, N.D., Guild, E., Cyr, A.-A., Jones, R.S.P., and Clare, L. (2016). The benefits of errorless learning for people with amnestic mild cognitive impairment. Neuropsychological Rehabilitation 1–13.

Rüsseler, J., Kuhlicke, D., and Münte, T.F. (2003). Human error monitoring during implicit and explicit learning of a sensorimotor sequence. Neuroscience Research *47*, 233–240.

Sakai, K., and Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. Nature *354*, 152–155.

Scheffers, M.K., and Coles, M.G.H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. Journal of Experimental Psychology: Human Perception and Performance *26*, 141–151.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of Neurology, Neurosurgery & Psychiatry *20*, 11–21.

Seger, C.A., and Miller, E.K. (2010). Category learning in the brain. Annual Review of Neuroscience *33*, 203–219.

Smith, J.D., Beran, M.J., Crossley, M.J., Boomer, J.T., and Ashby, F.G. (2010). Implicit and explicit category learning by macaques (Macaca mulatta) and humans (Homo sapiens). Journal of Experimental Psychology: Animal Behavior Processes *36*, 54–65.

Smith, J.D., Boomer, J., Zakrzewski, A.C., Roeder, J.L., Church, B.A., and Ashby, F.G. (2013). Deferred feedback sharply dissociates implicit and explicit category learning. Psychological Science *25*, 447–457.

Squires, E.J., Hunkin, N.M., and Parkin, A.J. (1997). Errorless learning of novel associations in amnesia. Neuropsychologia *35*, 1103–1111.

Vogels, R., Sary, G., Dupont, P., and Orban, G.A. (2002). Human brain regions involved in visual categorization. NeuroImage *16*, 401–414.

Walsh, M.M., and Anderson, J.R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. Neuroscience & Biobehavioral Reviews *36*, 1870–1884.

Wang, L., Saalmann, Y.B., Pinsk, M.A., Arcaro, M.J., and Kastner, S. (2012). Electrophysiological low-frequency coherence and cross-frequency coupling contribute to BOLD connectivity. Neuron *76*, 1010–1020.

Warden, M.R., and Miller, E.K. (2010). Task-dependent changes in short-term memory in the prefrontal cortex. Journal of Neuroscience *30*, 15801–15810.

de Werd, M.M., Boelen, D., Rikkert, M.G.O., and Kessels, R.P. (2013). Errorless learning of everyday tasks in people with dementia. Clinical Interventions in Aging *8*, 1177–1190.

Wessel, J.R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. Frontiers in Human Neuroscience *6*, 88.

Wessel, J.R., Danielmeier, C., and Ullsperger, M. (2011). Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. Journal of Cognitive Neuroscience *23*, 3021–3036.

Williams, Z.M., and Eskandar, E.N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. Nature Neuroscience *9*, 562–568.

Wirth, S., Avsar, E., Chiu, C.C., Sharma, V., Smith, A.C., Brown, E., and Suzuki, W.A. (2009). Trial outcome and associative learning signals in the monkey hippocampus. Neuron *61*, 930–940.

Wirth, S., Yanike, M., Frank, L.M., Smith, A.C., Brown, E.N., and Suzuki, W.A. (2003). Single neurons in the monkey hippocampus and learning of new associations. Science *300*, 1578–1581.

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

# CHAPTER 3: DIFFERENT LEVELS OF CATEGORY ABSTRACTION BY DIFFERENT RHYTHMS IN DIFFERENT PREFRONTAL AREAS

Andreas Wutz†, Roman Loonis†, Jefferson E. Roy, Jacob A. Donoghue and Earl K. Miller*

The Picower Institute for Learning & Memory and Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139, USA.

Department of Anatomy and Neurobiology, Boston University, 72 East Concord, Boston, MA 02116, USA

† These authors contributed equally.

* Correspondence to: ekmiller@mit.edu

## Abstract

Categorization is a hallmark of cognition. Categories can be grouped either by shared sensory attributes (i.e. cats) or by a more abstract rule (i.e. animals). We aimed to explore the dynamic brain networks underlying categorization at different levels of abstractness. We recorded from multi-electrode arrays in prefrontal cortex (PFC) while monkeys performed a dot-pattern categorization task. We found that different oscillatory rhythms in different PFC regions process different levels of category abstractness. Gamma rhythms in ventral PFC built less abstract categories based on bottom-up sensory inputs,

whereas beta rhythms in dorsal PFC extracted the category's essence (top-down) for higher levels of abstractness. Our results suggest a two-stage, rhythm-based model in distinct anatomical brain regions for "the genesis of abstract ideas".

## Results

Categorization is the capacity to organize items based on shared characteristics. These characteristics can vary by level of abstractness. Sometimes category membership is more feature-based with members looking similar (e.g., housecats), and other times they can be more conceptual with members looking quite different (e.g., cats and elephants). But how are these different levels of abstractness achieved by the brain? Do different levels of abstractness engage different mechanisms and different networks? We used a dot-pattern categorization task (Posner and Keele, 1968), in which the level of abstractness is controlled by the degree of spatial distortion of each exemplar pattern from its category prototype. Low distortion exemplars look alike. High distortion exemplars require greater abstraction of the category's "essence" (Figs. 3.1A and B). Monkeys distinguished two new categories in each session in a delayed-match-to-category paradigm (Fig. 3.1C). We recorded local field potentials (LFPs) from chronically implanted multi-electrode arrays in dorsolateral and ventrolateral prefrontal cortex (dlPFC, vlPFC; Fig. 3.1D), where the neural correlates of categorization have been reported (Wallis et al., 2001; Cromer et al., 2010). We found that category abstractness was organized by area and oscillatory rhythm. vlPFC-gamma oscillations were more

engaged for lower-level category abstractness, whereas dlPFC-beta oscillations took over for higher levels of abstractness.

There was a task-related increase in LFP oscillatory power over baseline levels in different frequency bands in dlPFC vs vlPFC (Wilcoxon sign-rank z-score relative to -1 to -.75 s before sample onset). In dlPFC, it was mainly in the beta band during the sample and memory delay epoch (Fig. 3.2A; 13-30 Hz for 0-1s and 13-28 Hz for 1-2 s). By contrast, in the vlPFC (Fig. 3.2B), gamma power increased and beta power decreased during stimulus presentation (63-110 & 13-33 Hz for 0.1-0.3 s and 46-172 & 11-33 Hz for 2.5-2.7 s). The increase in gamma power after sample onset (0.1-0.3 s) was maximal between 80-120 Hz (Fig. 3.S2). Over time (between 0-3 s), beta (10-35 Hz) and gamma power (60-160 Hz) were negatively correlated within each area (dlPFC *mean r* ± *SD*: -.3 ± .4, $t(29)= -4.1$, $p< 2.8\text{x}10^{-4}$; vlPFC: -.83 ± .1, $t(29)= -36$, $p< 1.3\text{x}10^{-25}$) and across areas (dlPFC beta with vlPFC gamma: -.64 ± .2, $t(29)= -20$, $p< 1.4\text{x}10^{-18}$).

We assessed category information in beta and gamma by computing the explained variance in the LFP signal by category ($\omega^2$-statistic, 0.5 to 1.5 s after sample onset). Figures 3.2C and D plot category information over time and frequency for electrodes with significant information in either beta or gamma (Fig. 3.S3; 25% electrodes in dlPFC beta, 49% in vlPFC beta, 71% in vlPFC gamma; $p< .001$, permutation test). There was significant category information in dlPFC beta (Fig 3.2C, 10-42 Hz for 0-1s and 14-27 Hz for 1-2 s). In vlPFC (Fig. 3.2D), category information was found in both low and high frequencies (1-25 & 35-200 Hz for 0-1s and 12-200 Hz for 1-2 s) but it was maximally concentrated between 80-120 Hz (Fig. 3.S4). There was both greater power (Fig. 3.2E) and category information (Fig. 3.2F) for beta in dlPFC, and for gamma in vlPFC. This yielded highly significant interactions between frequency bands and areas (power: $F(1,29)= 365$, $p< 1.1\text{x}10^{-16}$; information: $F(1,29)= 31.8$, $p< 4.3\text{x}10^{-6}$).

Gamma power has been associated with bottom-up, stimulus-evoked activity. Indeed, there was nearly a 3-fold stronger stimulus-evoked potential in vlPFC compared to dlPFC ($t(29)= 14.9$, $p< 4.3\text{x}10^{-15}$, Fig. 3.2G). In vlPFC, there was a significant positive correlation between the time course of the evoked response and gamma power and a negative correlation with beta power (vlPFC beta *mean r* ± *SD*: -.19 ± .1, $t(29)= -7.8$, $p< 1.4\text{x}10^{-8}$; vlPFC gamma: .21 ± .2, $t(29)= 6.9$, $p< 1.3\text{x}10^{-7}$). By contrast, in the dlPFC there was no significant correlation between the beta time course and the evoked response (-0.007 ± .2, $t(29)= -0.2$, $p< .8$). Furthermore, evoked activity in vlPFC carried significantly more category information compared to dlPFC (sample: $t(29)= 3$, $p< .006$;

delay: $t(29)= 2.9$, $p< .008$; Fig. 3.2H, also evident in the transient, low-frequency information in Fig. 3.2D). Thus, vlPFC (which had stronger gamma power) was more driven by stimulus-evoked activity than dlPFC.

Monkeys learned new categories each day and often performed better on one. In 2/3 of the sessions, there was a difference by more than 5% between preferred and non-preferred categories (Figs. 3.3A and B). This behavioral preference was reflected in each region's dominant frequency. There was a significant positive correlation between the difference in behavioral performance and the difference in beta (dlPFC) or gamma (vlPFC) power between the categories. In order to calculate this correlation, we arbitrarily subtracted the performance and power for one category from those of the other, revealing changes in performance from about -20 to +20% and in power from -10% to +10% (normalized to overall performance/ power). Figure 3.3C shows the significantly positive correlation between the differences in dlPFC beta power and performance ($p< .002$, see also Figs. 3.S5 and 3.S6). The correlation was based on stronger dlPFC beta power for the preferred category (Fig. 3.3D, $p< .006$). There was no significant correlation between dlPFC gamma power and performance (Fig. 3.S7). Conversely, in vlPFC there was a significant, positive correlation between gamma power and performance ($p< .02$, Figs. 3.3E, 3.S5 and 3.S6) because gamma was stronger for the preferred category (Fig 3F, $p< .024$). A further contrast with the dlPFC was that beta power in vlPFC negatively correlated with category preference during the delay epoch

($p< .016$, Fig. 3.S7). Thus, the greater the difference in performance between categories, the greater the difference in power in each area's dominant frequency.

Category abstractness had a greater effect on behavior for the preferred category. Performance for the preferred category as a function of abstractness (dot pattern distortion) was best fit with a decreasing sigmoid function with a sharp inflection point ($R^2= .87$, inflection ± *CI* at 1.1 ± .0004° visual angle). This was less so for the non-preferred category ($R^2= .5$, inflection at 1.2 ± .008°; Fig. 3.4A). We used the inflection point for the preferred category to separate exemplars into high vs low abstractness. This revealed main effects for category preference and abstractness on performance (Fig. 3.4B; preferred vs non-preferred, $F(1,29)= 41.2$, $p< 5.1x10^{-7}$; low vs high abstractness, $F(1,29)= 17.7$, $p< 2.2x10^{-4}$) and a significant interaction ($F(1,29)= 4.6$, $p< .042$). The performance difference between low and high abstractness levels was significant for the preferred category ($t(29)= 4.2$, $p< 2.5x10^{-4}$) but not for the non-preferred category ($t(29)= 1$, $p< 0.33$; Fig. 3.4B). Category information ($\omega^2$-statistic) was compared for the 10% of electrodes with the most category information (Fig. 3.S3). During the delay, dlPFC beta power only carried category information for high abstractness ($p< .001$, permutation test), which was significantly greater than that for low abstractness (low vs high: $t(29)= -3.5$, $p< .002$; Fig. 3.4C). By contrast, in vlPFC gamma showed significantly more category information for low than high abstractness during the sample epoch ($t(29)= 2.2$, $p< .034$; Fig. 3.4D). vlPFC beta showed no difference between the abstractness levels (Fig. 3.S8). The amount of category information (averaged between 0-2 s) on different levels of

abstractness interacted significantly between the frequency bands and PFC areas: vlPFC gamma had more category information for low abstractness, whereas dlPFC had more information for high abstractness ($F(1,29)= 13.3$, $p< .001$).

## Discussion

We found a dissociation between PFC sub-region (vlPFC and dlPFC), oscillatory frequency (gamma vs beta) and level of category abstractness. Gamma increased in vlPFC and beta did in dlPFC. Gamma has been associated with bottom-up and beta with top-down processing. For example, learned rules are expressed in beta (Engel and Fries, 2010; Buschman et al., 2012; Antzoulatos and Miller, 2014) while gamma oscillations are involved in encoding bottom-up information in working memory (Buschman and Miller, 2007; Fries et al., 2001; Lundqvist at al., 2016). Correspondingly, we found stimulus-evoked (bottom-up) potentials in the vlPFC. There, gamma carried more information at low category abstractness. In contrast, dlPFC beta was less tied to stimulus onset and carried more information at high category abstractness. This suggests a two-stage, rhythm-based model for category abstraction in PFC. Lower-level, feature-based categories are first extracted from ventral stream inputs via gamma network interactions in vlPFC. Then, dlPFC uses beta rhythms to encode more abstract categories that transcend appearance and depend more on top-down input. This model is also consistent with notions of feedforward and feedback connectivity. Gamma rhythms are thought to support the feedforward flow of cortical information while beta rhythms support feedback (Jensen et al., 2015; Bastos et al., 2015).

Animals could have divided exemplars in two categories ("A vs. B") or used an alternative strategy in which one category was dominant ("A") and, accordingly, all exemplars would be judged against it ("A vs not-A"). Our results suggest the latter. Behaviorally, one category was often preferred ("A") and, on this category, performance varied between low and high abstractness levels. Physiologically, the power increases in the dominant frequency in each PFC region were greatest for the preferred category ("A"), supporting its special treatment. There was a negative correlation between vlPFC beta power changes and behavioral preference. However, vlPFC beta was suppressed during the task and this suppression was stronger for the preferred category.

Disorders like autism are marked by a decreased capacity to categorize and those like schizophrenia by a confusion between bottom-up and top-down signaling (Gastgeb et al., 2012; Uhlhaas and Singer, 2010). Our results support an anatomically distinct, rhythm-based model for category abstraction in the PFC. They might guide the way to new key insights into the underlying pathology and therapy of psychiatric disorders, as well as into the creation of abstract ideas by the brain.

## Methods

### *Prototype and exemplar generation*

The visual stimuli were composed of 7 randomly located dots on a black background. To generate the categories, we followed previously published procedures (Posner and Keele,

1968; Antzoulatos and Miller, 2011; Vogels et al., 2002). Figure 3.1A shows two example categories. Every day, two novel prototypes were created at random. These prototypes (as would be the exemplars) were generated as 7 arbitrarily positioned, 0.35° DVA dots on a grid of 7° by 7° DVA. In order to control for difficulty and ease, these arbitrarily constructed prototypes had to obey a number of rules: (1) They had no dot centers that fell within 0.7° DVA of one another. (2) The average position of the prototype was at the center of the grid. (3) No dots from each exemplar fell within a 0.5° DVA margin on the edge. And, (4) the maximum Euclidean distance (summed across all pairs of dots) between each exemplar and each prototype was 10° DVA.

In order to generate the exemplars, the prototype dot patterns were distorted according to a procedure first established by Posner and colleagues (Posner and Keele, 1968). We first defined 5 concentric annular regions around each dot, which were spaced apart radially by 0.35° DVA. Region 1 refers to the annulus immediately surrounding the dot center, 1 dot-diameter away, and region 5 refers to the annulus 5 dot-diameters away. Next, each dot was shifted away from its prototypical location by at least 1 region. Whether any particular dot was moved to regions 2 through 5 depended on the distortion level desired. Posner et al. defined different levels of distortion based on the probability of a dot-shift to each concentric region. Distortion level 1 was used in this task. At distortion level 1, 88% of dots were shifted to region 1, 10% to region 2, 1.5% to region 3, 0.4% to region 4, and 0.2% to region 5. To ensure that each exemplar was unique no more than 2 dots from each exemplar could be less than 0.5° DVA away from any other exemplar's dots. Across

all trials used here, the minimum and maximum Euclidean distance summed over all pairs of dots between each exemplar and its corresponding prototype (distortion) was 0.82° and 2.17° DVA, respectively. The minimum and maximum Euclidean distance summed over all pairs of dots between each exemplar and the nearest dots in the other prototype (distance between categories) was 3.15° and 5.94° DVA, respectively. Figure 3.1B shows the distributions of all available trials as a function of distance between the presented exemplar and its own category prototype (distortion, yellow) and to the other category prototype (between-category distance, green).

Overall, the use of these visual stimuli provided us with a number of advantages: (1) The categories were not imbued with any overt meaning to the animal, for they held no apparent relationship to objects seen in daily life. (2) The categories could not be distinguished by any simple rule. (3) The perceptual distance between categories was controlled over sessions and exemplars from different categories were different enough to ensure above chance performance. (4) The exemplars from each category, which could in fact look distinctively different from one another, were always perceptually related and averaged out to the original prototype. And (5), the stimuli provided parametric control over the similarity of sensory features between each exemplar and its prototype (distortion) and thus allowed us to vary the level of required abstraction.

*Task*

In each session, animals had to classify numerous category exemplars into their respective categories (delayed match-to-category task, Fig. 3.1C). To initiate each trial, each animal had to fixate within 2.5° degrees of visual angle (DVA) of a centrally located, red dot (0.2° DVA in diameter) for 0.5 s. After this fixation, an exemplar of one of the two categories was presented at the center of the screen (7° by 7° DVA) for 1s. If the animal continued to fixate through this sample epoch, and a subsequent delay of at minimum 0.85 s (plus an additional jitter of max. 0.4s), then the central fixation dot disappeared and two new exemplars from either category (match vs. non-match) were presented on the left and right side of the screen (9° from the center of the screen). Once the test exemplars appeared, the animal had the opportunity to freely view both of the exemplars presented, and make the choice. The animal indicated its choice by fixating for 0.7s on one of the two peripherally presented exemplars. Neither category was tied to any particular location. If the animal made the correct choice, the white dots of the chosen exemplar turned green and the animal received juice. If the animal made the wrong choice, the chosen exemplar turned red and no juice was given. Depending on the animal, the length of timeout incurred on error trials varied from 5-16 s.

*Block Design*

To facilitate category learning, each session was organized into blocks. The blocks were defined by a progressively growing pool of available exemplars. In block 1, there were two exemplars per category. The pool of available exemplars grew by accretion; “new”

exemplars were added to a bank of "familiar" ones, so that the total number of available exemplars for each category was equal to $2^{block}$. The terms novel and familiar are not an indication for how familiar any exemplar was to an animal, but simply a reflection of when it became available in the pool of potentially usable exemplars. As the blocks progressed, the chances for only seeing novel exemplars increased substantially, and above chance performance on these novel exemplars suggested successful categorization. In fact, block transition was not possible without successful categorization, and the overlap of available exemplars between blocks favored a smooth learning process. In order to pass from one block to another, each animal had to successfully complete 70% of the previous 10 trials for each potential condition (Category A – on left, Category A – on right, Category B – on left, and Category B – on right). This behavioral criterion ensured that the animals were able to categorize stimuli from an increasing pool of different exemplars and supposedly learned the underlying category rule by the end of the training blocks. We only used correct trials above training block 5 with a minimum of 64 different exemplars per category for the analysis of neurophysiological data presented here. In addition, the behavioral criterion limited idiosyncratic biases of the animals for either choosing a particular location and/or a particular category. Because of these behavioral criteria, not all available exemplars were presented in each block (see Bias correction below). An additional restraint was imposed on the pool of available exemplars presented in block 1. Because both animals struggled to pass block one, in which two exemplars from each category were presented, the two exemplars from each category had to have a summed Euclidean distance of less than 1° DVA apart. This constraint reduced the

difficulty of the first block, promoted rapid block passage, and ultimately favored category abstraction. Following block one, there was no limitation on the presented exemplars.

### *Bias Correction*

As stated above, each of the animals attempted suboptimal strategies (i.e. exhibited biased behavioral choices) and, if left to their own devices, they would fail to learn to categorize stimuli. To avoid these aberrant behaviors, we detected the animals' biases, and scaled the probability that any particular condition was shown to counteract these "easier," inefficient strategies. In order to assess bias in any one of the four conditions enumerated above, we compared performance in each of the four conditions to one another, and computed a Mann Whitney U test statistic for each comparison. From this test statistic, we obtained the area under the curve, subtracted 0.5 to obtain a bias measure, and remapped this bias measure to a value between [0-1] by dividing it by 0.5. We then used this measure to scale the probability that any particular condition would be seen (i.e. we forced more choices for the non-preferred condition by showing it more often). We only implemented this bias correction algorithm after 20 trials were performed in each block. The bias correction ensured that the animals' performance was above chance for exemplars from both categories. Despite this bias correction, the animals still maintained preferences for a particular category in each session.

*Recordings*

Stimulus presentation and reward delivery were controlled by custom software written in Matlab (The MathWorks, Natick, MA) using PsychToolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented on an LCD screen at 144 Hz (ViewSonic VG2401mh 24" Gaming Monitor). Eye movements and pupil size were monitored using EyeLink II at 1000 Hz sampling. Four 8x8 channel Blackrock Cereport arrays with 1mm long electrodes were placed within dorosolateral prefrontal cortex (dlPFC), and ventrolateral prefrontal cortex (vlPFC). Each electrode was separated by 400 μm. vlPFC and dlPFC were defined by anatomical landmarks following a large craniotomy. 3D MRI brain reconstructions and plastic models were used to guide the surgical implants of the array. The vlPFC array was placed 1 mm ventral to the principal sulcus and was centered at 9-12 mm anterior to the genu of the arcuate sulcus. In contrast, the dlPFC array was positioned slightly more rostral, 12-15 mm anterior to the genu of the arcuate and 1 mm dorsal to the principal sulcus. Figure 3.1D shows the approximate anatomical locations. Signals were recorded through a headstage (Blackrock Cereplex M and Cereplex E), sampled at 30 kHz, band-passed between 0.3 Hz and 7.5 kHz (1st order Butterworth high-pass and 3rd order Butterworth low-pass), and digitized at a 16-bit, 250 nV/bit. All arrays had units present on at least 5, if not typically a large proportion of channels. Local field potentials (LFPs) were recorded with a sampling frequency of 1 kHz, referenced to ground and AC-coupled.

Data from 15 recording sessions was analyzed for each of the two monkeys. For the analysis of neurophysiological data, we used equal proportions of trials from the two categories in each session (by drawing a random sub-sample of trials equal to the minimum trial number across categories). Further, we used only correct trials above training block 5 (see section Block design). Across sessions, on average 269 trials were used from monkey P (min= 94, max= 520) and 293 trials for monkey G (min= 140, max= 572). For the analysis on category information across different levels of distortion, we equated the trial numbers for each session between low and high distortion levels (see below for details), which reduced the number of available trials (monkey P: 185 trials, min= 48, max= 364; monkey G: 200 trials, min= 80, max= 360).

## Data analysis

### *Behavioral data*

Data was analyzed using custom Matlab code (The MathWorks, Natick, MA) and the Fieldtrip toolbox (Oostenveld et al., 2001). Behavioral and neurophysiological results were very similar between monkeys and therefore pooled across animals. Unless indicated otherwise, all analytical measures were calculated for each recording session separately. This yielded repeated measures between the tested conditions in each session and, thus, the statistical contrasts were calculated for a dependent-samples design across sessions. For instance, all error bars reflect the standard error of the mean for repeated measures (Morey, 2008).

Behavioral performance (percent correct trials) was well above chance (50 %) in every session (*mean* ± *SD*: 78 ± 6 %, *t*-value vs. 50 %: $t(29)= 27$, $p< 5.3x10^{-22}$, Fig. 3.3A). Performance was analyzed as a function of distortion level for preferred and non-preferred categories. Category preference was based on the proportion of correct trials for each category on any given recording day. The better-performed category was defined as "preferred". Performance was then pooled over preferred and non-preferred categories across recording days. Averaging across all sessions revealed a highly significant difference in performance between preferred vs non-preferred categories ($t(29)= 7$, $p< 1.1x10^{-7}$; Fig. 3.3B). For the effect of exemplar distortion on behavioral performance, we sorted all trials across all recording days by the summed Euclidean distance of the shown exemplar to its category prototype (distortion, see Prototype and exemplar generation above). Performance curves as a function of exemplar distortion were calculated by convolving the distortion-sorted performance vectors with a sliding-average window, encompassing 10% of the trials (width for preferred categories: 1243 trials; non-preferred categories: 1403 trials). Performance remained largely unchanged across a wide range of distortion levels but then decreased sharply at a critical distance from the prototype. Generalization across different exemplars and sharp distinctions with increasing category distance are hallmarks of categorization. Performance curves were fitted with a generalized logistic function (sigmoid S, Eqn. 1) with four free parameters (A = lower asymptote, B = upper – lower asymptote, C = steepness, x0 = mid-point) to estimate the inflection point (x0; see Fig. 3.4A).

**Eqn. 3.1**: $S = A + B / (1 + e^{(C * (x - x0))})$

The coefficient of determination ($R^2$) was used to determine the goodness of the fit, and the 95%-confidence intervals of the parameter estimates were calculated. In order to directly compare performance between low and high distortion levels and preferred and non-preferred categories, we split the data sets for each session at the estimated inflection point for preferred category trials (1.1° DVA). The effects of category preference and exemplar distortion (low vs. high split at the inflection point) across sessions were then tested with a two-way, repeated-measures analysis of variance (ANOVA) and the interaction effect explored with post-hoc dependent-samples t-tests (Fig. 3.4B).

### *LFP pre-processing*

The continuous local field potential (LFP) for each of the 64 electrodes on each recording array (vlPFC, dlPFC) was cut into trials between -2 s to 4 s around the sample onset. For the evoked response analysis, the LFP signal from each area was referenced to the same common reference (ground) and band-pass filtered between 1 and 15 Hz with a zero-phase Butterworth filter (4th order) applied in the forward and reverse direction. Before filtering, each trial was zero-padded to a length of 10 s to avoid edge artifacts. Stimulus-evoked activity was derived by averaging the LFP signal at each electrode across trials and baseline correcting it to the pre-trial interval between -1 to -0.75 s relative to sample onset. For the analysis of oscillatory power, LFPs were re-referenced to the array average, subtracting out the common signal components across all electrodes. LFPs were then band-pass filtered between 1 and 250 Hz and band-stop filtered around line noise

frequencies (60, 120, 180, 240 ± 1 Hz, same filter settings as above). In order to obtain induced activity without the contribution from stimulus-evoked LFP components, the trial-average was subtracted from each single trial.

Time-frequency representations were calculated using a Fourier transform applied to short sliding time windows in steps of 10 ms in the time interval between -1 to 3 s relative to sample onset and in the frequency range between 1 to 200 Hz. Fourier estimates were computed by means of a multi-taper transformation (discrete prolate spheroidal sequences (dpss), 3 tapers) applied to single trial data. The squared absolute value of the Fourier estimate gave the LFP signal power for each electrode across different frequencies and time points. For time-frequency representations and frequency spectra, we used a fixed 200 ms window width with a fixed amount of spectral smoothing (±10 Hz for frequencies between 1-200 Hz in steps of 1 Hz). This procedure yielded a good resolution in the frequency domain (see Figs. 3.2A-D, 3.S2, 3.S4). For the power time courses, we opted for a better temporal resolution (especially for higher frequencies) and used a frequency-dependent window width (5 cycles per frequency between 1-59 Hz in 1 Hz steps, between 60-99 Hz in 5 Hz steps, between 100-200 Hz in 10 Hz steps) and smoothing (0.4 times the frequency of interest). Subsequently, we averaged over the respective frequency bands to derive the time course of beta (10-35 Hz) and gamma power (60-160 Hz) and information (see Figs. 3.3D & F, 3.4C & D, 3.S5, 3.S7, 3.S8). The two methods yielded very similar results apart from the mentioned pay-off between spectral and temporal resolution.

*Task-related changes in LFP power*

LFP signal power during the sample presentation (0-1 s), the memory delay epoch (on average 1-2 s, jittered between 1 to 1.85-2.25 s) and at the test epoch (>2.25 s) was compared to the pre-trial baseline epoch (-1 to -0.75 s relative to sample onset) by means of a Wilcoxon signed-rank test. This time epoch was chosen as baseline because it was free from stimulus-evoked and eye-movement related activity by the onset of the fixation dot and the associated saccade. Baseline activity for each trial was calculated by averaging power between -1 to -0.75 s for every frequency bin on each electrode. Single-trial baseline values were then compared to each time-frequency bin during the task epoch. The sum of the signed rank difference across trials (Wilcoxon test statistic) was converted into a z-score for a standard normal distribution. The resulting time-frequency z-score maps for each electrode were averaged over sessions.

For Figs. 2A and B, the time-frequency z-score maps were averaged over all electrodes in each area and masked at a conservative threshold of $z = \pm 3.29$ (corresponding to $p < .001$, two-sided). For the array topographies (Fig. 3.S1) the z-score maps were averaged over the time interval between 0-3 s and either the beta (10-35 Hz) or the gamma band (60-160 HZ). For the power spectra (Fig. 3.S2) the z-score maps were averaged over all electrodes in each area and over the time intervals between 0-1 s / 1-2 s for dlPFC and 0.1-0.3 s / 2.5-2.7 s for vlPFC. Both z-score power spectra and topographies were Bonferroni corrected for multiple comparisons (200 frequency bins, 64 electrodes). The locus of maximal power changes in vlPFC was explored by separating the broadband

gamma spectrum into four equally spaced frequency ranges between 40 and 200 Hz (41-80 Hz, 81-120 Hz, 121-160 Hz, 161-200 Hz). The average z-score over each of these frequency ranges was then tested between adjacent sub-bands with a dependent-samples t-test. The resulting r-coefficients for each session were compared against 0 with a t-test. In order to obtain a single value representing the power modulations relative to baseline in each session, as shown in Figure 3.2E, we averaged the z-scores over each frequency band, the time epoch from 0-3 s and all electrodes per area. This power modulation value was used to explore the interaction between the frequency bands across areas with a two-way, repeated-measures ANOVA.

The degree of correlation between the beta and gamma power time-courses within and across areas was tested by averaging power changes over all electrodes per area and in the respective frequency band (beta: 10-35 Hz, gamma: 60-160 Hz) and calculating the Pearson-correlation over the time interval between 0-3 s after sample onset.

*Category information in LFP power*

We assessed category selectivity in LFP power in vlPFC and dlPFC using a percentage of explained variance statistic ($\omega^2$-statistic). The $\omega^2$-statistic reflects how much variance in the LFP signal can be explained by the category membership of a particular presented exemplar in each trial (Eqn. 2, where $SS_{between}$ is the sum of squared residuals between categories $SS_{between} = \sum_{category} N_{category} * (mean_{category} - mean_{total})^2$, $SS_{total}$ is the total sum of squared residuals across all trials $SS_{total} = \sum_{trials} (x_{exemplar} - mean_{total})^2$, $df_1$ is the degrees of

freedom between categories (i.e: 1, number of categories – 1), MSE is the mean squared error MSE = $1/df_2 * \sum_{trials} (xx_{exemplar} - \bar{x}\ mean_{category})^2$, where $df_2$ is the degrees of freedom of the error (i.e: number of trials – number of categories).

**Eqn. 3.2:** $\omega^2 = (SS_{between} - df_1 * MSE) / (SS_{total} + MSE)$

$$\omega^2 = \frac{SS_{between} - df1 * MSE}{SS_{total} + MSE}$$

$\omega^2$ is an unbiased measure of explained variance (Olejnik et al., 2003) and results in a zero-mean statistic when there is no category information (see baseline interval from -0.5 to 0 s in Figures 3.2C & D, 3.2H, 4C & D, 3.S8).

Category information expressed in the $\omega^2$-statistic was calculated for each recording session, in each of which a new set of categories was presented, and then averaged over sessions. A permutation test was used to determine significant category information ($\omega^2$) in the LFP-signal. To this end, the association between neural activity and category membership was broken up by randomly shuffling the category labels across trials. The $\omega^2$-statistic was recorded after each permutation run, generating a reference distribution of $\omega^2$-statistics under the null hypothesis of no category information in the LFP signal (approximated with a Monte Carlo procedure of 1000 permutations). Akin to the analysis of the observed data, the reference distributions were generated for each session separately and subsequently averaged over sessions. The observed $\omega^2$-statistic was then compared with this null distribution. A given electrode, time- or frequency sample was defined as carrying category information, if its associated, observed $\omega^2$-statistic exceeded

the 99.9% - quantile of the corresponding reference distribution ($p < .001$). For example, we determined the percentage of category-informative electrodes (see Fig. 3.S3) in each area (vlPFC, dlPFC) for each frequency band (beta: 10-35 Hz; gamma: 60-160 Hz). To this end we averaged power across the respective frequency band and the time interval between 0.5 to 1.5 s after sample onset and calculated for each electrode the $\omega^2$-statistic (and its corresponding null-distribution). An electrode was then defined as carrying category information, if its associated, observed $\omega^2$-statistic exceeded the 99.9% - quantile of the corresponding reference distribution.

Figs. 3.2C and D show the time-frequency maps averaged over all category-informative electrodes in either frequency band in each area and masked at a conservative threshold of $p < .001$. Fig. 3.S4 shows category information as a function of frequency averaged over all category-informative electrodes and either the sample (0-1 s) or delay epoch (1-2 s). Significant frequency ranges were defined for observed $\omega^2$-statistics with $p < .001$. The locus of maximal category information in vlPFC was explored by separating the broadband gamma spectrum into four equally spaced frequency ranges between 40 and 200 Hz (41-80 Hz, 81-120 Hz, 121-160 Hz, 161-200 Hz). The average $\omega^2$-statistic over each of these frequency ranges was then tested between adjacent sub-bands with a dependent-samples t-test. For each array (vlPFC, dlPFC), we calculated a single value representing the category information in each frequency band (beta, gamma, averaged between 0.5-1.5 s, see above). We averaged the $\omega^2$-statistics for each frequency band across all electrodes per area and converted it into a z-score for each session relative to

the session's permutation distribution (i.e. subtracting the mean of the distribution and dividing by its standard deviation, see Fig. 3.2F). This category information value was used a) to assess the amount of category information per frequency band and cortical area and b) to explore their interaction with a two-way, repeated-measures ANOVA and post-hoc dependent-samples t-tests.

### *Evoked activity*

The evoked potential was calculated as the trial-average LFP signal between 1-15 Hz (see above). The evoked potentials for each electrode were baseline corrected for the average amplitude in the baseline epoch (-1 to -0.75 s before sample onset) and then averaged over all electrodes per area. The absolute amplitude time-locked to sample onset (averaged between 0.1 to 0.3 s after sample onset) was compared between areas (vlPFC, dlPFC) with a dependent-samples t-test (Fig. 3.2G). Category information in the evoked activity ($\omega^2$-statistic) in each area was calculated for the band-bass filtered single trial data (averaged over all electrodes) and its statistical significance assessed with a permutation test (as described above for LFP power). The information time courses were smoothed with a Gaussian filter (width: 100 ms, sigma: 25 ms) for better illustration (Fig. 3.2H). The results were identical between raw and smoothed time courses. Significant differences in information between areas were tested independently for the sample (0-1 s) and memory delay epoch (1-2 s) with a dependent-samples t-test for the average $\omega^2$-statistic over the respective time interval.

*Correlation between category performance and LFP power changes*

We tested the relationship between the behavioral preference for a particular category in a given session with the task-related change in LFP power in the beta and gamma bands. To this end, we computed an index of behavioral category preference as the difference in the proportion of correct trials (pcor) between the two categories divided by the overall proportion of correct trials per session (Eqn. 3). Likewise, the LFP power differences between categories in the beta (averaged between 10-35 Hz) and gamma band (60-160 Hz) were quantified by the difference in the average LFP power (pow) between the two categories divided by the overall LFP power in that band per session (Eqn. 4).

**Eqn. 3.3**: $(pcor_A - pcor_B) / pcor_{all}$

**Eqn. 3.4**: $(pow_A - pow_B) / pow_{all}$

The performance and power differences were divided by overall performance/power, in order to bring the magnitude of the behavioral and neural effects onto the same scale. LFP power was averaged over all electrodes for each array (vlPFC, dlPFC) and the Pearson-correlation (Pearson-$r_{sessions}$) between behavioral and power differences per frequency band was calculated for each time point between 0-2 s.

In order to correct for multiple comparisons at multiple time samples, we used a nonparametric cluster-based permutation test (Maris and Oostenveld, 2007). First, clusters of temporally adjacent supra-threshold correlation (Pearson-correlation exceeding $p < .05$, two-sided) were identified. Within one cluster, r coefficients were

summed up to obtain a cluster-level test statistic. Then, random permutations of the data were drawn by exchanging the session labels and therefore breaking up the relationship between behavioral category preference and LFP power change between categories in each session. The maximum cluster level statistic was recorded after each permutation run, generating a reference distribution of cluster-level statistics (approximated with a Monte Carlo procedure of 1000 permutations). Cluster-level p-values were then estimated as the proportion of values in the corresponding reference distribution exceeding the cluster-level statistic obtained in the actual data. The cluster-level statistic represents the significant correlation over a time interval, which is effectively controlled for multiple comparisons at multiple time samples (see Figs. 3.S5 and 3.S7).

Figures 3.3C and E show the power differences between categories, averaged over these significant time intervals, plotted against behavioral category preference. We estimated the percentage of electrodes in each array with a significant correlation during those significant time intervals. To this end, the LFP power change was averaged within the time intervals of interest (0.23-1.17 s for dlPFC beta; 0.1-0.3 s for vlPFC gamma) and the Pearson-correlation with behavioral preference was calculated for each electrode. The cluster-based permutation procedure (see above) was used for multiple comparison correction at multiple electrodes (Fig. 3.S6). In order to form sensor clusters, the electrode neighborhood on the array was defined by Delaunay triangulation. We tested the difference in power between preferred and non-preferred categories (defined based on behavior, see Behavioral data above) for the average power across the electrodes with a

significant correlation. Dependent-samples t-statistics ($p < .05$, two-sided) were calculated for the beta and gamma power difference between preferred and non-preferred categories for the time course between 0-2 s and the cluster-based permutation procedure (see above) was used for multiple comparison correction at multiple time samples (Figs. 3.3D and F).

*Category information as a function of exemplar distortion*

In order to compare category information across different levels of exemplar distortion, we split the data sets for each recording day at the critical distortion level (1.1° DVA, based on the inflection point for behavioral performance on the preferred category, see Behavioral data). We calculated the $\omega^2$-statistic separately for each sub-sets of trials containing either exemplars on a low (< inflection point) or high distortion level (>= inflection point). Although the mean of $\omega^2$ is unbiased (around zero when there is no information), the distribution of observed values still varies with the number of observations (the skew of the distribution). Therefore, the data sub-sets (low vs. high distortion) were balanced in trial numbers for each session (by drawing a random sub-sample of trials equal to the minimum trial number across conditions, see section Recordings above).

The reduced number of trials for this part of the analysis (about one third less) led to less statistical power. For this reason, we restricted our analyses to the top 10% of electrodes that carried the most category information between 0.5 to 1.5 s after sample onset (see green dots in Fig. 3.S3). Category information ($\omega^2$-statistic) was averaged between the

balanced low and high distortion data sub-sets and then the six most informative electrodes in each area and each frequency band were selected. This ensured that we simultaneously captured effects in the sample and delay epoch on those informative electrodes without any bias to either low or high distortion levels. Category information over time in each data sub-set (low and high distortion levels) was determined with a permutation test (as described above in the section Category information in LFP power). The information time courses were smoothed with a Gaussian filter (width: 100 ms, sigma: 25 ms) for better illustration (Figs. 3.4C & D, 3.S8). The results were identical between raw and smoothed time courses. Significant differences in information between low and high distortion levels were tested independently for the sample (0-1 s) and memory delay epoch (1-2 s) with a dependent-samples t-test for the average $\omega^2$-statistic over the respective time interval. The interaction between low/high distortion levels and each frequency band (beta band in dlPFC, gamma band in vlPFC) was tested with a two-way repeated-measures ANOVA for the average category information between 0-2 s.

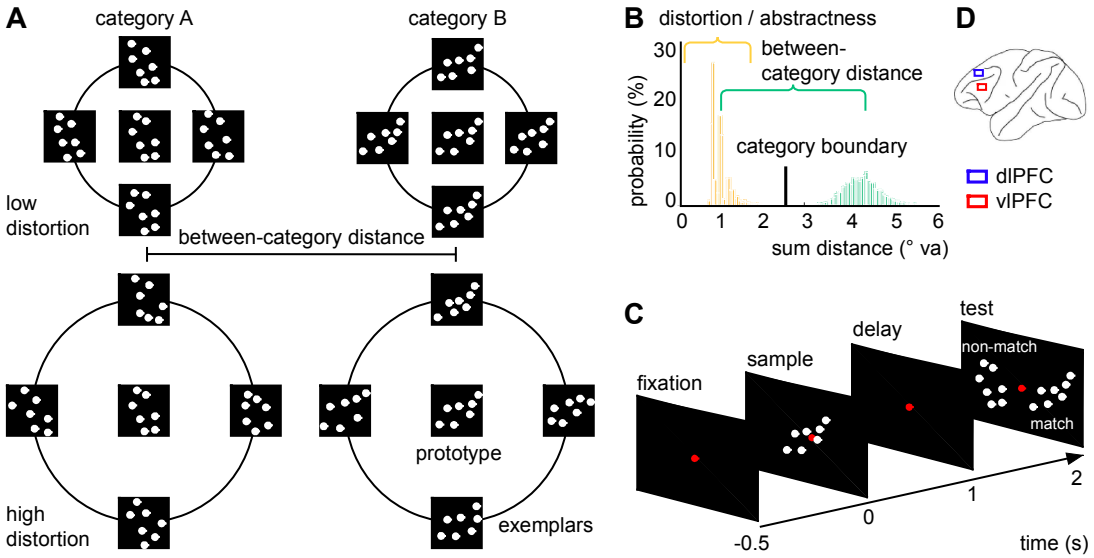


**Figure 3.1** Dot-pattern category stimuli, task and recording locations. **A.** Two dot-pattern categories (left vs right) on low and high distortion (abstractness) levels (up vs down). **B.** Summed Euclidean distance (distortion) in degrees of visual angle between exemplars and prototypes (distance to same category in yellow; distance to other category in green). **C.** Trial sequence of the delayed match-to-category paradigm. **D.** Multi-electrode array locations in dlPFC (blue) and vlPFC (red).

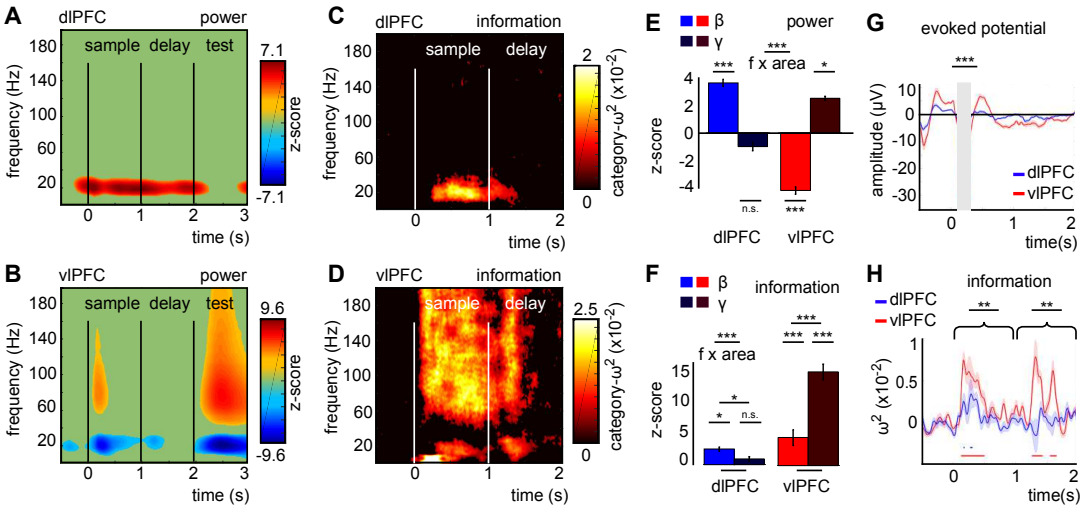


**Figure 3.2** LFP-power, category information and evoked activity. **A, B.** Power change (*z*-score) relative to baseline as a function of frequency and time in dlPFC (A) and vlPFC (B). Only z-scores with $p < .001$ are shown. These effects appeared in a large proportion of electrodes (Fig. S1). **C, D.** Category information ($\omega^2$) in power as a function of frequency and time in dlPFC (C) and vlPFC (D). Only $\omega^2$ with $p < .001$ are shown. **E, F.** Power change (E) and information *z*-scores (F) for the beta (light hue) and gamma band (dark hue) in dlPFC (blue) and vlPFC (red). Error bars show +/-1 *SE*. Asterisks indicate the significance level for each *z*-score (vs baseline (E) or random permutations (F)), between the *z*-scores per frequency in each area (F, beta vs gamma dlPFC: $t(29) = 2.7$, $p < .013$; vlPFC: $t(29) = -4.8$, $p < 4.2 \times 10^{-5}$) and for the interaction between frequency bands and areas (with * $p < .05$, *** $p < .001$). **G, H.**Evoked potential (G) and its category information ($\omega^2$) over time (H) in dlPFC (blue) and vlPFC (red). Shaded areas show +/-1 *SE*. Horizontal lines show significant time intervals ($p < .001$). Asterisks indicate the significance level for the difference between areas averaged between 0.1-0.3 s ((G), gray area) and 0-1 s or 1-2 s ((H), with ** $p < .01$, *** $p < .001$).

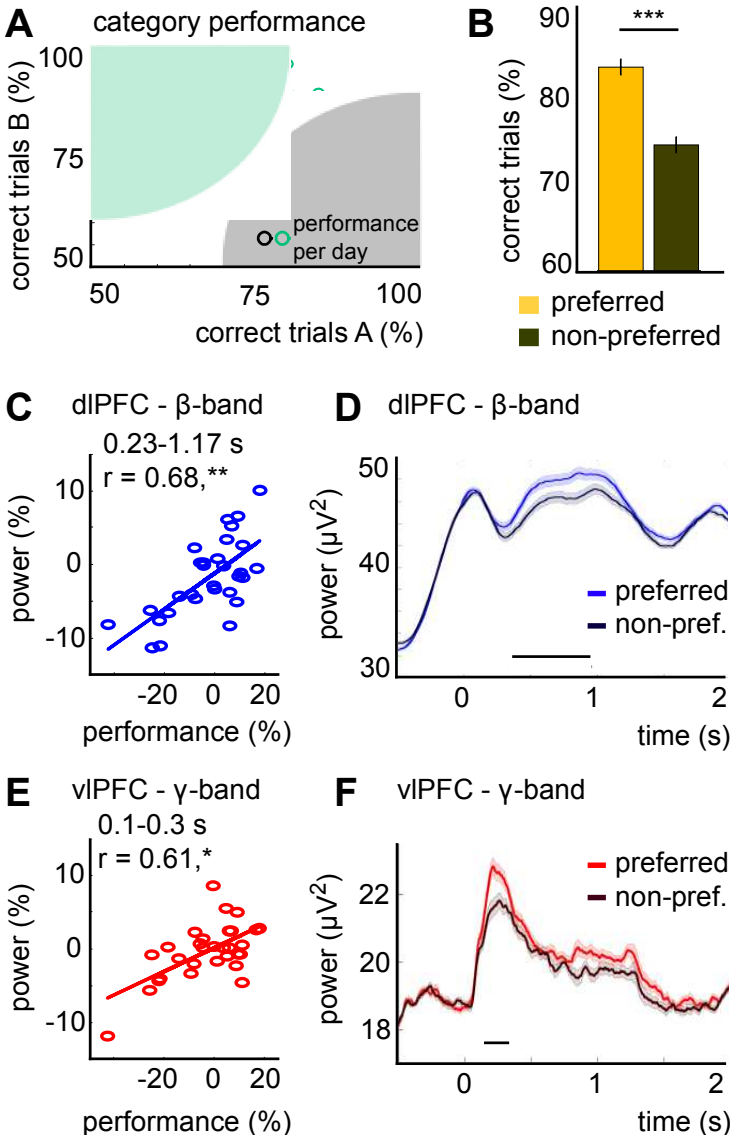


**Figure 3.3** Category preference in behavior and LFP-power. **A.** Performance (percent correct trials) per category for each recording day (circles). One category was preferred on each day (A in black or B in green). **B.** Performance for preferred (yellow) and non-preferred categories (brown). Error bars show +/-1 *SE*. Asterisks indicate the significance level (with *** $p < .001$). **C.** dlPFC-beta power difference between categories, averaged over significant time intervals (0.23-1.17 s, Fig. S5), plotted against the performance difference per session (circles). Straight lines show the linear fit. Asterisks indicate the significance level (with * $p < .05$, ** $p < .01$). (D) dlPFC-beta power as a function of time (averaged over electrodes with a significant correlation, Fig. S6) for preferred (light hue) and non-preferred categories (dark hue). Shaded areas show +/-1 *SE*. Horizontal lines show significant time intervals ($p < .05$, time-cluster corrected). (E, F) The same as in (C, D) for gamma power in vlPFC with significant effects between 0.1-0.3 s.

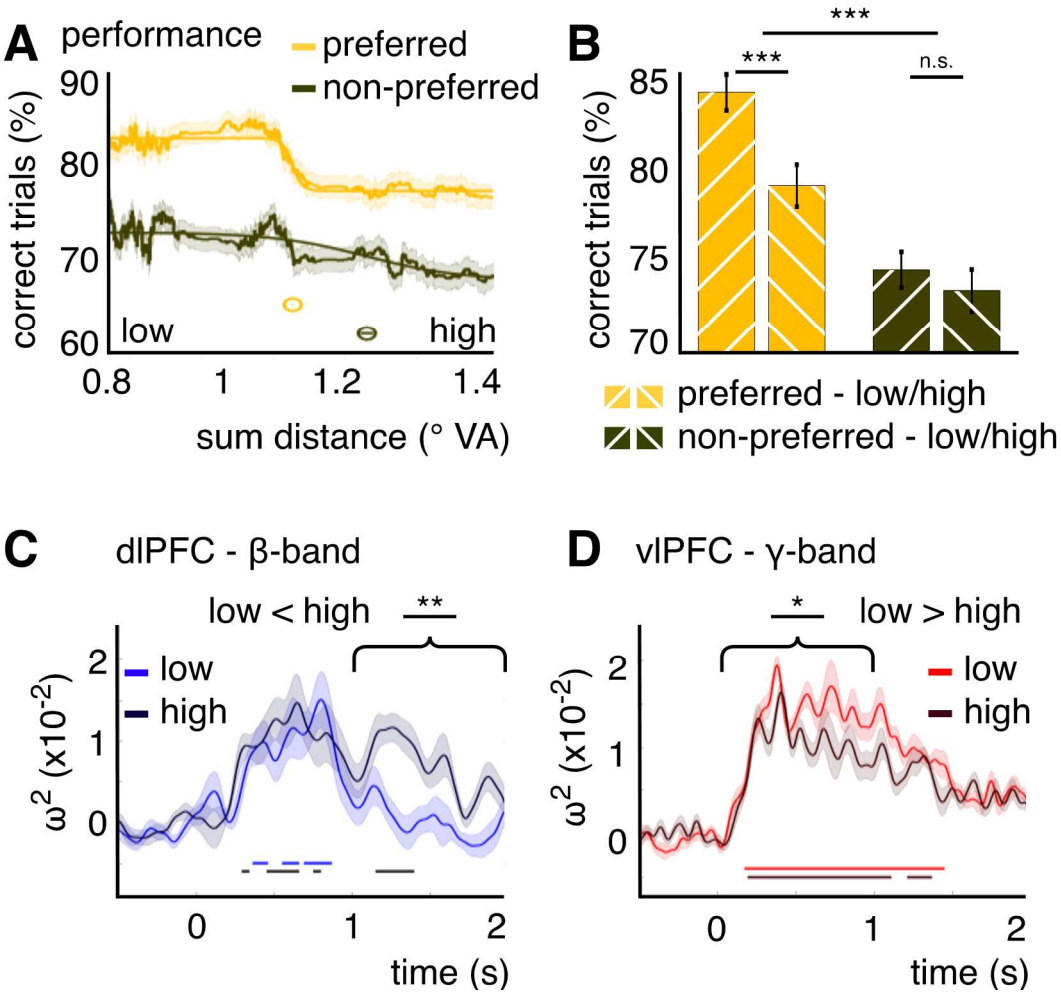


**Figure 3.4** Category abstractness in behavior and LFP-power **A.** Performance (percent correct trials) as a function of exemplar distortion for preferred (yellow) and non-preferred categories (brown). Shaded areas show +/-1 *SE*. Straight lines show the sigmoid fit and inset circles show the inflection point. **B.** Performance for preferred (yellow) and non-preferred categories (brown) for low (upward) and high distortion levels (down). Error bars show +/-1 *SE*. Asterisks indicate the significance level (with *** $p < .001$). **C,D.** Category information ( $\omega^2$) in power as a function of time for low (light hue) and high distortion levels (dark hue) in beta in dlPFC (C, blue) and in gamma in vlPFC (D, red). Shaded areas show +/-1 *SE*. Horizontal lines show time intervals with significant $\omega^2$ ($p < .001$). Asterisks indicate the significance level between low vs. high distortion levels averaged between 0-1 s (D) or 1-2 s ((C), with * $p < .05$, ** $p < .01$). There was no significant difference for dlPFC beta between 0-1 s ($t(29) = -0.2$, $p < .8$) and for vlPFC gamma between 1-2 s ($t(29) = 1.7$, $p < .11$).

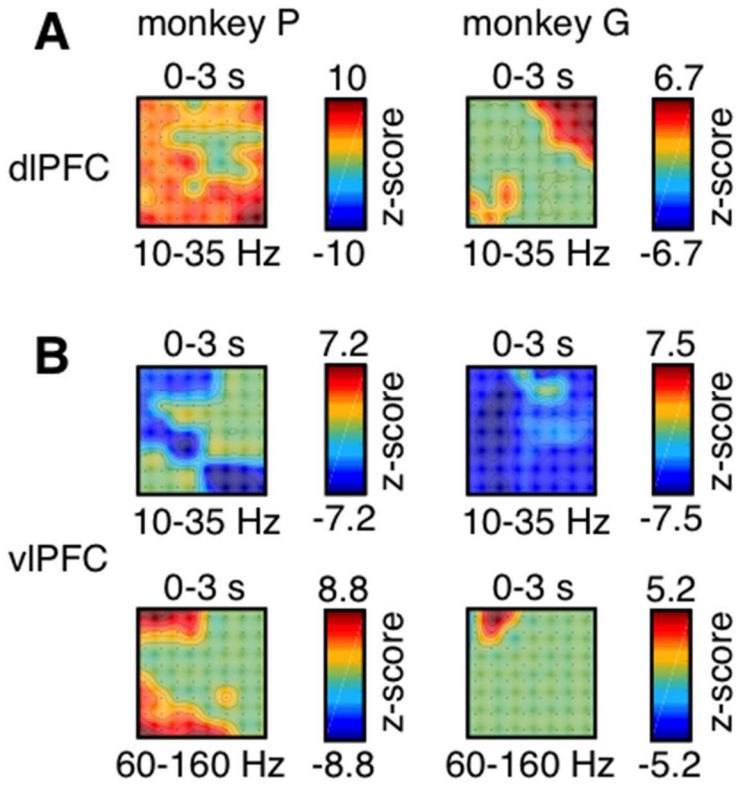


**Figure 3.S1** LFP-power topographies relative to pre-trial baseline **A,B.** Array topographies for power changes (z-score) relative to baseline averaged over the time epoch from 0-3 s after sample onset and either the beta (10-35 Hz) or gamma frequency band (60-160 Hz) for dlPFC (A) or vlPFC (B). Non-significant z-scores are masked ($p < .05$, Bonferroni corrected). For monkey P (left column) 80% of the electrodes showed an increase in the beta band in dlPFC, 42% electrodes showed a decrease in the beta band in vlPFC and 34% electrodes showed an increase in the gamma band in vlPFC. For monkey G (right column) 23% of the electrodes showed an increase in the beta band in dlPFC, 95% electrodes showed a decrease in the beta band in vlPFC and 5% electrodes showed an increase in the gamma band in vlPFC.

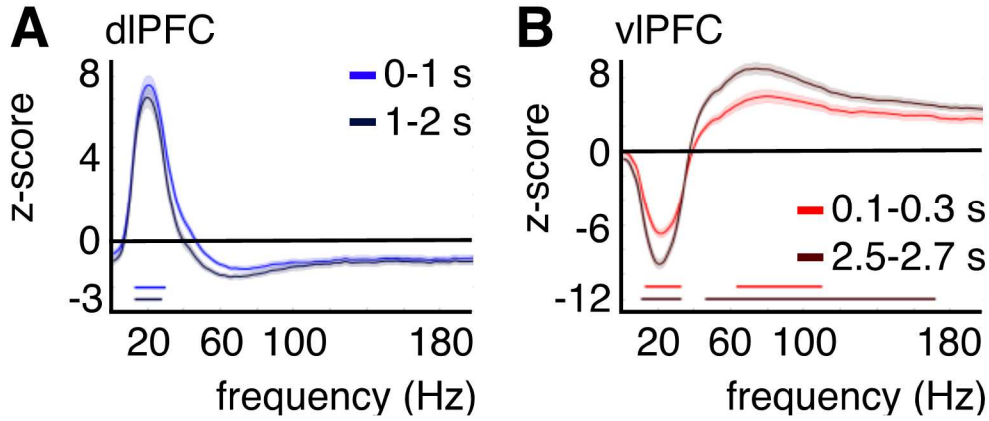


**Figure 3.S2** LFP-power over frequency relative to pre-trial baseline **A,B.** Power change (z-score) relative to baseline (-1 to -0.75 s) as a function of frequency averaged over time intervals of interest (light, dark hue) and all electrodes in dlPFC (blue, A) and vlPFC (red, B). Shaded areas show +/-1 SE. Horizontal lines show significant frequency ranges ($p < .05$, Bonferroni corrected). The change in vlPFC gamma power (B) after sample onset (0.1-0.3 s, light red) was maximal between 80-120 Hz and then decreased with increasing frequency (tested between four adjacent gamma sub-bands; 41-80 vs 81-120 Hz: $t(29) = -9.5$, $p < 2.2 \times 10^{-10}$; 81-120 vs 121-160 Hz: $t(29) = 9.7$, $p < 1.3 \times 10^{-10}$; 121-160 vs 161-200 Hz: $t(29) = 8$, $p < 7.7 \times 10^{-9}$).

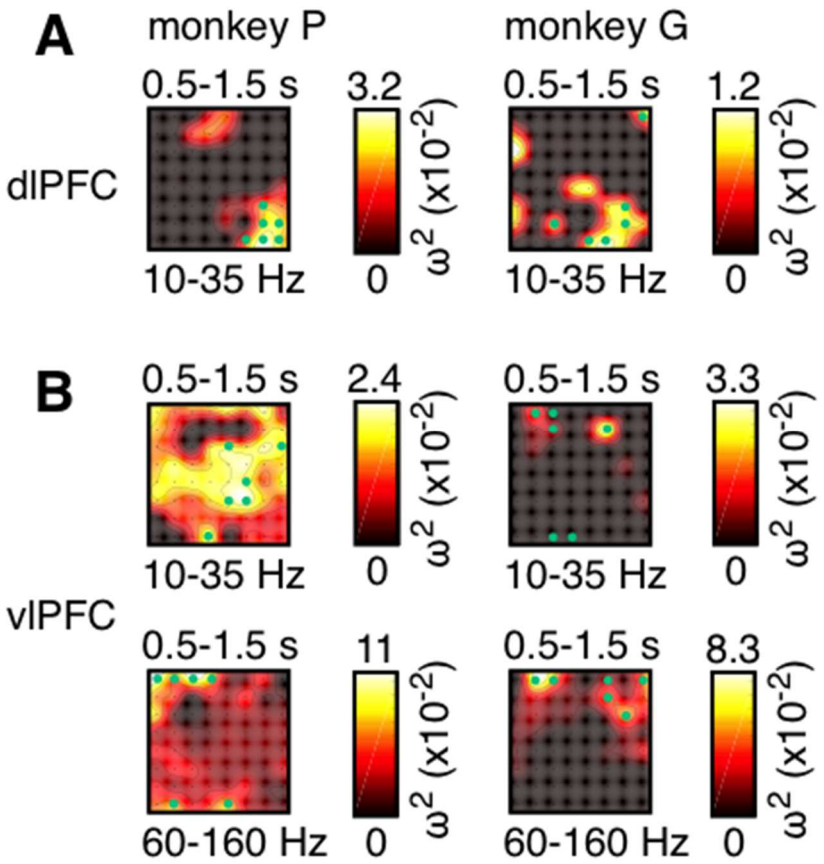


**Figure 3.S3** Electrode topographies for category information **A,B.** Array topographies for category information ($\omega^2$) in LFP-power averaged over the time epoch from 0.5-1.5 s after sample onset and either the beta (10-35 Hz) or gamma frequency band (60-160 Hz) for dlPFC (A) or vlPFC (B). Non-significant effects are masked ($p < .001$). For monkey P (left column) 25% of the electrodes showed an effect in the beta band in dlPFC, 86% electrodes showed an effect in the beta band in vlPFC and 95% electrodes showed an effect in the gamma band in vlPFC. For monkey G (right column) 25% of the electrodes showed an effect in the beta band in dlPFC, 13% electrodes showed an effect in the beta band in vlPFC and 47% electrodes showed an effect in the gamma band in vlPFC. The green dots show the 6 electrodes (10%) with most category information in each frequency band and area, which were used as regions of interest for the distortion analysis.

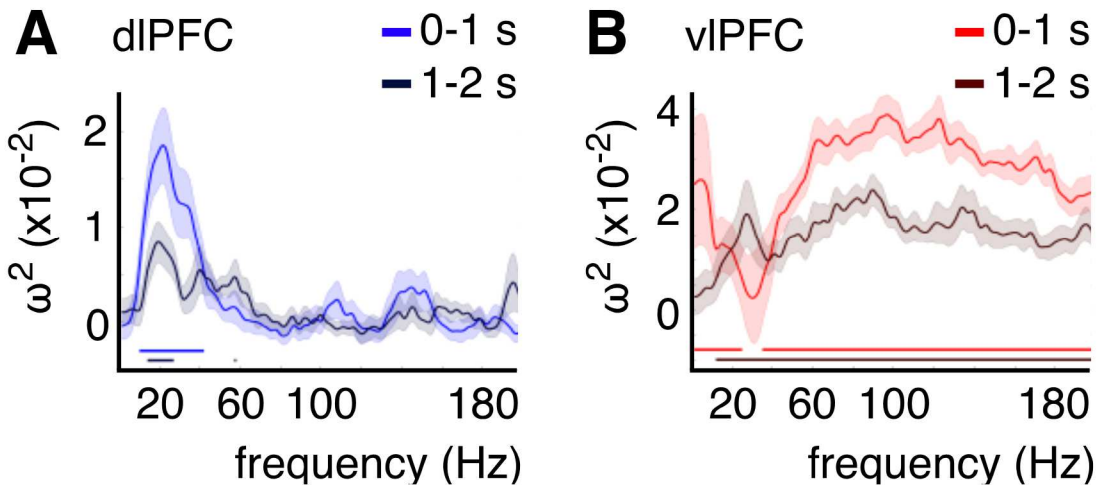


**Figure 3.S4** Category information over frequency **A,B.** Category information ($\omega^2$) in power as a function of frequency averaged over significant electrodes in dlPFC (blue, A) and vlPFC (red, B) separately for sample (0-1 s; light hue) and delay epochs (1-2 s; dark hue). Shaded areas show +/- 1 SE. Horizontal lines show significant frequency ranges ($p < .001$). Category information in vlPFC gamma power (B) during the sample epoch (0-1 s, light red) was maximal between 80-120 Hz and then decreased with increasing frequency (tested between four adjacent gamma sub-bands; 41-80 vs 81-120 Hz: $t(29) = -2.7$, $p < .012$; 81-120 vs 121-160 Hz: $t(29) = 2.4$, $p < .025$; 121-160 vs 161-200 Hz: $t(29) = 2.9$, $p < .007$).

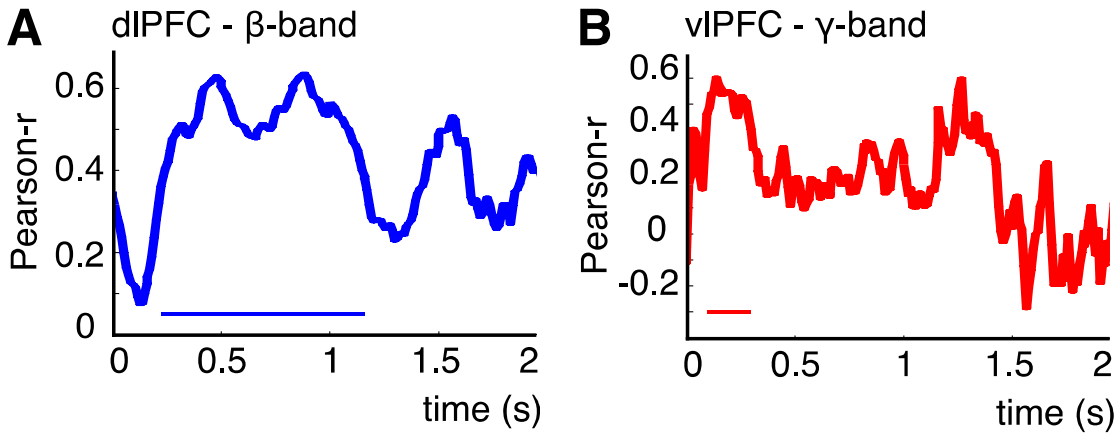


**Figure 3.S5** Correlation between behavioral and power differences between categories in the dominant frequency bands per area **A,B.** Pearson-correlation coefficient (r) over time between behavioral performance and power differences between categories averaged in the beta band (10-35 Hz) and all electrodes in dlPFC (A, blue) and the gamma band (60-160 Hz) in vlPFC (B, red). Horizontal lines show significant time intervals (time-cluster corrected; for dlPFC beta $p < .002$ for 0.23-1.17 s; for vlPFC gamma $p < .02$ for 0.1-0.3 s).

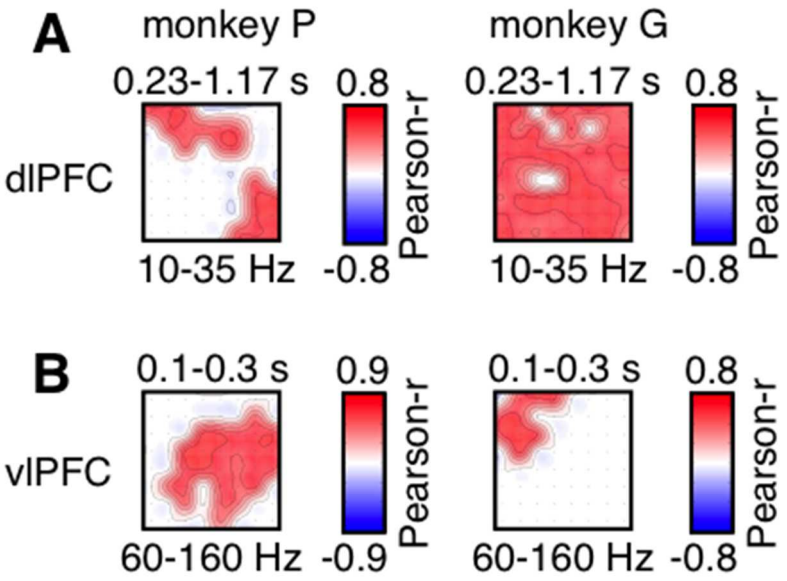


**Figure 3.S6** Electrode topographies for the correlation between behavioral and power differences between categories **A,B.** Array topographies for the correlation (Pearson-r) between behavioral and power differences between categories averaged over the beta band (10-35 Hz) in dlPFC (A) or the gamma band (60-160 Hz) in vlPFC (B) and over the time interval of interest (0.23-1.17 s for dlPFC beta; 0.1-0.3 s for vlPFC gamma). Non-significant Pearson-r coefficients are masked ($p < .05$, channel-cluster corrected). For monkey P (left column) 30% of the electrodes showed an effect in the beta band in dlPFC in two distinct clusters ($p < .018$ and $p < .028$) and 42% electrodes showed an effect in the gamma band in vlPFC ($p < .006$). For monkey G (right column) 92% of the electrodes showed an effect in the beta band in dlPFC ($p < .002$) and 14% electrodes showed an effect in the gamma band in vlPFC ($p < .006$).

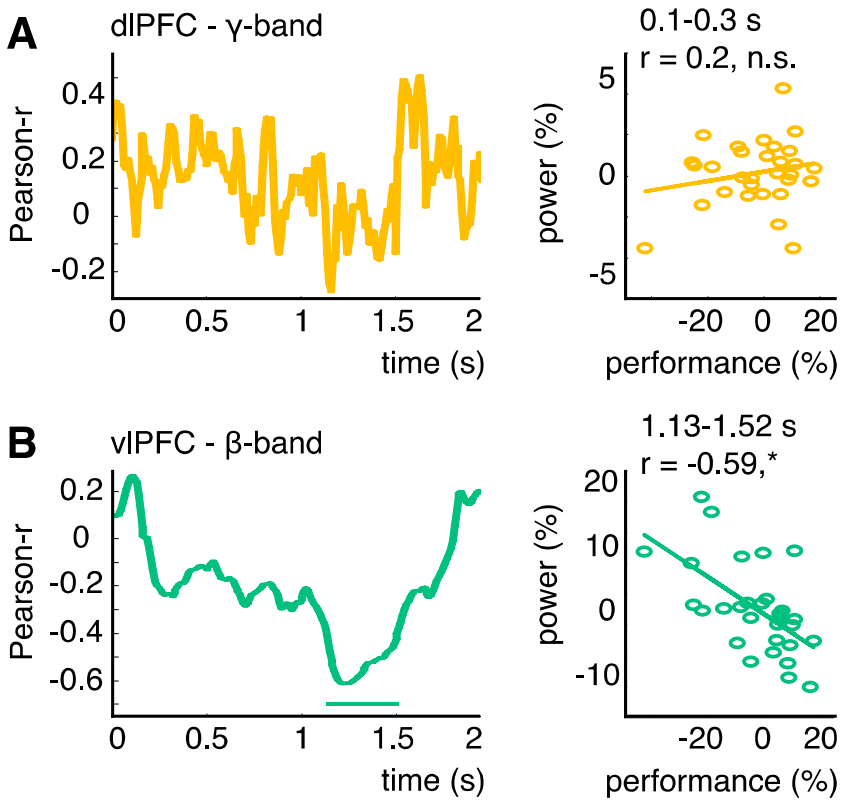


**Figure 3.S7** Correlation between behavioral and power differences between categories in the non-dominant frequency bands per area **A,B.** Left panels: Pearson-correlation coefficient (r) over time between behavioral performance and power differences between categories averaged in the gamma band (60-160 Hz) and all electrodes in dlPFC (A, yellow) and the beta band (10-35 Hz) in vlPFC (B, green). Horizontal lines show significant time intervals (time-cluster corrected; for vlPFC beta $p < .016$ for 1.13-1.52 s). Right panels: Power difference between categories, averaged over significant time intervals (see left panel), plotted against the performance difference per session (circles). Straight lines show the linear fit. Asterisks indicate the significance level (with * $p < .05$). For gamma in dlPFC (A, yellow), there was no significant correlation over time. For comparison, we show the correlation between behavioral and power differences averaged over the time interval, in which we found a significant correlation for gamma in vlPFC (0.1-0.3 s).

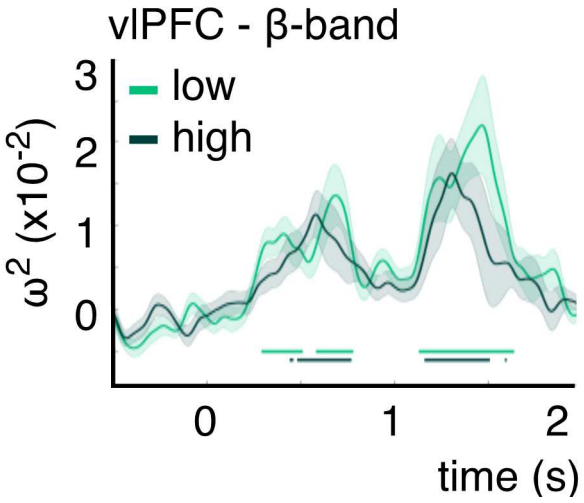


**Figure 3.S8** Category information in vlPFC beta power Category information ($\omega^2$) in power as a function of time for low (light hue) and high distortion levels (dark hue) averaged over the 10% most informative electrodes in the beta band in vlPFC. Shaded areas show +/-1 SE. Horizontal lines show time intervals with significant $\omega^2$ ($p < .001$). There was no significant difference in category information between low and high distortion levels, neither in the sample (0-1 s; $t(29) = 0.5$, $p < .62$) nor the delay epoch (1-2 s; $t(29) = 0.9$, $p < .4$).

## Bibliography

Antzoulatos, E.G., and Miller, E.K. (2014). Increases in functional connectivity between prefrontal cortex and striatum during category learning. Neuron *83*, 216–225.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., de Weerd, P., Kennedy, H., and Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron *85*, 390–401.

Brainard, D.H. (1997). The psychophysics toolbox. Spatial Vision *10*, 433–436.

Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., and Miller, E.K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron *76*, 838–846.

Cromer, J.A., Roy, J.E., and Miller, E.K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. Neuron *66*, 796–807.

Engel, A.K., and Fries, P. (2010). Beta-band oscillations — signalling the status quo? Current Opinion in Neurobiology *20*, 156–165.

Fries, P., Reynolds, J.H., Rorie, A.E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science *291*, 1560–1563.

Gastgeb, H.Z., Dundas, E.M., Minshew, N.J., and Strauss, M.S. (2012). Category formation in autism: can individuals with autism form categories and prototypes of dot patterns? Journal of Autism and Developmental Disorders *42*, 1694–1704.

Jensen, O., Bonnefond, M., Marshall, T.R., and Tiesinga, P. (2015). Oscillatory mechanisms of feedforward and feedback visual processing. Trends in Neurosciences *38*, 192–194.

Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., and Miller, E.K. (2016). Gamma and beta bursts underlie working memory. Neuron *90*, 152–164.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods *164*, 177–190.

Morey, R. (2008) Confidence intervals from normalized data: a correction to Cousineau (2005). Tutorial in Quantitative Methods for Psychology *4,* 61-64.

Olejnik, S., and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological Methods *8*, 434–447.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2010). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience *2011*, e156869.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision *10*, 437–442.

Posner, M.I., and Keele, S.W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology *77*, 353–363.

Uhlhaas, P.J., and Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. Nature Reviews Neuroscience *11*, 100–113.

Vogels, R., Sary, G., Dupont, P., and Orban, G.A. (2002). Human brain regions involved in visual categorization. NeuroImage *16*, 401–414.

Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956.

# CHAPTER 4: LAMINAR-SPECIFIC ACTIVITY IN FRONTAL CORTEX SUGGESTS MECHANISMS FOR CONTROL OF WORKING MEMORY

André M. Bastos*, Roman Loonis*, Simon Kornblith, Michael Lundqvist, Earl K. Miller#

*Equal contributions
# - To whom correspondence should be addressed

The Picower Institute for Learning & Memory and Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139, USA.

## Abstract

All of the cerebral cortex has some degree of laminar organization. These different layers are composed of neurons with distinct connectivity patterns, embryonic origins, and molecular profiles. But there is little data on the laminar specificity of cognitive functions in frontal cortex. We recorded neuronal spiking/local field potentials (LFPs) using laminar probes in frontal cortex (PMd, SEF, SMA, ACC, 46d/v, 8) of monkeys performing working memory (WM) tasks. LFP power in gamma (50-250 Hz) was strongest in superficial layers and in alpha/beta (4-22 Hz) in deep layers. Memory-delay activity, including spiking and stimulus-specific gamma bursting, was predominately in superficial layers. LFPs from superficial and deep layers synchronized in the alpha/beta bands. This was primarily unidirectional with alpha/beta in deep layers driving superficial-layer activity. The phase of deep-layer alpha/beta modulated superficial gamma bursting associated with working memory encoding. Thus, alpha/beta

rhythms in deep layers may regulate the superficial-layer gamma and hence the maintenance of the contents of working memory.

## Introduction

Working memory has long been associated with neural activity bridging a memory delay. This has been proposed to be due to recurrent connections between columns of pyramidal neurons in superficial cortical layers (Goldman-Rakic et al., 1996). Support for this has been mixed. One study found that delay activity was shared across superficial and deep layer neurons (Sawaguchi et al., 1990). However, another reported delay-activity neurons ("late storage units") at more superficial depths (Markowitz et al., 2015). This uncertainty may be due, in part, to the prior use of single-contact electrodes, which render it difficult to assess the depth of the recorded signals.

Another related question is whether the frontal cortex would show similar layer-specific properties as those observed in visual cortex (Bollimunta et al., 2011; Buffalo et al., 2011; van Kerkoerle et al., 2014; Maier et al., 2010). In visual cortex, multiple-contact "laminar" electrodes have revealed that gamma (> 30 Hz) oscillations are more prominent in superficial and middle layers while slower oscillations (alpha/beta, 10-30 Hz) are prominent in deep layers (Maier et al., 2010; Buffalo et al., 2011). Moreover, it has been shown that deep layer alpha activity drives superficial alpha activity (Bollimunta et al., 2011; Kerkoerle et al., 2014), and the phase of deep layer alpha modulates the amplitude of the superficial layer gamma (Spaak et al., 2012). Similar tests

in one frontal area, the supplemental eye fields, have reported gamma LFP power in superficial layers (Godlove et al., 2014; Ninomiya, 2012). But one of these studies failed to find evidence that deep-layer low frequency oscillations coupled with superficial gamma (Ninomiya et al., 2012), leading to the conclusion that frontal cortex laminar dynamics might fundamentally differ from those of the rest of cortex. Neither study has examined working memory-related activity in frontal cortex with laminar electrodes.

To clarify this, we recorded both spiking and local field potential (LFP) activity with multi-laminar electrodes in five frontal cortex areas (PMd, SEF, SMA, ACC, 46d/v, 8) in three WM tasks. This revealed layer-specific dynamics in frontal cortex that were similar to those in visual cortex, suggesting a layer-specific micro-circuit motif that is common across cortex. Further, examining this in the context of a working memory task provided evidence that these dynamics may be used to control access to working memory.

## Results

*Gamma power peaks in superficial layers, and alpha/beta peaks in deep layers*

Three monkeys performed three different working memory tasks (Fig. 4.1A-C). During these tasks, laminar probes with 100 to 200 μm spacing between contacts recorded local field potentials (LFPs) and neuronal spiking from all cortical layers (Figure 4.1D-F). These electrodes were lowered as perpendicular as possible to the cortex to ensure an even sampling of the different layers. We completed a total of 60 recordings in frontal cortex (Fig. 4.1G). The middle cortical layer was identified, using current source density

(CSD) analysis, by the presence of a current sink in response to the presentation of a visual stimulus (see Methods). We aligned all of the data from all electrodes to the middle layer (i.e., the contact with the first significant CSD sink; see Figure 4.S1 for the average CSD profile), and pooled all of the data together.

We found that high frequency oscillations peaked in superficial layers and low frequencies peaked in deep layers. Figure 4.1H plots examples of low- and high-passed LFPs from one laminar recording, and Fig 4.1I is a power spectrum, illustrating how low frequency (alpha/beta) vs high frequency (gamma) power became more apparent in deep vs superficial layers, respectively. To quantify these changes in power across each laminar probe, we normalized power at each frequency and each contact (1-500 Hz) by the maximal power at that frequency across contacts. In other words, for each frequency and each session, the contact with maximal power had a value of 1 and other contacts had values relative to this maximum. For example, 0.8 means that the power at that contact was 80% of the maximal power observed at that frequency.

Figure 4.2A shows the mean power profile across all electrodes. The middle cortical layer (bottom of layer 3/layer 4) is at depth 0, negative depths are superficial layers (layers 1-3), and positive depths are deep layers (layers 5-6). At each frequency and cortical depth, red colors indicate the maximal power and blue the minimal. The superimposed black line represents the mean distance from the middle of the cortex at which the maximal power occurred. It shows that lower frequencies (4 to 22 Hz) had

their maximal power below the sink (contact at depth zero), while a continuous band of higher frequencies (58-260 Hz) had their maximal power above the sink (sign test across sessions, Bonferroni corrected, $p<0.05$).

To further illustrate this, we collapsed Figure 4.2A into two separate profiles by averaging across the alpha/beta (4-22 Hz) and gamma (58-260 Hz) frequency bands (Figure 4.2B). The peak gamma power (blue line) occurred in superficial layers, 400 um above the sink, and the peak beta power (red line) occurred in deep layers, 800 um below the sink. The cross-over point between the profiles (the intersection of the blue and red lines) occurred between -100 and -200um, nearly identical to the location of the CSD sink. Thus, gamma power was prominent in superficial layers, and alpha/beta in deep. In middle layers (from 200um above to 300 um below the sink) there was a transition zone where neither gamma nor beta predominated. These results were present in each of the tasks (Figure 4.2C-E).

*Delay period multi-unit activity is modulated in superficial layers*

Working memory has long been correlated with the persistent modulation of delay-period spiking activity in frontal cortex, but whether or not this activity is laminar specific is unclear. We found that delay period multi-unit activity was largely localized to superficial layers. To measure this multi-unit activity, we used rectified, high-passed signals greater than 500 Hz. To assess the delay period modulation, we took the absolute value of the mean change in the multi-unit activity (MUA) between the delay and the

baseline, and z-scored it by the standard deviation of delay-period MUA across trials (see Methods). This normalization step ensured that differences in the overall level of MUA activity were de-emphasized and, instead, each session contributed with equal weight after pooling. We used the absolute values because activity in working memory delays could increase or decrease relative to baseline (Miller et al., 1996)

Figure 4.3 A-C shows the mean delay period MUA modulation averaged across electrodes and sessions for each task. Red colors indicate a change in MUA from baseline; blue colors indicate no or little change. In all three tasks, the largest change in MUA occurred in superficial layers. Figure 4.3 D-F illustrates the average delay period modulation, and demonstrates that the greatest modulations occurred in the superficial layers. Averaging across all three tasks, Figure 4.3G shows that delay MUA peaked 400um above the sink and dropped in deeper cortical layers (Fig. 4.3G, $P<0.002$, sign test across sessions). Moreover, the average modulation of delay-period MUA across layers was positively correlated (Spearman rank correlation, $R=0.84$, $p=2E-6$) with gamma power (compare to Figure 4.2B) and negatively correlated with beta power (Spearman rank correlation, $R=-0.67$, $P=3E-4$). It was also seen in each task individually (Spearman rank correlation, Visual Search, gamma: rho = 0.79, $p<1E-5$, alpha/beta: rho = -0.57, $p=0.003$; Masked Delayed Saccade, gamma: rho = 0.80, $p<1E-5$, alpha/beta: rho = -0.66, $p<0.001$; Delayed Saccade, gamma: rho = 0.86, $p<1E-5$, alpha/beta: -0.68, $p<0.001$).

The rectified, high-passed signal we used to measure MUA was advantageous because it is thought to capture the mean of all spiking activity within the local vicinity of the recording contact (Siegel et al., 2015). However, some spectral overlap between it and the gamma power was seen in superficial layers, raising the possibility that the greater MUA was merely due to the higher gamma power (or vice-versa). To address this, we used a thresholded signal to measure the spike rate. This signal was more conservative than the MUA signal, capturing only small groups of units with large spikes near the contact (henceforth referred to as "units"). Using this technique, we identified a total of 420 units, and confirmed that delay-period modulation of spiking is higher in superficial layers, consistent with the analysis based on the more inclusive rectified, high-passed MUA signal.

Figure 4.4A shows the number of units recorded at each depth (in green), summed across sessions, and the number of units that were modulated (in blue). Modulation was determined by testing whether each unit's firing rate during the delay was significantly different from its baseline firing rate by a two-tailed t-test at $p<0.01$. Figure 4.4B shows the proportion of units with delay activity (i.e., units modulated/units recorded), which was significantly different between superficial vs deep layers (Figure 4.4B inset, 53% vs. 34%, Chi-squared proportion test, p=3E-4). The proportion of modulated units by layer (Figure 4.4B) positively correlated with the gamma LFP profile and negatively correlated with the beta LFP profile (Spearman rank correlation, gamma, rho = 0.83, p<1E-6, alpha/beta, rho = -0.60, p=0.001). Finally, the greater proportion of modulated units in

superficial layers was not the result of a poor signal to noise ratio or lack of units in deep layers. In fact, baseline firing rates were higher in the deep layers (Figure 4.4C), as was the unit yield (Figure 4.4D).

*Gamma bursts in superficial layers encode stimulus information during the delay*

Recent work has shown that gamma bursting in the PFC is associated with the encoding of stimulus information in working memory (Lundqvist et al., 2016). We tested for its layer-specificity. To distinguish this from the power analyses described above, we defined bursts as periods when the power in the alpha/beta (4-22Hz) and gamma (50-150 Hz) bands exceeded the mean power at each frequency band by 2 SDs for three cycles (see Methods). The gamma and beta burst rate during the delay and baseline periods matched the power profiles at their respective frequencies (Figure 4.5A, B). They were positively correlated to their respective power profiles (Spearman rank correlation, gamma: Rho=0.76, p<2E-5, beta: Rho: 0.8, P<4E-6). The average gamma burst rate increased during the delay relative to the baseline (Figure 4.5A, t-test, p=0.002), and the beta burst rate decreased (Figure 4.5C, t-test, p<2E-9).

We tested whether delay-period gamma and beta bursts carried information about which cue was held in WM by calculating the PEV between the burst rate and cued object/location during the delay (see Methods). Figure 4.5B shows the profile of information in gamma-bursts by layer, and 4.5D for beta-bursts. The red lines to the right of each plot represents statistical significance at $p < .01$, uncorrected for multiple

comparisons. Over sessions, gamma bursting in superficial layers was more informative than in deep ($p=0.018$, sign test over sessions, inset of Figure 4.5D). Furthermore, the amount of gamma-bursting information per layer was strongly correlated to gamma power (Spearman rank correlation, $rho = 0.95$, $p<5E-6$). Information in beta-bursting was weaker, as might be expected given its very low burst rate during the delay. Information in beta-bursting was not significantly different between deep and superficial layers ($p=0.52$, sign test over sessions), but trended towards an increase in deep layers (Figure 4.5D). The amount of information in beta bursting strongly correlated with the beta power profile (Spearman rank correlation, $rho = 0.84$, $p<5E-6$).

*Alpha/beta oscillations in deep layers modulate superficial layers*

We next tested whether there were interactions between oscillatory activity and layers. To assess directed interactions, we applied nonparametric Granger causality (GC) analysis to LFPs within supragranular and infragranular layers. This analysis measures the extent to which the past values of time-series A can predict the present state of time-series B, above and beyond what the past values of B already predict about its present state. We performed this for every frequency using the phase and amplitude of past values to assess predictions about the present.

We found that the GC spectrum had peaks in the alpha-beta range, and that directed interactions were asymmetric. Deep layer alpha-beta drove superficial layer alpha-beta more than the other way around (4-22 Hz, $p=0.002$, sign test, Figure 4.6A). We also

tested whether GC influences were stronger in the delay period compared to baseline. We found no significant differences in GC influences in the delay vs. baseline contrast ($p > 0.2$).

GC is a linear measurement of inter-laminar interactions, in the sense that interactions are only tested between different channels at the same frequency. To test for cross-frequency interactions, we next investigated whether there was Phase-Amplitude Coupling (PAC) between superficial and deep layers. To test whether the phase of the slower frequency band (alpha/beta band) coupled with the amplitude of the higher frequency band (gamma band) we used the modulation index (Tort et al., 2008). The modulation index is a measure of how non-uniformly distributed the amplitudes of one frequency band are across the phase space of another. We systematically calculated PAC at every possible combination of (alpha/beta) phase-providing channel and (gamma) amplitude providing channel. We found that deep alpha/beta phase modulates superficial gamma amplitude and that this ascending (deep to superficial) influence was stronger than in the reverse direction (Figure 4.6B).

Figure 4.6C plots the modulation index value for each of the possible laminar interactions: superficial-deep, deep-superficial, deep-deep, and superficial-superficial. The solid black lines indicate significant differences in phase amplitude coupling between each condition (t-test, $p < .001$). We found that the modulation of superficial gamma was greatest by deep and superficial alpha/beta (Fig. 4.6C). In contrast, coupling

between deep gamma and both superficial and deep alpha/beta was significantly reduced (Fig. 4.6C). Moreover, PAC was significantly lower during the delay, relative to baseline (sign test over sessions, $p<0.0005$; Figure 4.6D, PAC during delay is shown in Figure 4.S2). Taken together, these results suggest that deep-layer alpha-beta regulated superficial alpha-beta which, in turn, regulated superficial layer gamma.

## Discussion

### *Shared Functional Motifs Across Cortex*

Superficial and deep layers of frontal cortex exhibited distinct dynamics. Gamma power peaked in superficial layers while alpha/beta power peaked in deep layers. Deep layer alpha/beta oscillations drove superficial alpha/beta. The phase of these deep-layer alpha/beta oscillations modulated the amplitude of superficial gamma. All of these dynamics were shared across a broad range of areas spanning from premotor to prefrontal cortex in our dataset and match closely with results from visual cortex (Bollimunta et al., 2011; Buffalo et al., 2011; Spaak et al., 2013; van Kerkoerle et al., 2014). The global consistency and the specificity of these physiological effects argue in favor neuronal dynamics that are shared in many regions (Bastos et al., 2012; Douglas and Martin, 1991).

*Delay activity in superficial layers*

Delay-period activity, both spiking and gamma bursting, was most prominent in superficial layers. The co-occurrence of these two phenomena is consistent with reports that gamma bursts are associated with spiking that encodes stimulus information in working memory (Lundqvist et al., 2016). Indeed, those delay-period gamma bursts we observed also carried stimulus-related information within superficial layers. The broad-band nature of the average power spectrum in the gamma range (which lacked a clear peak) does not necessarily imply the lack of an oscillatory phenomenon. Averaging of gamma bursts of varying frequency across individual trials could lead to this broad-band appearance (Buzsáki et al., 2012; Lundqvist et al, 2016).

During the working memory delay, both alpha/beta bursts and coupling between deep-layer alpha/beta and superficial layer oscillations decreased relative to baseline. Alpha/beta oscillations are purported to be an inhibitory rhythm responsible for suppressing behavioral dominant rules and disregarding distracting stimuli (Buschman and Miller, 2007; Buschman et al., 2012; Jensen et al., 2015; Haegens et al., 2011). Low frequency coupling between deep and superficial layers, therefore, may serve a control function, gating information into superficial layers by regulating the gamma bursting associated with working memory encoding. A decrease in coupling might release inhibition from deep to superficial layers, and allow cue information to be maintained within superficial layers. There, this information could be stored through recurrent lateral connections that result, on average, in sustained neuronal activity but within a single trial

as short-lived gamma bursts and spiking (Lunqvist et al., 2016). This is consistent with known recurrent networks within supragranular layers of prefrontal cortex (Goldman-Rakic, 1996; Luebke, 2017). Working memory activity, therefore, reflects the continuous reactivation of its contents in supragranular layers.

In Figure 4.7, we summarize this model of working memory. In this diagram, we note that both superficial and deep layers are comprised of networks of deeply interconnected excitatory pyramidal (black) neurons and inhibitory (red) interneurons. Circuits in both layers are capable of oscillating within the alpha/beta range (the red sine wave below, the blue line above), but the drive is directional. Deep layers (as seen in the red arrows) drive superficial layers to resonate within the alpha/beta frequency. These alpha/beta oscillations are coupled with superficial layer gamma oscillations. Increasing deep to superficial layer coupling and/or deep-layer alpha/beta suppresses gamma-related activity that stores bottom-up information in working memory (the small blue lines in superficial layers). This suppression in the default state allows for a number of different excitatory regimes (i.e. different networks responsible for cue information) to co-exist. However, in the memory delay, we propose that this default suppression of gamma band activity is released and, as a result, the recurrent connectivity of layer 3 neurons (as indicated by the loop arrow) is allowed to persist. This recurrent activity is not only responsible for the persistence of gamma activity, but also for the dominance of a particular excitatory circuit (i.e. one encoding the cue information).

Previously, we linked gamma-band dynamics with both feedforward and bottom-up mechanisms (Bastos et al., 2015; Buschman and Miller, 2007). In this prior work, gamma was found to signal sensory stimuli from lower to higher visual cortex (Bastos et al., 2015; Michalareas et al., 2016; van Kerkoerle et al., 2014), and to drive stimulus-driven attention (Buschman and Miller, 2007). Here we find that gamma dynamics are associated with working memory maintenance, a cognitive function. In the visual system, the function of superficial layer cells, with gamma-band dynamics, is thought to involve feedforward information transmission (Bastos et al., 2015; van Kerkoerle et al., 2014; Roberts et al 2013). In prefrontal cortex, we find a preservation of this feature of the laminar circuit (superficial layer gamma-dominated dynamics). At the highest level of the cortical hierarchy, feedforward connections are by definition undefined (Markov et al., 2013). We suggest that the preservation of these laminar patterns in PFC in the absence of further levels to the hierarchy gives rise to a new function for these superficial layers, namely, the maintenance of bottom-up stimulus information in working memory.

## Methods

### *Tasks*

In order to investigate the mechanisms underlying working memory, we trained three monkeys on three different working memory tasks. Task one (Fig 4.1A) required memory for a sample object over a variable delay (0.5-1.2s). The other two tasks engaged working memory for a spatial location; however, they differed in the presentation of a masking stimulus during the delay. In task two (Fig 4.1B), after a location was cued by a

red dot, the animal had to maintain information as a visual mask (red dots at all possible cued locations) was presented through a variable delay (2.2 to 2.7s). In the third task, the monkey had to similarly maintain one of four possible cue locations after a short sample epoch (0.295s); however, during the subsequent 0.99s delay, there was no visual mask, only the fixation point was shown during the delay. The idea in comparing these tasks was to show that independent of both the content of the working memory and the presence or absence of distractors during the delay, the neural patterns were comparable. Those phenomena we observed, therefore, were general hallmarks of working memory, and for the purposes of this paper all of the data was pooled together (except for Figure 4.2, where LFP power is shown for each individual task, and in Figure 4.3, where the time course of the delay activity are shown separately for each task). Behavioral performance was high for each of the tasks/monkeys (monkey C: 85%, monkey S: 77%, monkey P: 88%). Only correct trials are included in the present analysis.

### *Recordings*

We acutely inserted between 1 and 3 laminar probes into cortex in every recording session. Our recordings included a large portion of the macaque frontal cortex, spanning dorsal premotor regions (area 6DR, SEF, area 9) to lateral prefrontal cortex (area 46, area 8). In Figure 4.1G, the colored dots represent recording sites from each of the monkeys and their respective frontal areas. In addition to the areas that are depicted, we also performed several recording sessions from deeper midline structures, such as the supplementary motor area (SMA) and the anterior cingulate cortex (ACC).

We used linear laminar probes from Plexon ("U probes" and "V probes") with a variety of inter-site spacing (100, 150, or 200um) and contact numbers (16, 24, or 32 contacts per probes). Probe geometry (inter-site spacing, channel count or U/V type) had no qualitative impact on the data we report here. Because the contact spacing ranged between 100 to 200um, we used cubic spline interpolation to up-sample the data and organize it into the same depth coordinates. To do so, we up-sampled to 100um spacing (if the spacing was already 100um, no interpolation was performed). Each individual probe was considered a unit of observation for our analyses in Figures 4.2, 4.3, and 4.5. For the analysis of thresholded units (Figure 4.4), the unit of observation was thresholded spikes.

In Figure 4.1C-E, MRIs are plotted with sample trajectories superimposed (red lines) in each of the monkeys. These trajectories were approximately perpendicular to the cortical sheet, allowing a relatively unbiased sampling of all cortical layers. All of the data was recorded through Blackrock headstages (Blackrock Cereplex M), sampled at 30 kHz, band-passed between 0.3 Hz and 7.5 kHz ($1^{st}$ order Butterworth high-pass and $3^{rd}$ order Butterworth low-pass), and digitized at a 16-bit, 250 nV/bit. All LFPs were recorded with a low-pass 250 Hz Butterworth filter, sampled at 1 kHz, and AC-coupled. In monkey C, the reference was the headpost, in monkey S, the reference was internal to the metal shaft surrounding the probe itself. In monkey P the reference was a nearby guide tube sitting on the dura. Some U/V probes had noisy channels (average power greater than 2 standard

deviations above the mean of all channels, this occurred on less than 5% of all channels), which were removed prior to analysis.

### *Lowering Procedure*

In order to place the contacts of the laminar electrode uniformly through the cortex, spanning from cerebrospinal fluid through the gray matter to the white matter, we used a number of physiologic indicators to guide our electrode placement. First, the presence of a slow 1-2 Hz signal, a heartbeat artifact, was often found as we pierced the pia mater and just as we entered the gray matter. Second, as the first contacts of the electrode entered the gray matter, the magnitude of the local field potential increased, and single units and/or neural hash became apparent, both audibly and visually with online spike thresholding. Once the tip of the electrode transitioned into the gray matter, electrodes were lowered slowly with minimal vibration an additional 1-2mm. At this point, the electrodes were allowed to settle for about 5 minutes, and then we began a visually evoked potential paradigm.

During this paradigm, we flashed a white screen on for 50ms with 500ms pauses while the animal's eye position was detected on the monitor. We repeated this about 200 times, and then calculated a power profile and Current Source Density (CSD) over the contacts of each of the implanted laminar probes. To compute the CSD, we cut the data into trials at the flash onset, obtained the evoked response to the flash across trials, and computed the evoked responses' second spatial derivative (Mitzdorf, 1985; Buzsaki et al., 1986).

The CSD reflects regions where ionic currents are flowing into neurons (caused by a net depolarization of surrounding neurons), and an early sink (where the CSD goes negative within 100ms of stimulus presentation) in response to visual input has been mapped to layer 4 in visual and auditory cortex (van Kerkoerle et al., 2014; Schroeder et al, 2002). Layer 4 in these regions receives the bulk of thalamic or bottom-up sensory inputs. In frontal cortex for areas that are dysgranular (area 6 and 9), an early sink corresponds to the bottom of layer 3 (Godlove et al., 2014), which receives visual sensory inputs (Shipp et al., 2005). In the case of granular or eulaminate prefrontal cortex (areas 46d/v) we expect the early sink to correspond to layer 4, which receives thalamic input from mediodorsal nucleus of the thalamus as well as sensory cortical areas (Giguere and Goldman-Rakic, 1988; Zikopoulos and Barbas, 2007).

We sought to position this current sink at approximately the middle contact of the probe. In addition to using the CSD for electrode alignment, which was more or less noisy, we also computed power across the alpha/beta (10-30 Hz), and gamma bands (30-80 Hz). We used these power estimates at each contact and normalized the power estimates by the maximal power within both of these bands. We found that the crossings of the normalized powers within the alpha/beta and gamma bands nearly always occurred within one or two sites of the earliest sink. Since these power profiles correlated strongly with the CSD results, we used these power profiles to also guide our decisions on whether or not we lowered the electrode any further. Finally, we were also careful not to penetrate by more than an electrode contact or two into the underlying white matter. The

white matter was characterized by the predominance of upward going spikes, and the absence of any CSD (i.e. there was markedly low LFP variability). Once the probe was fully lowered, we allowed it to settle for an hour, and then began a longer flashing sequence of 800-1000 trials for offline CSD analysis.

We determined each session's "zero point" (corresponding to layer 4/bottom of layer 3) as the site at which current sinks were detected within a time window of 40-180ms of flash onset. CSDs were calculated for each trial, taking a spatial integration of between 350-400 um, to be as consistent as possible given our probe geometries (i.e., using 100um inter-electrode spacing, we integrated every $4^{th}$ contact, for 200um spacing, every $2^{nd}$ contact, and for 150um, every $3^{rd}$ contact). We subtracted the CSD at the pre-flash baseline (50ms to flash onset), and then z-scored the CSD data over trials by dividing the raw CSD values by their standard error. We then assessed which CSD contact first achieved a z-score of less than -4 (negative CSD values correspond to current sinks), and which lasted at such a level for at least 6 ms. We assigned this contact as the first significant sink, and this provided the zero (the middle layer) for each penetration.

Relative to this zero point, the average distance to the CSF outside of gray matter was 0.9mm, estimated based on the total power of the LFP (when power across all frequencies decreased below 30% of the contact with maximal power, similar to Godlove et al., 2014). We also measured the cortical thickness for each of the 60 penetrations based on each individual monkey's MRI. To visualize the specific electrode tracks, we

obtained an MRI of each animal with their respective recording chamber and recording grid in place. Filled with water, the grid lumens could be tracked, and electrode trajectories projected onto the cortex and its folds (Fig. 4.1D-F). The mean cortical thickness across the penetrations in this study was 2.4mm. Thus to span the full average cortical distance in our analyses, laminar profiles were plotted from 0.9mm above the sink (the corresponding depth of the average gray matter / CSF transition) to 1.5mm below the sink (average gray matter / white matter transition). Because our recordings span a number of frontal areas that are both dysgranular and granular, we chose not to classify particular contacts to specific layers, as the laminar widths and organization could be slightly different over areas. Instead, we chose a more general classification system that could be applied to all areas, grouping contacts above the sink as superficial layers (corresponding to layers 1-3) and contacts below the sink as deep (corresponding to layers 5-6). The distance from the sink was a proxy for laminar position.

### *Analysis*

All analysis were performed with customized MATLAB scripts and with Fieldtrip software (Oostenveld et al., 2011). Given probe movements across a day, we smoothed across contacts on all analyses. This smoothing was symmetric, and across 200um above and below the contact of interest. This smoothing parameter equaled the distance between the contacts of the coarsest probe used in this study. Power was calculated based on 700 ms segments of data (200 ms prior flash and 500 ms post flash), which were tapered with Hanning windows. After these steps, we applied a fast Fourier transform. To compute

phase amplitude coupling (also known as cross-frequency coupling) we applied the Hilbert transform on two sets of band-passed filtered data: one, band-passed for lower frequencies (4-22 Hz), and the other band-passed at higher frequencies (50-250 Hz). The modulation index (Tort et al., 2008) was used to quantify the extent to which the faster frequency power deviated from a flat distribution over different phases (taking 18 non-overlapping equally spaced phase bins). We took the phase of the lower frequency Hilbert transform, and the amplitude of the higher frequency Hilbert transform.

Power and cross-frequency coupling analysis were calculated on unipolar data. Granger causality (GC) was computed through a nonparametric spectral matrix factorization of the Fourier transforms of bipolar data (Dhamala et al., 2008). The nonparametric estimation of GC has certain advantages over parametric approaches in that it does not require the specification of a particular autoregressive model order. Bipolar derivation is a recommended pre-step prior to Granger causality analysis, as the presence of a common reference can lead to spurious results (Bastos and Schoffelen, 2016; Trongnetrpunya et al., 2016). However, when bipolar derivations are too close to one another and cortical sources are synchronous, the effect of bipolar derivations is to remove the oscillation of interest. Therefore, bipolar derivations were performed between adjacent probes, always taking the contacts that were aligned to the same depth (relative to the sink). This was only possible when two probes had been simultaneously lowered to adjacent cortical areas (between 2-4 mm apart). To control for possible contributions of signal-to-noise differences in driving GC, we performed time-reversed Granger testing (Vinck et al.,

2015). In all cases, time-reversing the signals either reduced the dominant directionality or flipped it (from ascending to descending), confirming that signal-to-noise differences could not explain the ascending dominance of GC.

For the analysis of gamma and alpha/beta bursts, we used wavelets to capture deviations in power over time that could change quickly in both time and frequency, reflecting the non-stationarities of interest. Power was computed at equally-spaced frequency bins between 4 and 150 Hz. For each frequency, we used wavelets with a width of 5 cycles, and estimated power every 10ms. Bursts were detected as epochs when power exceeded a threshold of mean +2 standard deviations for at least three cycles, given the center frequency of that burst. Bursts in the gamma-band were detected between 50 to 150 Hz, reflecting the fact that previous findings have shown this to be the upper bound for oscillatory gamma responses (Lundqvist et al., 2016). Bursts in the alpha/beta range were detected between 4-22 Hz, chosen according to frequencies which had peak power in the deep layers. To minimize the contribution of spikes to the gamma bursts, we removed bursts that contained significant increases in power across all gamma frequencies (50-150Hz), reflecting the fact that spikes are expected to contribute power to this entire band (Ray and Maunsell, 2011). To calculate whether the delay-period bursting rate contained WM information, we summed the number of bursts around a sliding window (200ms for gamma and 400ms for alpha/beta), and assessed whether the burst rate reliably distinguished between the different cues using an unbiased measure of information,

percent explained variance (omega-squared; Olejnik and Algina, 2003). Information was then averaged across the delay period.

For the analysis of the analog multi-unit activity (MUA, Figure 4.3) we band-pass filtered the raw, unfiltered, 30kHz sampled data into a wide band between 500-5,000Hz, the power range dominated by spikes. The signal was then low-pass filtered at 250Hz and re-sampled to 1,000 kHz. For the analysis of thresholded spikes, we re-referenced each contact's signal to the global mean across all contacts, applied a $6^{th}$ order 250Hz, high-pass Butterworth filter, and then z-scored each signal by its own mean and standard deviation. We next identified spiking by identifying those time periods when the z-scored signal fell below 5 standard deviations of the mean noise floor. In order to ensure that all of these spikes were appropriately captured in time, we further extracted 10 samples before and 24 samples after threshold crossing from the original non-thresholded, 30 kHz signal. This new data was up-sampled by 2 using cubic splines, and the global minimum of each of these snippets was identified. If any spike was counted twice, because of more than a single threshold crossing within this time interval, they were rejected. Finally, to rule out spurious thresholds due to random noise or unit drift, we included units for further analysis if they maintained a firing rate of at least 2 Hz for at least 25 trials.

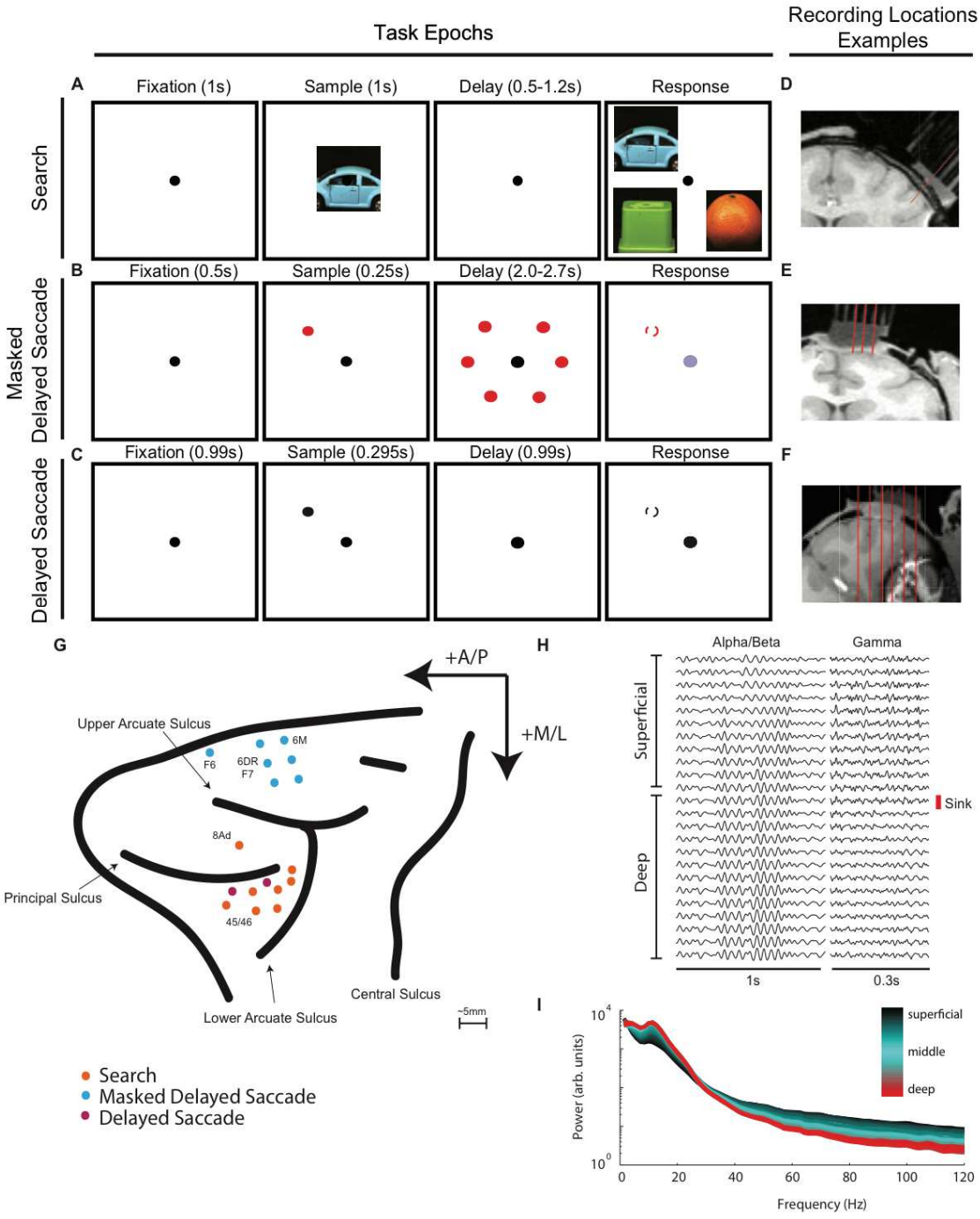


**Figure 4.1** (see caption on next page)

**Figure 4.1** (see figure on previous page) **A.** Visual search task: A match between sample and test image was chosen after a variable delay (0.5-1.2s) by making a saccade to the same image amongst a panel of 3 presented peripherally. Each image was positioned randomly at any 1 of 4 possible locations (upper right, lower right, upper left, and lower left). **B.** Masked delayed saccade: After a sample period, during which a single spatial location was cued (1 of 6 possible locations), the animal had to hold fixation through variable delay (2.2 to 2.7s) and the presentation of visual mask. After this delay, and when the fixation point color changed, the animal had to saccade to the previously cued location. **C.** Delayed saccade: After a sample period, during which a single spatial location was cued (1 of 4 possible locations), the animal had to hold fixation through a fixed delay (0.99s) and saccade to the cued location when the fixation dot disappeared. **D-E.** The small red lines indicate sample trajectories that were chosen to be as perpendicular as possible to cortex. **F.** The small red lines indicate sample trajectories that were possible given the recording hardware. Only the 3$^{rd}$ trajectory from the left was used for laminar recordings. **G.** We recorded across frontal cortex. The different colored dots indicate the task, and the letters the corresponding anatomical region. In addition to those labeled, we recorded from the anterior cingulate cortex (ACC) and the supplementary motor area (SMA). **H.** Sample LFP recordings were band-passed filtered using 3$^{rd}$ order Butterworth filters (on the left, 10-25 Hz; on the right, 40-160 Hz). The red line marks the location of the first significant CSD sink and the distinction between supra- and infragranular layers. **I.** A sample power spectrum with a clear alpha/beta bump (between 10 and 25 Hz) and broadband gamma (> 40 Hz). The variations across layers are plotted as a color gradient (black superficial, and red deep).

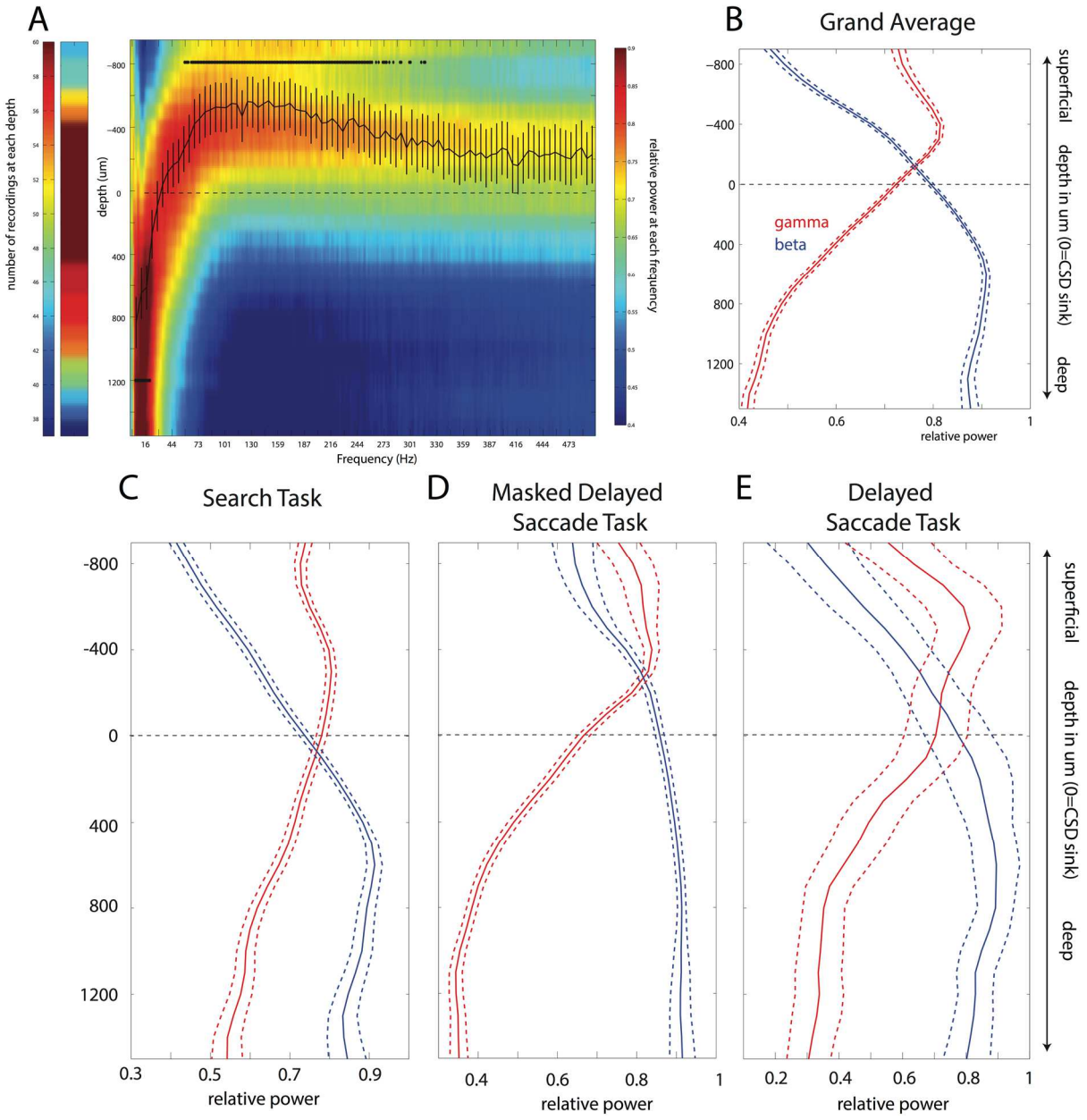


**Figure 4.2 A.** Normalized power averaged across laminar multi-contact probes across cortical depth and frequency. Red colors indicate greater power at a particular depth, and blue less. The black line is the average depth where the power at each frequency peaks. Error bars ± 1 SEM. The black stars indicate frequency bins at which the mean depth was significantly superficial or deep (Bonferroni corrected for multiple comparisons). ***B.*** Normalized power averaged across low frequencies (4-22 Hz, blue line), and high frequencies (50-250 Hz, red line). Error bars ± 1 SEM. ***C-E.*** Normalized power profiles across low and high frequencies for each task (Visual-Search, Masked Delayed-Saccade, Delayed-Saccade). Error bars ± 1 SEM.

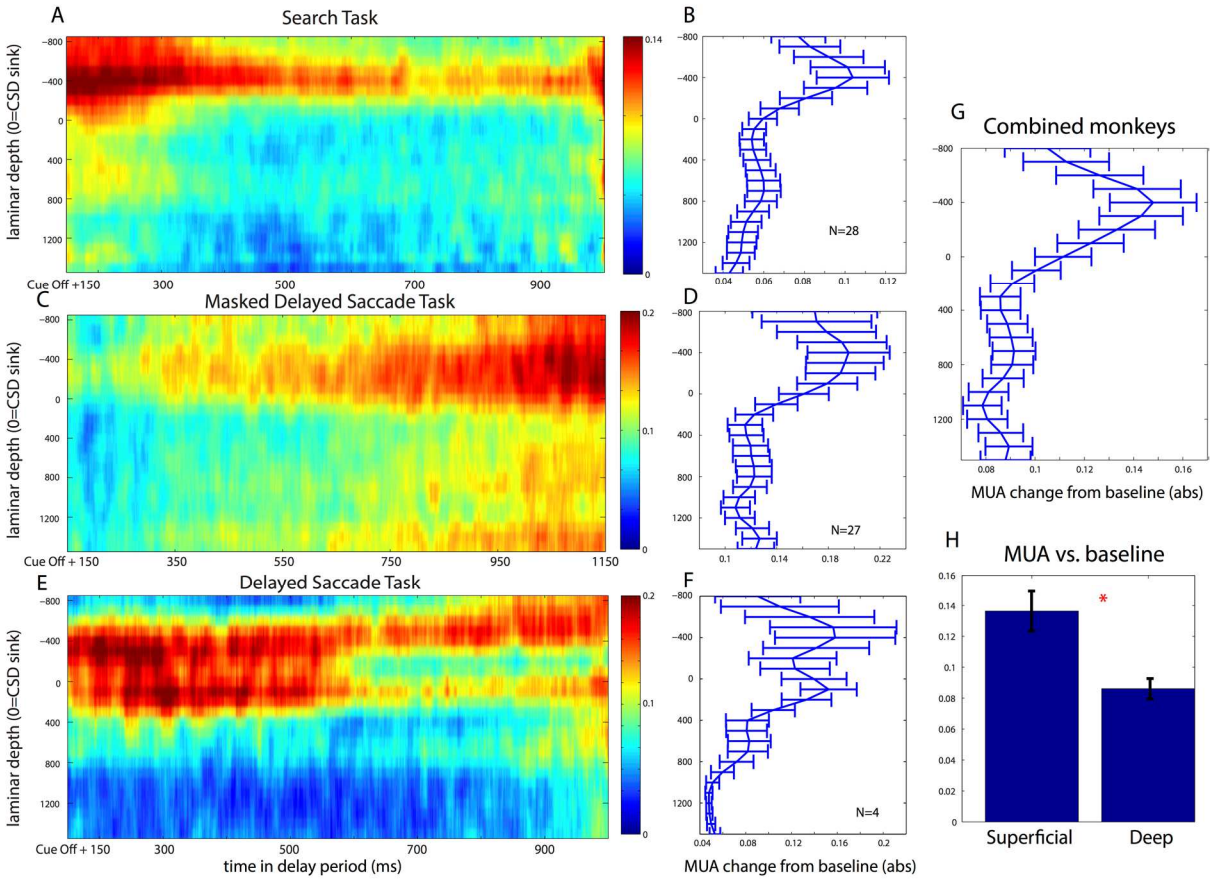


**Figure 4.3** ***A,C,E.*** Delay-period, MUA modulation across cortical depth and time. Plotted from 150ms after sample offset until ~1s into the delay, for the visual search, masked delayed-saccade, and delayed saccade tasks. ***B,D,F.*** The mean delay-period MUA modulation across the entire delay and cortical depth for each task. Error bars ± 1 SEM. *G.* The delay-period MUA modulation averaged across all tasks and plotted across cortical depth. Error bars ± 1 SEM. *H.* The mean MUA modulation averaged across all tasks, and all superficial or deep contacts. Error bars ± 1 SEM.

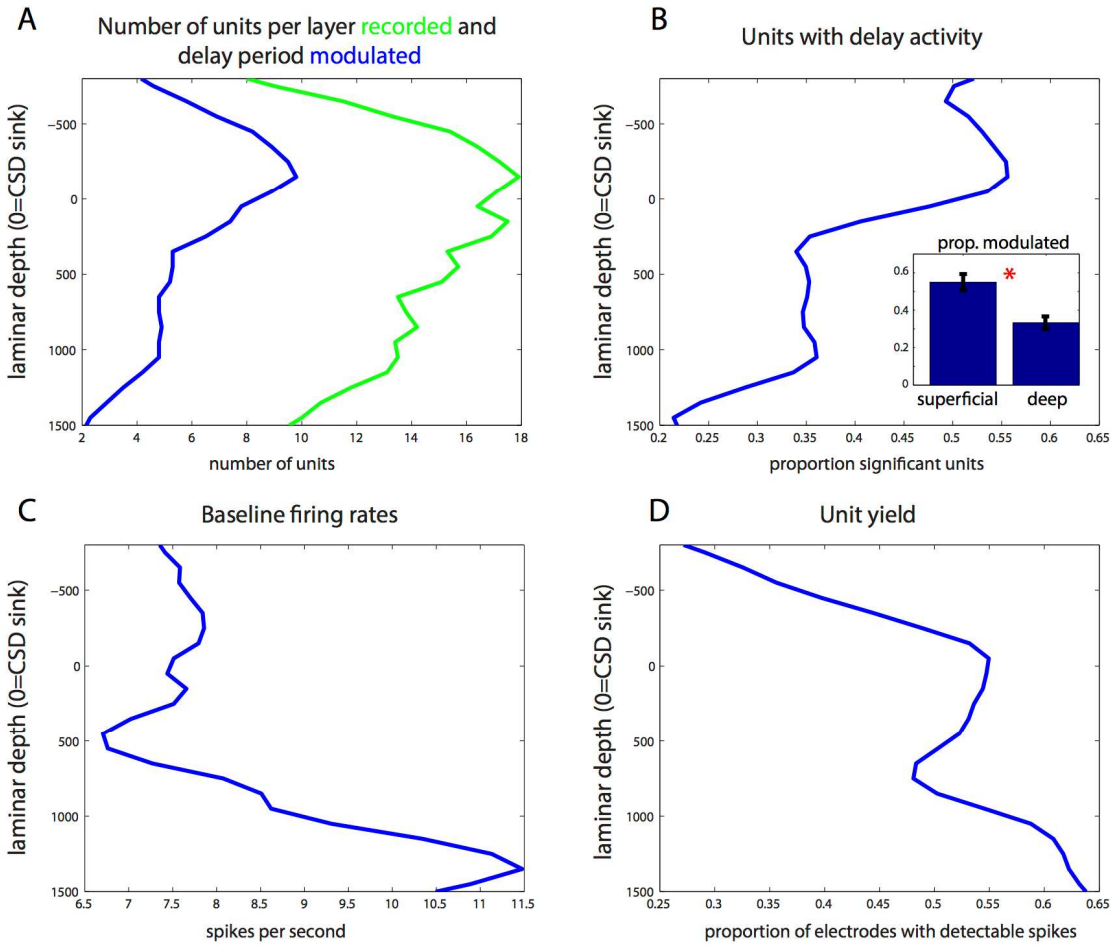


**Figure 4.4** ***A.*** Number of units recorded and the absolute number of units significantly modulated during the delay (t-test, $p<.01$) across laminar depths. ***B.*** The proportion of units that were significantly modulated during the delay. (inset) A bar graph comparing the proportion modulated in superficial vs. deep layers. Error bars ± 1 SEM. ***C.*** The mean baseline firing rates across laminar depths. ***D.*** The proportion of electrodes with detectable units.

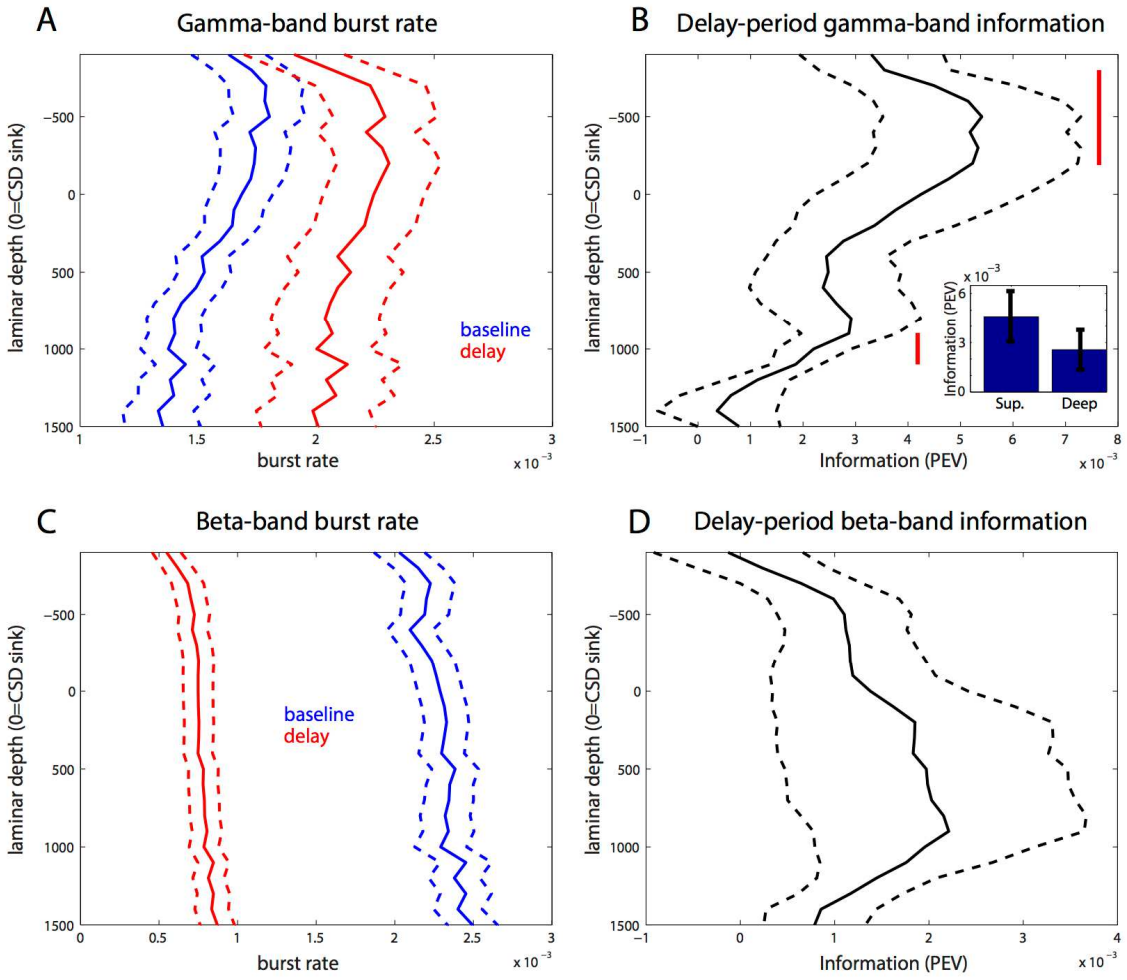


**Figure 4.5** ***A.*** Gamma burst rates at baseline (blue) and during the delay (red). ***B***. The percent explained variance (omega-squared) of the gamma bursts across different cortical depths. (inset) The mean PEV across all superficial and deep contacts, respectively. Error bars ± 1 SEM. ***C***. Beta burst rates at baseline (blue) and during the delay (red). ***D***. The PEV of beta bursts across different cortical depths. Error bars ± 1 SEM.

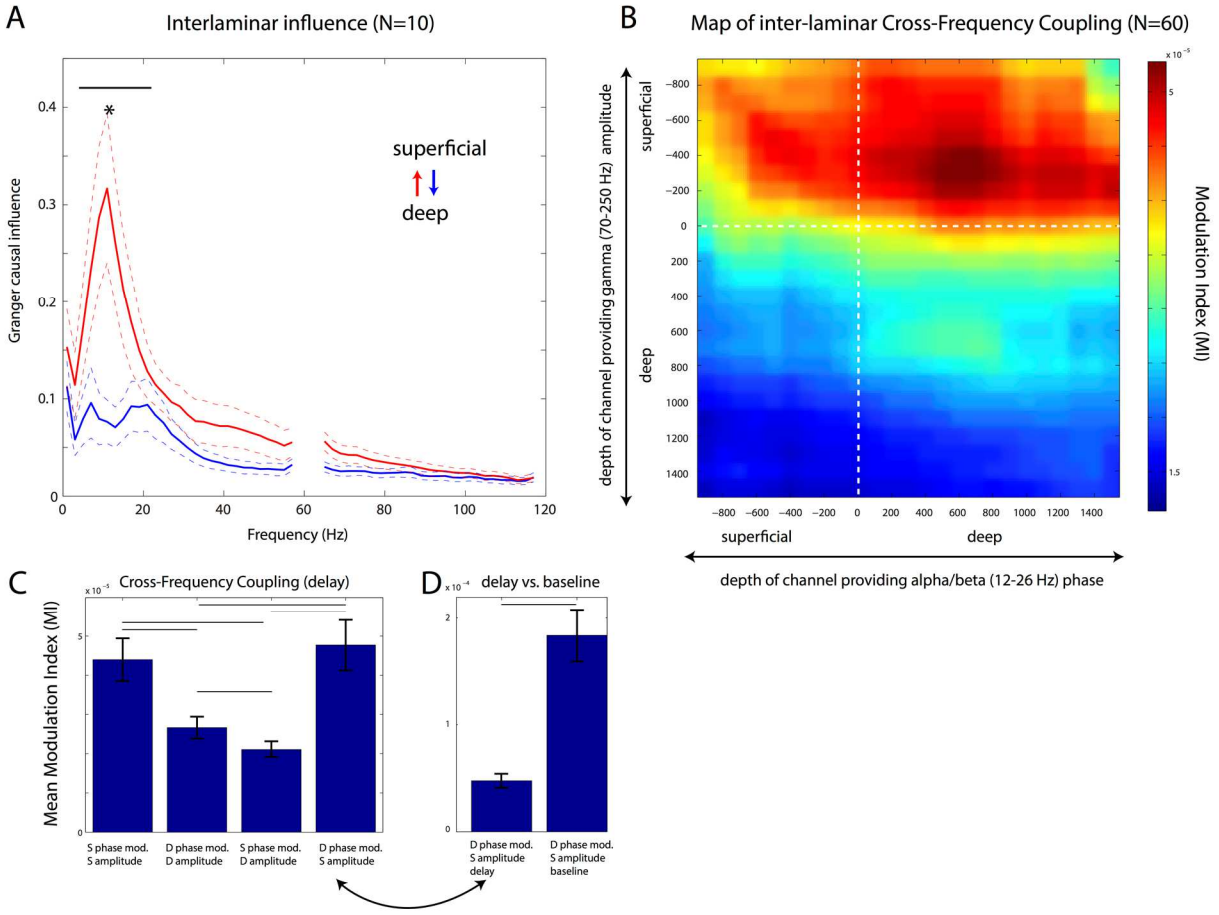


**Figure 4.6 A.** The Granger causal influence across frequency during the delay period. The red line is the Granger influence of deep to superficial layers, and the blue is the reverse. Only sessions in which 2 laminar probes were placed within 2-4 mm of one another were used (see Methods). **B.** Phase amplitude coupling between the phase of alpha/beta oscillations and the amplitude of gamma oscillations. Plotted across both axes is the PAC between specific cortical depths. **C.** The mean PAC across all four possible conditions during the delay-period: superficial phase to superficial amplitude (left), deep phase to deep amplitude (middle-left), superficial phase to deep amplitude (middle-right), and deep phase to superficial amplitude (right). **D.** PAC between deep phase to superficial amplitude during the delay (left) and the baseline (right).

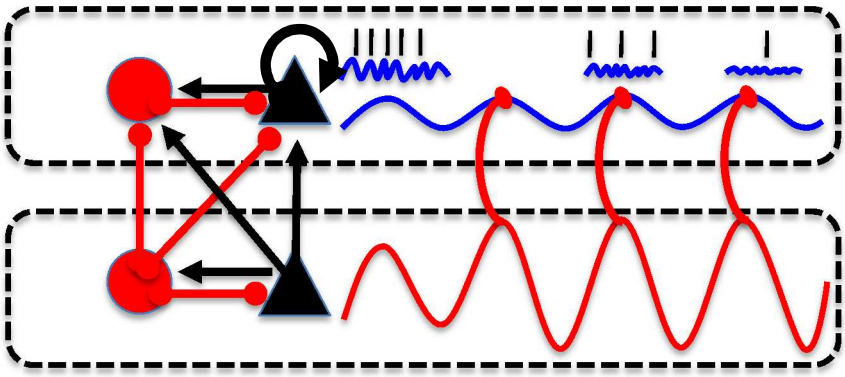

**Figure 4.7** *A model of working memory.* Denoted by two rectangular, dashed boxes, two cortical compartments, superficial and deep, are made up of densely interconnected pyramidal (black) and inhibitory (red) neurons. Inhibitory connections are line segments with a red, rounded end, and excitatory connections are line segments with a black, arrow end. The looping arrow returning on itself is reflects the recurrent connectivity found within layer 3 pyramidal cell networks in prefrontal cortex. The sinusoidal red-line in deep layers reflects the predominance of alpha/beta oscillations deep and their driving influencing on superficial alpha/beta oscillations (the sinusoidal blue line). Both the superficial and deep alpha/beta oscillations are coupled with gamma oscillations (blue squiggly lines), and these gamma oscillations organize informative spiking (straight black marks). Over time, moving from left to right in the figure, the deep alpha/beta suppress both superficial gamma and spiking.

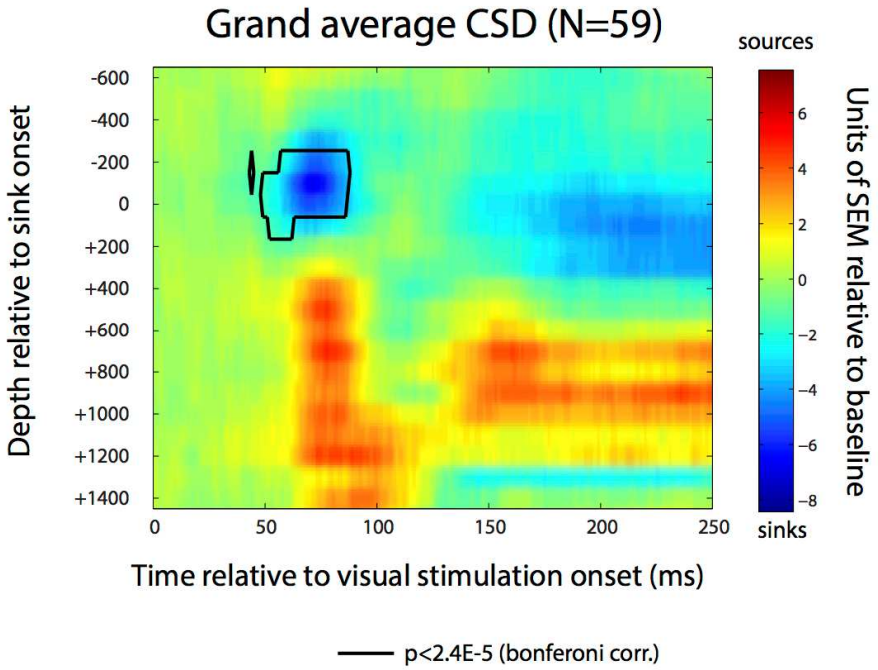


**Figure 4.S1** The grand average current source density plot after flash onset and after alignment to the first significant sink on each electrode. The black contour reflects significance at $p < 2.4E\text{-}5$ after Bonferroni correction.

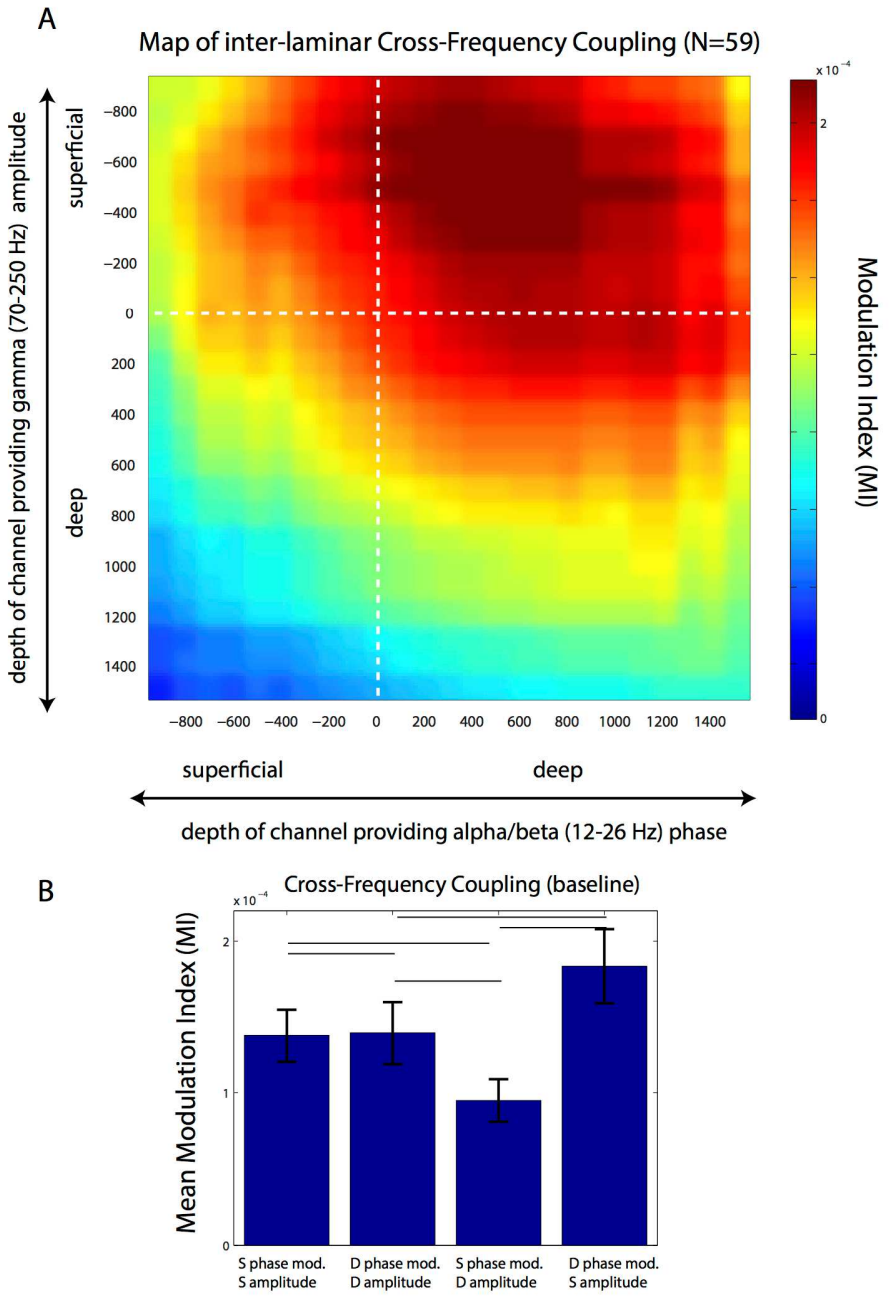


**Figure 4.S2** ***A.*** Phase amplitude coupling between the phase of alpha/beta oscillations and the amplitude of gamma oscillations during the baseline period. Plotted across both axes is the PAC between specific cortical depths. ***B.*** The mean PAC across all four possible conditions during the baseline period: superficial phase to superficial amplitude (left), deep phase to deep amplitude (middle-left), superficial phase to deep amplitude (middle-right), and deep phase to superficial amplitude (right).

## Bibliography

Bastos, A.M., and Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Frontiers in Systems Neuroscience *9*, 175.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron 76, 695–711.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., and Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron *85*, 390–401.

Bollimunta, A., Mo, J., Schroeder, C.E., and Ding, M. (2011). Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations. Journal of Neuroscience *31*, 4935–4943.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. PNAS *108*, 11262–11267.

Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., and Miller, E.K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron *76*, 838–846.

Buschman, T.J., and Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science *315*, 1860–1862.

Buzsáki, G., Czopf, J., Kondákor, I., and Kellényi, L. (1986). Laminar distribution of hippocampal rhythmic slow activity (RSA) in the behaving rat: current-source density analysis, effects of urethane and atropine. Brain Research *365*, 125–137.

Dhamala, M., Rangarajan, G., and Ding, M. (2008). Analyzing information flow in brain networks with nonparametric Granger causality. NeuroImage *41*, 354–362.

Douglas, R.J., and Martin, K. (1991). A functional microcircuit for cat visual cortex. Journal of Physiology *440*, 735-769.

Giguere, M., and Goldman-Rakic, P.S. (1988). Mediodorsal nucleus: areal, laminar, and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. Journal of Computational Neurology *277*, 195–213.

Godlove, D.C., Maier, A., Woodman, G.F., and Schall, J.D. (2014). Microcircuitry of agranular frontal cortex: testing the generality of the canonical cortical microcircuit. Journal of Neuroscience *34*, 5355–5369.

Goldman-Rakic, P.S. (1996). Regional and cellular fractionation of working memory. PNAS *93*, 13473–13480.

Haegens, S., Nácher, V., Luna, R., Romo, R., and Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. PNAS *108*, 19377–19382.

Jensen, O., Bonnefond, M., Marshall, T.R., and Tiesinga, P. (2015). Oscillatory mechanisms of feedforward and feedback visual processing. Trends in Neurosciences *38*, 192–194.

van Kerkoerle, T. van, Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., Togt, C. van der, and Roelfsema, P.R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. PNAS *111*, 14332–14341.

Luebke, J.I. (2017). Pyramidal neurons are not generalizable building blocks of cortical networks. Frontiers in Neuroanatomy *11*, 11.

Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., and Miller, E.K. (2016). Gamma and beta bursts underlie working memory. Neuron *90*, 152–164.

Maier, A., Adams, G.K., Aura, C., and Leopold, D.A. (2010). Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. Frontiers in System Neuroscience *4*.

Markov, N.T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., et al. (2013). The anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. Journal of Computational Neurology *522*, 225-259.

Markowitz, D.A., Curtis, C.E., and Pesaran, B. (2015). Multiple component networks support working memory in prefrontal cortex. PNAS *112*, 11084–11089.

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., and Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. Neuron *89*, 384–397.

Miller, E.K., Erickson, C.A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. Journal of Neuroscience *16*, 5154–5167.

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiological Reviews *65*, 37–100.

Ninomiya, T., Dougherty, K., Godlove, D.C., Schall, J.D., and Maier, A. (2015). Microcircuitry of agranular frontal cortex: contrasting laminar connectivity between occipital and frontal areas. Journal of Neurophysiology *113*, 3242–3255.

Olejnik, S., and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological Methods *8*, 434–447.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience *2011*, 156869.

Ray, S., and Maunsell, J.H.R. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLOS Biology *9*, e1000610.

Roberts, M.J., Lowet, E., Brunet, N.M., Ter Wal, M., Tiesinga, P., Fries, P., and de Weerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. Neuron 78, 523–536.

Sawaguchi, T., Matsumura, M., and Kubota, K. (1990). Catecholaminergic effects on neuronal activity related to a delayed response task in monkey prefrontal cortex. Journal of Neurophysiology *63*, 1385–1400.

Schroeder, C.E., and Foxe, J.J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cognitive Brain Research *14*, 187–198.

Shipp, S. (2005). The importance of being agranular: a comparative account of visual and motor cortex. Philosophical Transactions of the Royal Society of London B: Biological Sciences *360*, 797–814.

Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355.

Spaak, E., Bonnefond, M., Maier, A., Leopold, D.A., and Jensen, O. (2012). Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. Current Biology *22*, 2313–2318.

Tort, A.B.L., Kramer, M.A., Thorn, C., Gibson, D.J., Kubota, Y., Graybiel, A.M., and Kopell, N.J. (2008). Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task. PNAS *105*, 20517–20522.

Trongnetrpunya, A., Nandi, B., Kang, D., Kocsis, B., Schroeder, C.E., and Ding, M. (2015). Assessing Granger causality in electrophysiological data: removing the adverse effects of common signals via bipolar derivations. Frontiers in Systems Neuroscience *9*, 189.

Vinck, M., Huurdeman, L., Bosman, C.A., Fries, P., Battaglia, F.P., Pennartz, C.M.A., and Tiesinga, P.H. (2015). How to detect the Granger-causal flow direction in the presence of additive noise? NeuroImage *108*, 301–318.

Zikopoulos, B., and Barbas, H. (2007). Circuits for multisensory integration and attentional modulation through the prefrontal cortex and the thalamic reticular nucleus in primates. Reviews in the Neurosciences *18*, 417–438.

# CHAPTER 5: DISCUSSION

Overall, we investigated how different neural oscillations varied within the context of learning, working memory, and categorization. Combining all three studies, we found evidence that alpha/beta oscillations reflected top-down, feedback control, and that gamma oscillations reflected bottom-up, feedforward influences.

In learning, we found that 3 different tasks could be differentiated as a function of the animal's capacity to respond to negative feedback (i.e. errors). In both Match tasks, CM and OM, animals continued to perform at a high level following an error, and their PFC exhibited a strong evoked response to the error. In contrast, in the CS task, the animal's performance dropped substantially following an error, and there was no ERN in the PFC. We found that these dissociations between tasks qualified each of them separately as either explicit or implicit. Explicit learning is supervised, hypothesis-based, and hippocampal-dependent (Ashby and Maddox, 2011; Reber, 2013). This type of learning therefore requires the integration of errors in order to assess the accuracy of potential hypotheses. In contrast, implicit learning is understood as unconscious and automatic (Reber, 2013). Errorless methods enhance the retention of implicit skills, and errors are generally believed to be more often disruptive in such skills (Poolton et al., 2005; Donaghey et al., 2010). In others words, explicit learning is a top-down feedback driven process, whereas implicit learning may be dominated by bottom-up, feedforward processes. As a result, we characterized the Match tasks as explicit and top-down, and the Saccade tasks as implicit and bottom-up.

Those differences we found in LFP synchrony during the feedback epoch, therefore, are likely attributable to these differences in top-down, feedback vs. bottom-up, feedforward processes. In both explicit tasks, there was a prevalence of alpha/beta oscillations that changed with learning (increasing and then decreasing). In the implicit task, there was a prevalence of theta oscillations that changed with learning (monotonically decreasing). Despite a comparable quantity of alpha/beta synchrony, there were no changes in alpha/beta synchrony with learning in the implicit task. Overall, the increase in alpha/beta synchrony with both learning stage and performance in the Match tasks strongly supported our predictions regarding alpha/beta oscillations as top-down control signals. In other words, we had thought that as concepts are learned, top-down control should increase, and there should be a concordant increase in alpha/beta synchrony. However, we observed that late in learning there was a drop in alpha/beta synchrony. We attributed this drop to the end (or winding down) of PFC-guided learning. At that point, there was a decreased need for PFC to exert control on the activity of cortices more closely tied to task execution.

While this study did not examine how gamma oscillations varied across these three learning tasks, it did find a dissociation between alpha/beta and theta synchrony. In order to be consistent with our general interpretation of feedback vs. feedforward, theta synchrony should reflect a feedforward process. In a study examining the LFP correlates of feedforward vs feedback processes, Bastos et al. found that both the theta and gamma

bands carried feedforward signals across the different levels of the visual hierarchy (Bastos et al., 2015). A follow-up study confirmed that theta oscillations indeed reflect a feedforward process, and that they are down-regulated by top-down control (Spyropoulos et al., 2017).

During categorization, we found two distinct patterns associated with alpha/beta and gamma oscillations. In the VLPFC, relative to baseline, there was both a drop in alpha/beta power and a coincident increase in gamma power. Gamma oscillations in this area carried significant category information, which correlated with behavioral performance, and this information was greatest for categories of low abstractness. There was also in VLPFC significant category information in both the evoked response and in the theta band. In contrast, in DLPFC, relative to baseline, there was a significant increase in alpha/beta power and a concurrent drop in gamma power. The alpha/beta power increase correlated with behavioral preference, and carried significant category information. There was a greater amount of category information for categories of high abstractness.

Oscillatory activity in VLPFC and DLPFC corroborated our main hypotheses regarding alpha/beta and gamma oscillations in two different ways. One, alpha/beta oscillations occurred at moments requiring the greatest amount of top-down control, i.e. on high abstractness trials, while gamma oscillations occurred at moments requiring the least amount of top-down control, i.e. low abstractness trials. Two, gamma oscillations

appeared more tied to stimulus onset, occurring rapidly after stimulus presentation and arising in VLPFC where there was a large, short-latency evoked response that carried category information. In contrast, alpha/beta oscillations lasted throughout the delay in DLPFC, where there was an evoked response which did not carry significant category information. In short, bottom-up processes should arise rapidly in response to stimuli, and decrease in periods of greater cognitive demand. This was true here for gamma oscillations. In contrast, top-down processes should increase in periods of greater cognitive demand, and be less tied to the stimulus. This was true here but for alpha/beta oscillations.

During working memory, we found that gamma and alpha/beta oscillations were localized to layers within frontal cortex that corresponded to layers associated with feedforward and feedback connectivity. Van Essen and others have previously found that feedforward neurons originate from supragranular regions, while feedback neurons originate from both infragranular regions and subcortical regions. Consistent with these findings, we found that gamma peaked within superficial layers and alpha/beta peaked deep. Moreover, consistent with the role of feedback connections in modifying bottom-up activity, according to the biased competition model, we found that feedback layers had a modulatory role on feedforward ones, and not vice versa. Specifically, we found that deep layer alpha/beta oscillations both entrained superficial alpha/beta, and coupled with superficial gamma. According to our working memory model, deep alpha/beta oscillations serve to inhibit superficial neuronal ensembles i.e. gamma oscillations, thus

restricting the dominance of any particular ensemble. However, during the working memory delay, a decrease in alpha/beta-mediated inhibition leads to an excitation-inhibition imbalance. This imbalance allows for the persistent excitation of one cue-specific ensemble. Working memory, therefore, reflects the careful pruning of feedforward signals by inhibitory, feedback layers.

In summary, we found strong evidence to suggest that, within PFC, alpha/beta oscillations reflect top-down control mediated by feedback connections, while gamma (and possibly theta) oscillations reflect bottom-up influences via feedforward connections. These results are consistent with EEG findings in a number of psychiatric disorders. For instance, Alzheimer's disease is characterized by a progressive dementia and the loss of explicit learning, despite the maintenance of implicit-based perceptual learning (Fleischman, 2005). In line with earlier findings of this thesis, the increasing reliance on implicit learning in Alzheimer's patients leads to an increase in delta/theta oscillations and a reduction of alpha/beta (Jeong, 2004). Further, some hallmarks of schizophrenia include delusions and auditory hallucinations, where a patient can "lose touch" with reality. In these individuals, the most prominent finding is a reduction in gamma oscillations during cognitive tasks (Uhlhass and Singer, 2013). This reduction may lead to the relative prevalence of internally driven feedback processes and, hence, a confusion between top-down and bottom-up inputs.

## Future Work

Future work elucidating the cellular mechanisms underlying these rhythms will be necessary to better understand their functional role. An improved mechanistic understanding should lead to a more appropriate and biologically-based subdivision of those broad frequency bands hereto combined into the alpha/beta and gamma bands. Moreover, tracking these oscillations in vivo, and directly modulating them via external interventions may provide therapeutic relief for those psychiatric diseases associated with abnormal top-down/bottom-up control. In fact, preliminary evidence suggests that these oscillations are amenable to external intervention. For instance, we applied transcranial alternating current stimulation in phase with ongoing alpha oscillations within the frontal cortex of the monkeys studied here, and we found that we could enhance alpha/beta power in a period extending past the stimulation. This did not occur when stimulating using an arbitrarily-timed 12 Hz signal or a previously recorded brain signal. If successful, the external modulation of oscillations either through enhancement or suppression may allow us to better guide learning, store the contents of working memory, or generalize concepts. While providing potential therapeutic benefits, these electrophysiological interventions will also test the causal role these oscillations may have in cognitive functioning. Testing this causality is of tantamount importance, for most studies on neural oscillations are correlational in nature. Finally, a better appreciation of these oscillatory dynamics within the electroencephalogram (EEG) of humans could lead to the early identification of progressive neurological disorders, such

as schizophrenia and Alzheimer's, or alternatively of those learning disabilities where explicit learning is impaired.

## BIBLIOGRAPHY

Aizenstein, H.J., MacDonald, A.W., Stenger, V.A., Nebes, R.D., Larson, J.K., Ursu, S., and Carter, C.S. (2000). Complementary category learning systems identified using event-related functional MRI. Journal of Cognitive Neuroscience *12*, 977–987.

Antzoulatos, E.G., and Miller, E.K. (2011). Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. Neuron *71*, 243–249.

Antzoulatos, E.G., and Miller, E.K. (2014). Increases in functional connectivity between prefrontal cortex and striatum during category learning. Neuron *83*, 216–225.

Antzoulatos, E.G., and Miller, E.K. (2016). Synchronous beta rhythms of frontoparietal networks support only behaviorally relevant representations. eLife *5*, e17822.

Asaad, W.F., Rainer, G., and Miller, E.K. (1998). Neural activity in the primate prefrontal cortex during associative learning. Neuron *21*, 1399–1407.

Ashby, F.G., and Maddox, W.T. (2005). Human category learning. Annual Review of Psychology *56*, 149–178.

Ashby, F.G., and Maddox, W.T. (2011). Human category learning 2.0. Annals of the New York Academy of Sciences *1224*, 147–161.

Ashby, F.G., and O'Brien, J.B. (2005). Category learning and multiple memory systems. Trends in Cognitive Sciences *9*, 83–89.

Babapoor-Farrokhran, S., Vinck, M., Womelsdorf, T., and Everling, S. (2017). Theta and beta synchrony coordinate frontal eye fields and anterior cingulate cortex during sensorimotor mapping. Nature Communications *8*, 13967.

Bastos, A.M., and Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Frontiers in Systems Neuroscience *9*, 175.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron 76, 695–711.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., and Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron *85*, 390–401.

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P.L., Gioanni, Y., Battaglia, F.P., and Wiener, S.I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. Neuron *66*, 921–936.

Berger, H. (1929) Uber das Elektrenkephalogramm des Menschen (On the human electroencephalogram). Archiv für Psychiatrie und Nervenkrankheiten *87*, 527–70.

Bollimunta, A., Mo, J., Schroeder, C.E., and Ding, M. (2011). Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations. Journal of Neuroscience *31*, 4935–4943.

Brainard, D.H. (1997). The psychophysics toolbox. Spatial Vision *10*, 433–436.

Brincat, S.L., and Miller, E.K. (2015). Frequency-specific hippocampal-prefrontal interactions during associative learning. Nature Neuroscience *18*, 576–581.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. PNAS *108*, 11262–11267.

Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., and Miller, E.K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron *76*, 838–846.

Buschman, T.J., and Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science *315*, 1860–1862.

Buzsáki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nature Reviews Neuroscience *13*, 407–420.

Buzsáki, G., Czopf, J., Kondákor, I., and Kellényi, L. (1986). Laminar distribution of hippocampal rhythmic slow activity (RSA) in the behaving rat: current-source density analysis, effects of urethane and atropine. Brain Research *365*, 125–137.

Carpenter, K.L., Wills, A.J., Benattayallah, A., and Milton, F. (2016). A comparison of the neural correlates that underlie rule-based and information-integration category learning. Human Brain Mapping *37*, 3557–3574.

Chen, L.L., and Wise, S.P. (1995). Supplementary eye field contrasted with the frontal eye field during acquisition of conditional oculomotor associations. Journal of Neurophysiology *73*, 1122–1134.

Clare, L., Wilson, B.A., Breen, K., and Hodges, J.R. (1999). Errorless learning of face-name associations in early Alzheimer's disease. Neurocase *5*, 37–46.

Cleeremans, A., Destrebecqz, A., and Boyer, M. (1998). Implicit learning: news from the front. Trends in Cognitive Sciences *2*, 406–416.

Cohen, N.J., and Squire, L.R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science *210*, 207–210.

Colgin, L.L. (2013). Mechanisms and functions of theta rhythms. Annual Review of Neuroscience *36*, 295–312.

Cromer, J.A., Roy, J.E., and Miller, E.K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. Neuron *66*, 796–807.

DeCoteau, W.E., Thorn, C., Gibson, D.J., Courtemanche, R., Mitra, P., Kubota, Y., and Graybiel, A.M. (2007). Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task. PNAS *104*, 5644–5649.

Desimone, R., and Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. Annual Review of Neuroscience *18*, 193–222.

Dhamala, M., Rangarajan, G., and Ding, M. (2008). Analyzing information flow in brain networks with nonparametric Granger causality. NeuroImage *41*, 354–362.

Donaghey, C., McMillan, T., and O'Neill, B. (2010). Errorless learning is superior to trial and error when learning a practical skill in rehabilitation: a randomized controlled trial. Clinical Rehabilitation *24*, 195–201.

Douglas, R.J., and Martin, K. (1991). A functional microcircuit for cat visual cortex. Journal of Physiology *440*, 735-769.

Engel, A.K., and Fries, P. (2010). Beta-band oscillations — signalling the status quo? Current Opinion in Neurobiology *20*, 156–165.

Espenhahn, S., de Berker, A.O., van Wijk, B.C.M., Rossiter, H.E., and Ward, N.S. (2017). Movement-related beta oscillations show high intra-individual reliability. Neuroimage *147*, 175–185.

Evans, J.J., Wilson, B.A., Schuri, U., Andrade, J., Baddeley, A., Bruna, O., Canavan, T., Sala, S.D., Green, R., Laaksonen, R., et al. (2000). A comparison of "errorless" and "trial-and-error" learning methods for teaching individuals with acquired memory deficits. Neuropsychological Rehabilitation *10*, 67–101.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex *1*, 1–47.

Fleischman, D.A., Wilson, R.S., Gabrieli, J.D.E., Schneider, J.A., Bienias, J.L., and Bennett, D.A. (2005). Implicit memory and Alzheimer's disease neuropathology. Brain *128*, 2006–2015.

Frank, M.J., Woroch, B.S., and Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. Neuron *47*, 495–501.

Fries, P. (2015). Rhythms for cognition: communication through coherence. Neuron *88*, 220–235.

Fries, P., Reynolds, J.H., Rorie, A.E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science *291*, 1560–1563.

Gastgeb, H.Z., Dundas, E.M., Minshew, N.J., and Strauss, M.S. (2012). Category formation in autism: can individuals with autism form categories and prototypes of dot patterns? Journal of Autism and Developmental Disorders *42*, 1694–1704.

Gehring, W.J., and Willoughby, A.R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. Science *295*, 2279–2282.

Giguere, M., and Goldman-Rakic, P.S. (1988). Mediodorsal nucleus: areal, laminar, and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. Journal of Computational Neurology *277*, 195–213.

Godlove, D.C., Maier, A., Woodman, G.F., and Schall, J.D. (2014). Microcircuitry of agranular frontal cortex: testing the generality of the canonical cortical microcircuit. Journal of Neuroscience *34*, 5355–5369.

Goldman-Rakic, P.S. (1996). Regional and cellular fractionation of working memory. PNAS *93*, 13473–13480.

Gray, C.M., König, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature *338*, 334–337.

Gureckis, T.M., James, T.W., and Nosofsky, R.M. (2010). Re-evaluating dissociations between implicit and explicit category learning: an event-related fMRI Study. Journal of Cognitive Neuroscience *23*, 1697–1709.

Haegens, S., Nácher, V., Luna, R., Romo, R., and Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. PNAS *108*, 19377–19382.

Hargreaves, E.L., Mattfeld, A.T., Stark, C.E.L., and Suzuki, W.A. (2012). Conserved fMRI and LFP signals during new associative learning in the human and macaque monkey medial temporal lobe. Neuron *74*, 743–752.

Haegens, S., Cousijn, H., Wallis, G., Harrison, P.J., and Nobre, A.C. (2014). Inter- and intra-individual variability in alpha peak frequency. Neuroimage *92*, 46–55.

Herweg, N.A., Apitz, T., Leicht, G., Mulert, C., Fuentemilla, L., and Bunzeck, N. (2016). Theta-alpha oscillations bind the hippocampus, prefrontal cortex, and striatum during recollection: evidence from simultaneous EEG–fMRI. Journal of Neuroscience *36*, 3579–3587.

Hipp, J.F., Engel, A.K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron *69*, 387–396.

Histed, M.H., Pasupathy, A., and Miller, E.K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron *63*, 244–253.

Huerta, P.T., and Lisman, J.E. (1995). Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro. Neuron *15*, 1053–1063.

Hyman, J.M., Wyble, B.P., Goyal, V., Rossi, C.A., and Hasselmo, M.E. (2003). Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to the peak of theta and long-term depression when delivered to the trough. Journal of Neuroscience *23*, 11725–11731.

Jensen, O., Bonnefond, M., Marshall, T.R., and Tiesinga, P. (2015). Oscillatory

mechanisms of feedforward and feedback visual processing. Trends in Neurosciences *38*, 192–194.

Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Frontiers in Human Neuroscience *4*.

Jeong, J. (2004). EEG dynamics in patients with Alzheimer's disease. Clinical Neurophysiology *115*, 1490–1505.

Jutras, M.J., Fries, P., and Buffalo, E.A. (2009). Gamma-band synchronization in the macaque hippocampus and memory formation. Journal of Neuroscience *29*, 12521–12531.

Jutras, M.J., Fries, P., and Buffalo, E.A. (2013). Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. PNAS *110*, 13144–13149.

Kajikawa, Y., and Schroeder, C.E. (2011). How local is the local field potential? Neuron *72*, 847–858.

van Kerkoerle, T. van, Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., Togt, C. van der, and Roelfsema, P.R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. PNAS *111*, 14332–14341.

Kim, H., Ährlund-Richter, S., Wang, X., Deisseroth, K., and Carlén, M. (2016). Prefrontal parvalbumin neurons in control of attention. Cell *164*, 208–218.

Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition–timing hypothesis. Brain Research Reviews *53*, 63–88.

Knowlton, B.J., and Squire, L.R. (1993). The learning of categories: parallel brain systems for item memory and category knowledge. Science *262*, 1747–1749.

Kopell, N., Whittington, M.A., and Kramer, M.A. (2011). Neuronal assembly dynamics in the beta1 frequency range permits short-term memory. PNAS *108*, 3779–3784.

Kornblith, S., Buschman, T.J., and Miller, E.K. (2016). Stimulus load and oscillatory activity in higher cortex. Cerebral Cortex *26*, 3772–3784.

Lee, J.C., and Livesey, E.J. (2017). The effect of encoding conditions on learning in the prototype distortion task. Learning & Behavior *45*, 164–183.

Liebe, S., Hoerzer, G.M., Logothetis, N.K., and Rainer, G. (2012). Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. Nature Neuroscience *15*, 456–462.

Lisman, J.E., and Jensen, O. (2013). The theta-gamma neural code. Neuron *77*, 1002–1016.

Luebke, J.I. (2017). Pyramidal neurons are not generalizable building blocks of cortical networks. Frontiers in Neuroanatomy *11*, 11.

Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., and Miller, E.K. (2016). Gamma and beta bursts underlie working memory. Neuron *90*, 152–164.

Maier, A., Adams, G.K., Aura, C., and Leopold, D.A. (2010). Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. Frontiers in System Neuroscience *4*.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods *164*, 177–190.

Markov, N.T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., et al. (2013). The anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. Journal of Computational Neurology *522*, 225-259.

Markowitz, D.A., Curtis, C.E., and Pesaran, B. (2015). Multiple component networks support working memory in prefrontal cortex. PNAS *112*, 11084–11089.

Maxwell, J.P., Masters, R.S.W., Kerr, E., and Weedon, E. (2001). The implicit benefit of learning without errors. The Quarterly Journal of Experimental Psychology Section A *54*, 1049–1068.

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., and Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. Neuron *89*, 384–397.

Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annual Review of Neuroscience *24*, 167–202.

Miller, E.K., Erickson, C.A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. Journal of Neuroscience *16*, 5154–5167.

Milner, B., Corkin, S., and Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. Neuropsychologia *6*, 215–234.

Milton, F., Bealing, P., Carpenter, K.L., Bennattayallah, A., and Wills, A.J. (2016). The Neural Correlates of Similarity- and Rule-based Generalization. Journal of Cognitive Neuroscience *29*, 150–166.

Milton, F., and Pothos, E.M. (2011). Category structure and the two learning systems of COVIS. European Journal of Neuroscience *34*, 1326–1336.

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiological Reviews *65*, 37–100.

Morey, R. (2008) Confidence intervals from normalized data: a correction to Cousineau (2005). Tutorial in Quantitative Methods for Psychology *4,* 61-64.

Morrison, R.G., Reber, P.J., Bharani, K.L., and Paller, K.A. (2015). Dissociation of category-learning systems via brain potentials. Frontiers in Human Neuroscience *9*, 389.

Ninomiya, T., Dougherty, K., Godlove, D.C., Schall, J.D., and Maier, A. (2015). Microcircuitry of agranular frontal cortex: contrasting laminar connectivity between occipital and frontal areas. Journal of Neurophysiology *113*, 3242–3255.

Nunez, P.L., Srinivasan, R., (2006). Electric fields of the brain: the neurophysics of EEG. Oxford University Press, New York.

O'Connell, G., Myers, C.E., Hopkins, R.O., P, R., Gluck, M.A., and Wills, A.J. (2016). Amnesic patients show superior generalization in category learning. Neuropsychology *30*, 915–919.

O'Keefe, J., and Recce, M.L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus *3*, 317–330.

Olejnik, S., and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological Methods *8*, 434–447.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience *2011*, e156869.

Palmeri, T.J., and Mack, M.L. (2015). How experimental trial context affects perceptual categorization. Frontiers in Psychology *6*, 180.

Pasupathy, A., and Miller, E.K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature *433*, 873–876.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision *10*, 437–442.

Pesaran, B., Pezaris, J.S., Sahani, M., Mitra, P.P., and Andersen, R.A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. Nature Neuroscience *5*, 805–811.

Pfeiffer, B.E., and Foster, D.J. (2013). Hippocampal place cell sequences depict future paths to remembered goals. Nature *497*, 74–79.

Poolton, J.M., Masters, R.S.W., and Maxwell, J.P. (2005). The relationship between initial errorless learning conditions and subsequent performance. Human Movement Science *24*, 362–378.

Posner, M.I., Goldsmith, R., and Welton, K.E., Jr. (1967). Perceived distance and the classification of distorted patterns. Journal of Experimental Psychology *73*, 28–38.

Posner, M.I., and Keele, S.W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology *77*, 353–363.

Ray, S., and Maunsell, J.H.R. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLOS Biology *9*, e1000610.

Reber, P.J. (2013). The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. Neuropsychologia *51*, 2026–2042.

Reber, P.J., Gitelman, D.R., Parrish, T.B., and Mesulam, M.M. (2003). Dissociating explicit and implicit category knowledge with fMRI. Journal of Cognitive Neuroscience *15*, 574–583.

Reber, P.J., Stark, C.E.L., and Squire, L.R. (1998). Contrasting cortical activity associated with category memory and recognition memory. Learning & Memory *5*, 420–428.

Roberts, J.L., Anderson, N.D., Guild, E., Cyr, A.-A., Jones, R.S.P., and Clare, L. (2016). The benefits of errorless learning for people with amnestic mild cognitive impairment. Neuropsychological Rehabilitation 1–13.

Roberts, M.J., Lowet, E., Brunet, N.M., Ter Wal, M., Tiesinga, P., Fries, P., and de Weerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. Neuron 78, 523–536.

Rüsseler, J., Kuhlicke, D., and Münte, T.F. (2003). Human error monitoring during implicit and explicit learning of a sensorimotor sequence. Neuroscience Research *47*, 233–240.

Sakai, K., and Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. Nature *354*, 152–155.

Sawaguchi, T., Matsumura, M., and Kubota, K. (1990). Catecholaminergic effects on neuronal activity related to a delayed response task in monkey prefrontal cortex. Journal of Neurophysiology *63*, 1385–1400.

Scheffers, M.K., and Coles, M.G.H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. Journal of Experimental Psychology: Human Perception and Performance *26*, 141–151.

Schnitzler, A., and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. Nature Reviews Neuroscience *6*, 285–296.

Schroeder, C.E., and Foxe, J.J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cognitive Brain Research *14*, 187–198.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of Neurology, Neurosurgery & Psychiatry *20*, 11–21.

Seger, C.A., and Miller, E.K. (2010). Category learning in the brain. Annual Review of Neuroscience *33*, 203–219.

Shipp, S. (2005). The importance of being agranular: a comparative account of visual and motor cortex. Philosophical Transactions of the Royal Society of London B: Biological Sciences *360*, 797–814.

Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355.

Smith, J.D., Beran, M.J., Crossley, M.J., Boomer, J.T., and Ashby, F.G. (2010). Implicit

and explicit category learning by macaques (Macaca mulatta) and humans (Homo sapiens). Journal of Experimental Psychology: Animal Behavior Processes *36*, 54–65.

Smith, J.D., Boomer, J., Zakrzewski, A.C., Roeder, J.L., Church, B.A., and Ashby, F.G. (2013). Deferred feedback sharply dissociates implicit and explicit category learning. Psychological Science *25*, 447–457.

Sohal, V.S., Zhang, F., Yizhar, O., and Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature *459*, 698–702.

Spaak, E., Bonnefond, M., Maier, A., Leopold, D.A., and Jensen, O. (2012). Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. Current Biology *22*, 2313–2318.

Spyropoulos, G., Bosman, C.A., and Fries, P. (2017). A theta rhythm in awake macaque V1 and V4 and its attentional modulation. bioRxiv 117804.

Squires, E.J., Hunkin, N.M., and Parkin, A.J. (1997). Errorless learning of novel associations in amnesia. Neuropsychologia *35*, 1103–1111.

Swann, N.C., de Hemptinne, C., Aron, A.R., Ostrem, J.L., Knight, R.T., and Starr, P.A. (2015). Elevated Synchrony in Parkinson's Disease Detected with Electroencephalography. Annals of Neurology *78*, 742–750.

Thut, G., and Miniussi, C. (2009). New insights into rhythmic brain activity from TMS–EEG studies. Trends in Cognitive Sciences *13*, 182–189.

Tort, A.B.L., Kramer, M.A., Thorn, C., Gibson, D.J., Kubota, Y., Graybiel, A.M., and Kopell, N.J. (2008). Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task. PNAS *105*, 20517–20522.

Trongnetrpunya, A., Nandi, B., Kang, D., Kocsis, B., Schroeder, C.E., and Ding, M. (2015). Assessing Granger causality in electrophysiological data: removing the adverse effects of common signals via bipolar derivations. Frontiers in Systems Neuroscience *9*, 189.

Uhlhaas, P.J., and Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. Nature Reviews Neuroscience *11*, 100–113.

Uhlhaas, P.J., and Singer, W. (2013). High-frequency oscillations and the neurobiology of schizophrenia. Dialogues in Clinical Neuroscience *15*, 301–313.

Vinck, M., Huurdeman, L., Bosman, C.A., Fries, P., Battaglia, F.P., Pennartz, C.M.A., and Tiesinga, P.H. (2015). How to detect the Granger-causal flow direction in the presence of additive noise? NeuroImage *108*, 301–318.

Vogels, R., Sary, G., Dupont, P., and Orban, G.A. (2002). Human brain regions involved in visual categorization. NeuroImage *16*, 401–414.

Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956.

Walsh, M.M., and Anderson, J.R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. Neuroscience & Biobehavioral Reviews *36*, 1870–1884.

Wang, L., Saalmann, Y.B., Pinsk, M.A., Arcaro, M.J., and Kastner, S. (2012). Electrophysiological low-frequency coherence and cross-frequency coupling contribute to BOLD connectivity. Neuron *76*, 1010–1020.

Warden, M.R., and Miller, E.K. (2010). Task-dependent changes in short-term memory in the prefrontal cortex. Journal of Neuroscience *30*, 15801–15810.

de Werd, M.M., Boelen, D., Rikkert, M.G.O., and Kessels, R.P. (2013). Errorless learning of everyday tasks in people with dementia. Clinical Interventions in Aging *8*, 1177–1190.

Wessel, J.R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. Frontiers in Human Neuroscience *6*, 88.

Wessel, J.R., Danielmeier, C., and Ullsperger, M. (2011). Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. Journal of Cognitive Neuroscience *23*, 3021–3036.

Williams, Z.M., and Eskandar, E.N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. Nature Neuroscience *9*, 562–568.

Wirth, S., Avsar, E., Chiu, C.C., Sharma, V., Smith, A.C., Brown, E., and Suzuki, W.A. (2009). Trial outcome and associative learning signals in the monkey hippocampus. Neuron *61*, 930–940.

Wirth, S., Yanike, M., Frank, L.M., Smith, A.C., Brown, E.N., and Suzuki, W.A. (2003). Single neurons in the monkey hippocampus and learning of new associations. Science *300*, 1578–1581.

Xia, R., Guan, S., and Sheinberg, D.L. (2015). A Multilayered Story of Memory Retrieval. Neuron *86*, 610–612.

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

Zikopoulos, B., and Barbas, H. (2007). Circuits for multisensory integration and attentional modulation through the prefrontal cortex and the thalamic reticular nucleus in primates. Reviews in the Neurosciences *18*, 417–438.

# CURRICULUM VITAE