2018

The Neural Mechanisms Underlying Invariant Object Search In V4 And Inferotemporal Cortex

Noam Roth
University of Pennsylvania, noamroth@pennmedicine.upenn.edu

The Neural Mechanisms Underlying Invariant Object Search In V4 And Inferotemporal Cortex

Abstract

Finding a specific visual target, such as your car keys, requires the brain to combine visual information about objects in the currently viewed scene with working memory information about your target to determine whether your target is in view. This combination of context-specific signals with visual information is thought to happen via feedback of target information from higher brain areas to the ventral visual pathway. However, exactly where and how these signals are combined remains unknown. To investigate, we recorded neural responses in V4 and inferotemporal cortex (IT) while monkeys performed an invariant object search task, where targets could appear across variation in their size, position and background context. We applied two complementary approaches to this data to investigate the neural mechanisms underlying target search. The first approach (Chapter 2) is from a computational perspective: where and how are visual and target signals combined when searching for a target? Specifically, we found that while task-relevant modulations in V4 were large, they were larger in IT, suggesting that top-down context-specific modulations are integrated into the ventral visual pathway at multiple stages. In Chapter 3, we focused on the neural responses recorded from IT from the perspective of neural coding: we sought to understand how signal and noise combine to determine task performance. We found that while signals that report the solution for object search were much smaller than signals that act as noise for the task (nuisance modulations) in IT cortex, nuisance modulations had a small effect on task performance. This counterintuitive finding was due to large trial variability constrained by short, behaviorally relevant spike counting windows. Together, this body of work provides insight into where and how the brain combines context-specific signals with visual information during invariant object search.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Neuroscience

First Advisor

Nicole C. Rust

Subject Categories

Neuroscience and Neurobiology

THE NEURAL MECHANISMS UNDERLYING INVARIANT OBJECT SEARCH IN V4 AND INFEROTEMPORAL CORTEX

Noam Roth

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

Nicole Rust, PhD

Associate Professor of Psychology

Graduate Group Chairperson

Joshua Gold, PhD

Professor of Neuroscience

Dissertation Committee:

Yale Cohen, PhD, Professor of Otorhinolaryngology (Committee Chair)

Joshua Gold, PhD, Professor of Neuroscience

Kostas Daniilidis, PhD, Professor of Computer and Information Science

Marlene Cohen, PhD, Associate Professor of Neuroscience

THE NEURAL MECHANISMS UNDERLYING INVARIANT OBJECT SEARCH IN V4 AND INFEROTEMPORAL CORTEX

ACKNOWLEDGMENT

This work would not have been possible without the support and encouragement of an entire village of colleagues, mentors, friends, and family. First and foremost, I would like to thank my advisor and mentor Nicole Rust. Her wisdom, encouragement and passion have been an endless source of support. I am forever grateful to her dedication in every aspect of the research, but also in my growth as a scientist and as a human.

I would also like to thank the members of my thesis committee: Yale Cohen, Josh Gold, Kostas Daniilidis, and Marlene Cohen, for the invaluable advice and support over the years of my dissertation work.

I am very grateful to the wonderful community of people I had the pleasure of working with, both in the Rust lab and in the greater Neuroscience community at Penn. In particular, thank you to Krystal Henderson for her support, both in the lab and otherwise, to Sidd Joshi, Travis Meyer and Taka Doi for their electrophysiology wisdom and guidance. Thank you to Alex Smolyanskaya and Marino Pagan, whose conversations and time together reminded me of the importance of excitement and passion (for science, and everything.)

I would like to thank all of the members of the CNI for providing a friendly and supportive environment for the last year and a half of my thesis work. In particular, to Yunshu Fan, Arvind Iyer, Ben Chin, Michael Barnett, David White, Kyra Schapiro and Lingqi Zhang. A special thanks to Ben and Michael for their never-ending support during the last few months of dissertation writing.

A few friends in particular have been constant sources of love and encouragement. I am very grateful for the friendship, advice, and mentorship of Kate Christison-Lagay. Thank you to Isaac Perron and Leonardo Guercio for countless hours of laughter and support. A very special thank you to to Shachee Doshi for your humor, for grounding me and helping me see the good in everything. To Morgan Taylor for equal parts commiseration and nerding out over electrophysiology and computational neuroscience, but most of all for reminding me to take time for myself, and for steadfast love and encouragement. Lastly, I would not have been able to complete the dissertation without the support of Sarah Ly: thank you for your love, kindness, patience, humor, and most importantly, making the last few years feel like home.

Finally, I am so grateful to have my family to support me all of these years: to Ella for reminding me of the lighter side of life and for being mature beyond her years; to Edo, who is my strongest, and longest-lasting foundation. Lastly, but most importantly, I would like to thank my parents, for teaching me the value of hard work and perseverance, igniting my curiosity, and their endless patience and encouragement.

ABSTRACT

THE NEURAL MECHANISMS UNDERLYING INVARIANT OBJECT SEARCH IN V4 AND INFEROTEMPORAL CORTEX

Noam Roth

Nicole Rust

Finding a specific visual target, such as your car keys, requires the brain to combine visual information about objects in the currently viewed scene with working memory information about your target to determine whether your target is in view. This combination of context-specific signals with visual information is thought to happen via feedback of target information from higher brain areas to the ventral visual pathway. However, exactly where and how these signals are combined remains unknown. To investigate, we recorded neural responses in V4 and inferotemporal cortex (IT) while monkeys performed an invariant object search task, where targets could appear across variation in their size, position and background context. We applied two complementary approaches to this data to investigate the neural mechanisms underlying target search. The first approach (Chapter 2) is from a computational perspective: where and how are visual and target signals combined when searching for a target? Specifically, we found that while task-relevant modulations in V4 were large, they were larger in IT, suggesting that top-down context-specific modulations are integrated into the ventral visual pathway at multiple stages. In Chapter 3, we focused on the neural responses recorded from IT from the perspective of neural coding: we sought to understand how signal and noise combine to determine task performance. We found that while signals that report the solution for object search were much smaller than signals that act as noise for the task (nuisance modulations) in IT cortex, nuisance modulations had a small effect on task performance. This counterintuitive finding was due to large trial variability constrained by short, behaviorally relevant spike counting windows. Together, this body of work provides insight into where and how the brain combines context-specific signals with visual information during invariant object search.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1

Introduction

The problem of visual target search

Many everyday tasks require the brain to flexibly map incoming sensory information onto different behavioral responses based on context. One example is the task of finding a particular object, which requires the brain to solve two non-trivial computations. First, it requires the brain to form a representation of the object in view (i.e. visual signals). Next, it must compare this visual representation with a representation of the sought target (i.e. working memory signals). This comparison is thought to happen within the ventral visual pathway, where neurons not only underlie visual representations, but can also be strongly modulated by top-down, context-specific signals. However, exactly where and how this comparison happens is unclear. The aim of this dissertation was to investigate how these signals combine to support object search. We begin by taking a computational approach: where and how do these signals combine to compute a signal necessary to solve the task? Next, we focus on the same data from the perspective of neural coding: how do signal and noise combine to determine task performance? In this chapter, we review what is known about how visual signals in the ventral visual pathway are modulated by top-down, context-specific signals. We then discuss the role that signal and noise might play in determining neural task performance.

The largely feed-forward, sensory component of visual processing

A number of lines of evidence implicate the ventral visual pathway (Figure 1, cyan) in processing information about the identity of visual features and objects. This pathway consists of a series of hierarchically arranged cortical visual areas in the occipital and temporal lobe, including primary visual cortex (V1), secondary visual cortex (V2), area V4, and inferotemporal cortex (IT) (DiCarlo & Cox, 2007; Felleman & Van Essen, 1991; Ungerleider & Mishkin, 1982). The areas along this pathway are thought to underlie the encoding of visual information that gets progressively refined along each stage, and ultimately underlies representations of objects. This notion is supported by several lines of evidence. First, while lesions in monkey V1 cause blindness specific to the damaged portion of the visual field (reviewed by Stoerig & Cowey, 1997), lesions in V2 and V4 produce deficits in the ability to detect conjunctions of simple features (Merigan, Nealey, & Maunsell, 1993; Schiller, 1995) and lesions to IT produce specific deficits in the ability to distinguish among complex objects (Yaginuma Niihara, & Iwai, 1982; Holmes & Gross, 1984; although see Huxlin, Saunders, Marchionini, Pham, & Merigan, 2000).

Mirroring evidence from these lesion studies, responses of single neurons along this pathway reflect increases in selectivity for complex shapes and increases in invariance across small changes in position, size, and clutter (Hung, Kreiman, Poggio, & DiCarlo, 2005; Ito, Tamura, Fujita, & Tanaka, 1995; Kobatake & Tanaka, 1994). In addition, receptive fields become incrementally larger as visual information is pooled across wider portions of the visual field (Kobatake & Tanaka, 1994). For example, while

IT receptive fields are large (typically 5 degrees or larger in width) and extend into all four quadrants of the visual field, its main input area, V4 is organized quite differently. V4 receptive fields are smaller, retinotopically organized, and constrained to one visual field (Desimone & Schein, 1987; Gattass, Sousa, & Gross, 1988).

In parallel to these incremental complexities within single neuron responses, the population responses at each successive stage carry a progressively refined encoding of visual information (Rust & DiCarlo, 2010) which culminates in a robust representation of currently viewed objects in IT (Hung et al., 2005).

Top-down signals relay the representation of sought targets

Solving visual search requires the subject to actively hold the sought target in working memory while scanning the currently-viewed scene. Although the exact neural structures and mechanisms underlying working memory are still the subject of active debate (Curtis and Lee 2010, Barak and Tsodyks 2014), the brain areas most often implicated are found within the prefrontal cortex (Figure 1, red) (Barak, Tsodyks et al. 2010). A key line of evidence for the role of PFC in maintaining working memory is the experimental finding that neurons in PFC exhibit sustained responses that are selective for different targets even after the disappearance of the target, a phenomenon known as persistent activity (Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster & Alexander, 1971; Goldman-Rakic, 1996; Kubota & Niki, 1971; Miller, Erickson, & Desimone, 1996; Romo, Brody, Hernandez, & Lemus, 1999). The neural mechanism generally proposed to underlie persistent activity consists of multiple groups of neurons characterized by recurrent excitation and mutual inhibition (Machens, Romo, & Brody, 2005). After the initial activation of the group associated with the active target, recurrent excitatory synapses act to maintain the sustained activity, while inhibitory connections prevent other groups from becoming active as well. Despite the general prevalence of the theory of persistent activity through recurrent connections, it is worth noting the existence of alternative hypotheses postulating the involvement of short-term synaptic plasticity in the maintentance of working memory (Mongillo, Barak, & Tsodyks, 2008; Sugase-Miyamoto, Liu, Wiener, Optican, & Richmond, 2008).

Where in the brain do target-specific signals combine with visual information? Numerous sources of evidence suggest that working memory information is fed back directly into the same areas in the ventral visual pathway that are involved in visual processing, and in particular V4 and IT. First, these areas receive strong inputs from PFC (Markov et al., 2014). The functional role of these projections was directly demonstrated using monkeys who underwent a resection of posterior corpus callosum and anterior commissure. This procedure left intact only the anterior corpus callosum, which connects the prefrontal cortices (Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999). In these animals, neurons in IT and PRH responded to ipsilateral visual cues, which can only be explained by top-down signals from PFC, since visual information could only cross the hemispheres through the anterior corpus callosum. In sum, there is evidence supporting the idea that working memory signals reflecting the identity of the target are fed back to the ventral visual pathway. However, where exactly these signals combine to support invariant object search, as well as the format of the combined signals within the ventral visual pathway, is still unclear.

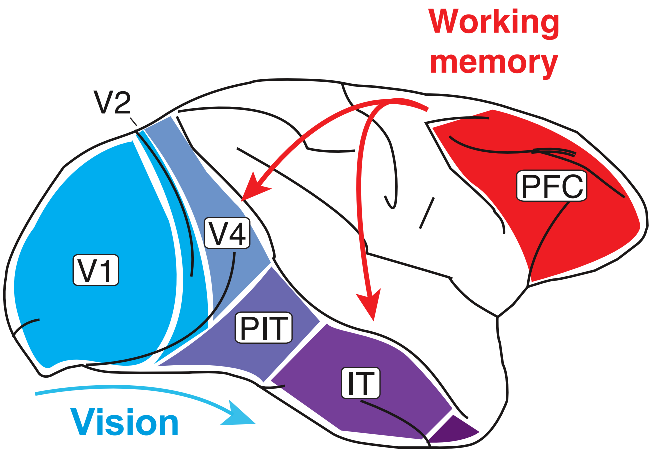

Figure 1-1. *Proposed pathways involved in visual and top-down working memory modulations during invariant object search.* Visual information is first encoded in the retina, where it reaches primary visual cortex (V1) through the lateral geniculate nucleus of the thalamus. Information about the identity of viewed objects is then extracted along the ventral visual pathway, composed by V2, V4, and IT ('Vision', cyan). Working memory information about the identity of the target is thought to be maintained in prefrontal cortex (red). Multiple sources of evidence suggest a top-down projection of this signal to mid to late stages of the ventral visual pathway (e.g. V4 and/or IT) during visual target search.

Attentional modulations in the ventral visual stream

The proposal that context-specific signals are fed back from PFC in order for the combination of contextual and visual signals to happen within visual cortical areas themselves (Figure 1) is further supported by a large body of work describing top-down modulations of visual signals within the ventral visual pathway. Below, we review evidence from single unit neurophysiology studies describing top down modulations in the context of different types of attentional tasks.

Spatial attention:

The most well documented top-down modulations of visual signals have been studied within the context of spatial attention. Allocating attention to a particular spatial location improves the perception of a stimulus at the attended location. This improved perception has been associated with changes in the way that neurons respond to that stimulus. Specifically, in experiments with a single stimulus presented in a neuron's receptive field, attention to that stimulus is usually associated with responses that are faster (Sundberg, Mitchell, Gawne, & Reynolds, 2012), stronger (Maunsell & Cook, 2002; McAdams & Maunsell, 1999) and less variable (Cohen & Maunsell, 2009; Mitchell, Sundberg, & Reynolds, 2007) compared with responses when attention is directed elsewhere. Pairwise correlations in the fluctuations of responses are also typically reduced with the allocation of attention (Cohen & Maunsell, 2009; Herrero, Gieselmann, Sanayei, & Thiele, 2013; Mitchell, Sundberg, & Reynolds, 2009; Zenon & Krauzlis, 2012).

Notably, the magnitude of attentional modulation by spatial cueing differs between visual areas. Specifically, attentional modulation seems to be weakest in the earliest stages of visual cortex, and strongest in later areas (Maunsell & Cook, 2002). However, this increase has not been studied extensively or systematically. A few studies (Luck, Chelazzi, Hillyard, & Desimone, 1997; McAdams & Maunsell, 1999; Moran & Desimone, 1985; though see Motter, 1993) compared modulations in early to mid stages of the ventral visual pathway, and found that firing rate responses in area V4 systematically increased to attended as compared to unattended stimuli (with increases ranging on the order of 25-60% across studies), while responses of cells in V1 reflected no significant or consistent effects of attention. Moran and Desimone (1985) also found increases in firing rates with attention in IT cortex (which were slightly smaller in magnitude to those in V4, possibly due to large variation in the visually-evoked responses in IT to the same stimuli as presented in V1 and V4).

Evidence from these spatial attention studies supports two lines of thought. First, the responses of neurons within the ventral visual pathway can be modulated by top-down contextual signals. Second, comparative studies between early and mid-stage areas imply that V4 might be an important locus of attentional modulation. Alternatively, attentional modulation might continue to gradually increase along the ventral visual pathway. To further examine this, direct comparisons must be made between V4 and downstream areas.

The allocation of spatial attention is only one component of visual search. In particular, attention to particular visual features is important for finding sought targets. Thus, next we review evidence for the involvement of V4 and IT in feature based-attention and visual search.

Feature-based attention and visual search in V4:

Most single unit studies of feature-based attention (attention allocated not to a particular location, but rather to a particular feature) in the ventral visual pathway have focused on V4. One of the first demonstrations of feature-based attentional modulations within V4 neurons was reported by Haenny, Maunsell, and Schiller (1988). Monkeys were shown a sample visual grating which was followed by a series of test visual gratings. They were trained to report when the orientation of the test grating in view matched the previously cued sample grating. The authors found that neural responses in V4 were not only modulated by the orientation of the grating in view (visual modulation) but also by whether the visual stimulus matched the target orientation. Importantly, this modulation by the target orientation is also observed if the animal is cued via a tactile stimulus (by feeling the orientation of a grooved plate which has been hidden from view) (Haenny et al., 1988; Maunsell, Sclar, Nealey, & DePriest, 1991). This was the first study to suggest that feature based attention is mediated by centers that are capable of generating an intermodal representation of orientation.

Motter (1994) trained monkeys to do a task in which they viewed arrays of mixed stimuli and had to attend to a subset of stimuli with a color or luminance that matched a cue stimulus. Most neurons in V4 showed increased responses to the same stimuli when that stimulus matched the cue. However, in these studies, it is possible that the modulation of neural activity depended on a mechanism that targeted spatial locations that were first identified as behaviorally relevant based on color or luminance, and then the effectiveness of the attended stimulus was enhanced relative to the representation of the distractor by spatial attention mechanisms alone. Bichot, Rossi, and Desimone (2005) avoided this issue by training a realistic visual search task where monkeys were allowed to freely move their eyes. In this study, responses in area V4 were recorded during the brief periods between saccades, when a known stimulus lay in the receptive field of the neuron being recorded. Critically, the authors analyzed responses when the target stimulus fell in the receptive field of the recorded neuron but was not yet detected by the animal, who made an eye movement elsewhere. They found enhancements in firing rates (median increase of 30%) when the stimulus in the receptive field matched the cued stimulus. Together, these studies show that the firing rates of V4 neurons can be modulated by target context in feature-based attention.

McAdams and Maunsell (2000) studied the effects of shifting attention between different feature dimensions (rather than specific values of a given feature). In particular, they recorded responses from V4 neurons with a stimulus of their preferred orientation in their receptive field. In one condition, the animal was required to attend to the orientation of another stimulus in a distant location. In a second condition the animal was required to attend to the color of an unoriented stimulus in a distant location. They found that shifting attention between orientations and colors affected the responses of most V4 neurons. This result demonstrates that neural representations of stimuli in parts of the visual field with no relevance to the task can be modulated by attention. This is consistent with the idea that feature based attention changes activity throughout the visual field representation in a way that would be useful for visual search.

Visual search within the late stages of the ventral visual pathway:

In IT and perirhinal cortex (PRH; a downstream area that receives most of its inputs from IT)), visual search has often been studied via delayed-match-to-sample (DMS) tasks, as briefly presented above (Haenny et al., 1988; Maunsell et al., 1991).

This task, first studied by Mishkin, Prockop, and Rosvold (1962), is designed to model the sequential search that commonly occurs when subjects look for an object. Subjects are first cued with a sample image of the sought target. The target then disappears and, after a temporal delay, subjects are required to respond to images that match the target (ignoring intermediately presented distractor images).

As seen in V4, strong task-relevant modulations during DMS have been reported in IT and PRH (Eskandar, Richmond, & Optican, 1992; Leuschow, Miller, & Desimone, 1994; Miller & Desimone, 1994; Pagan, Urban, Wohl, & Rust, 2013). In these areas, however, two subpopulations with opposite responses were described: one with enhanced responses and the other with suppressed responses to target matches as compared to distractor stimuli (Miller & Desimone, 1994). The authors hypothesized that the match-suppressed responses might have arisen as the result of passive, stimulus repetition of the target match following the cue, while the match enhanced neurons alone carry behaviorally relevant information (differentiating between target matches and distractors).

Comparisons between areas within the ventral visual pathway

How do the magnitude of these target search signals differ between V4 and IT? Only one systematic comparison exists between these two areas within the context of a target search task. Chelazzi and colleagues (Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 2001) trained monkeys to perform a visual search task, where monkeys were first cued to their target stimulus, which would subsequently appear at one of two possible locations within the receptive field of the neuron. The receptive field would always contain two stimuli: one that generated a relatively strong response and one that generated a relatively weak response. When the preferred stimulus matched the cue (compared to when the non-preferred stimulus matched the cue), the authors found that firing rate responses increased by similar magnitudes (V4: 63%; IT: 70%). Additionally, a comparison between IT and its projection area (PRH) reported that these areas also have matched amounts of total target match information. These lines of evidence suggest that V4 might act as a locus of combination of visual signals and top down context specific signals, and downstream areas (IT, PRH) simply inherit and reformat this combined task-relevant information.

However, the delayed match to sample tasks described above cued subjects to search for targets that always appeared at the same positions, sizes, and background context. In real world object search tasks, one doesn't know the context in which a target match might appear. Leuschow et al. (1994) trained monkeys to find objects that could appear at different sizes and locations, and found that neurons in IT showed similar modulations when target matches appeared at different sizes and locations. The authors did not compare these signals across areas in the ventral visual pathway, and thus could not determine whether IT inherits these task-relevant signals from V4.

In particular, how might V4 act as a locus for integrating visual and task-dependent signals for an invariant object search task? Flexibly finding different sought objects requires differential responses to the same visual inputs based on task context (as relayed by top-down contextual signals). Given that V4 receptive fields are small and retinotopically organized and consequently, that V4 lacks an explicit representation of object identity, how could the brain determine which subsets of neurons to target with these top-down contextual signals (Maunsell & Treue, 2006)? Whether these signals are

a) combined in V4 and inherited by later areas, or b) integrated gradually along the ventral visual pathway during invariant object search remains unknown.

In sum, despite the substantial impact of these early studies on our understanding of the neural mechanisms underlying target search, many questions still remain open. First, where do top-down target signals integrate into the ventral visual pathway? Does the IT population inherit its combined signals from V4, or do the amounts of these signals increase gradually along the pathway? Second, these areas are known to encode an explicit representation of currently viewed objects: how is this visual representation modulated by cognitive signals? Specifically, in what format do these modulations appear? We address these first two questions in Chapter 2.

How signal and noise might contribute object search performance

While Chapter 2 focuses on the locus and mechanism of combination of visual and context-specific signals, Chapter 3 considers these questions from a neural coding perspective. Specifically, how do signal and noise contribute to task performance and how might this inform how the brain multiplexes signals during invariant object search?

Performance on a particular task is determined not only by the amount of task-relevant signal reflected by neurons (e.g. information about whether a target match or distractor is in view), but also by the presence of noise, which can arise from multiple sources. Internal noise, or “trial variability” manifests as trial-by-trial variations in neural responses under seemingly identical conditions. External factors can also translate into noise, particularly when a task requires extracting a particular type of information from our environment amid changes in other task-irrelevant, nuisance parameters (Kim, Pitkow, Angelaki, & DeAngelis, 2016). Stated differently, for any given task, neurons in a brain area may be modulated by multiple experimental variables, but when viewed from the perspective of task performance, one type of modulation reflects the task-relevant signal, whereas other types of modulation act as noise.

Mixed selectivity in the ventral visual pathway:

The existence of cognitive modulations within the same areas known for solving strictly visual tasks (e.g. object recognition), also known as mixed selectivity, poses a potential challenge for the brain. When viewed from the perspective of task performance, one type of modulation reflects the task-relevant signal, whereas other types of modulations (e.g. modulations by visual information) act as noise. Specifically – a population of neurons whose responses are modulated by whether an object is a target match or a distractor is expected to perform worse at a simply visual discrimination task, and vice versa.

Outside of the realm of attentional modulations, growing evidence suggests that different types of signals are in fact mixed, both at the locus at which task-relevant solutions are computed as well as downstream (Freedman & Assad, 2009; Kobak et al., 2016; Mante, Sussillo, Shenoy, & Newsome, 2013; Meister, Hennig, & Huk, 2013; Raposo, Kaufman, & Churchland, 2014; Rigotti et al., 2013; Rishel, Huang, & Freedman, 2013; Zoccolan, Kouh, Poggio, & DiCarlo, 2007). A number of explanations have been proposed to account for mixed selectivity. Some studies have documented situations in which signal mixing is an inevitable consequence of the computations required for certain tasks, such as identifying objects invariant to the view in which they appear (Zoccolan et al., 2007). Others have suggested that mixed selectivity may be an essential component of the substrate required to maintain a representation that can rapidly and flexibly switch with changing task demands (Raposo et al., 2014; Rigotti et al., 2013). Still others have maintained that broad tuning across different types of parameters is important for learning new associations (Rigotti et al., 2013). Thus it may be the case that one or all of these benefits outweigh the performance costs associated with mixed selectivity. Alternatively, it may be that mixed selectivity is not as detrimental to task performance as it otherwise appears. As described in more detail in Chapter 3, our results support the latter assertion.

Overview

In Chapter 2, we sought to compare the responses of neurons in V4 and IT, to determine whether V4 acts as a singular locus of combination for context-specific and visual signals, or rather that these signals are injected at multiple stages of the ventral visual pathway. In this study, we provide evidence that during invariant object search, while context-specific modulations exist in both V4 and IT, they are larger in IT. Furthermore, we show that at the level of single units, these signals are formatted differently in V4 and IT. These results reveal that top-down, context-specific signals are integrated into the ventral visual pathway at multiple stages during invariant object search.

In Chapter 3, we focus on the responses in IT cortex during the invariant object search task. Specifically, we use the data from IT as a case study to answer a neural coding question. The task performance of neurons whose responses are modulated both by task-relevant signal (e.g. responses that differentiate between target matches and distractors) and also task-irrelevant factors that act as noise (nuisance variability; e.g. visual information about the object currently in view) should be limited by these nuisance modulations. Our results reveal that surprisingly, while nuisance modulation was large in IT cortex, its impact on task performance, both within single units and at the level of the population, was modest. This result could be explained by the existence of large trial variability constrained by short, behaviorally relevant spike counting window. In sum, our results reveal that when the brain operates in a fast processing, low spike count regime, nuisance modulations are largely inconsequential for task performance.

Finally, in Chapter 4 we discuss how our results relate to the existing literature, and speculate about possible future directions for this research.

CHAPTER 2

The multi-stage integration of visual and target signals during object search

ABSTRACT

Many everyday tasks require our brains to flexibly map incoming sensory information onto different behavioral responses based on context. One example is the act of searching for a specific object, which requires us to compare the items in view with a remembered representation of a sought target to determine whether a target match is present. During visual search, top-down modulations reflecting target identity are known to combine with feed-forward visual representations at mid-to-higher stages of the ventral visual or form processing pathway (e.g. V4 and inferotemporal cortex, IT). However, it remains unclear whether these top-down signals are inserted at a single locus (e.g. V4) or whether they are inserted at multiple stages (e.g. both V4 and IT). To investigate, we systematically compared neural responses in V4 and IT recorded as two monkeys performed a task that required them to identify when a target object appeared across variation in the objects' positions, sizes and background contexts. We found that while average context-specific modulation was considerable in V4 (35% the size of visual modulation), it was even larger in IT (72%), and consequently, total information about the target match solution was larger in the IT as compared to the V4 population. Additionally, in V4, modulations reflected changes in the identity of the sought target (i.e. working memory signals), whereas in IT they were a heterogeneous mixture of working memory signals and signals reflecting the task solution (i.e. whether an object is a target match or a distractor). Together, these results suggest that during object search, top-

down, task-relevant signals are combined with feed-forward visual information at multiple stages along the ventral visual pathway.

INTRODUCTION

Finding a specific sought object, such as our car keys, requires our brain to execute at least two types of non-trivial computations. First, we must determine the identities of the objects in view, across variation in details including their position, size, and background context. Additionally, we must perform a comparison of the visual representation of what we are looking at with a remembered representation of what we are looking for to determine whether our target is in view. Considerable evidence suggests that computations in the primate ventral visual pathway, including visual brain areas V1, V2, V4 and IT, support the process of invariant object recognition (reviewed by DiCarlo, Zoccolan, & Rust, 2012; Ungerleider & Mishkin, 1982). Evidence also suggests that signals reflecting the combination of visual and target information (e.g. differential responses to the same images presented as target matches versus as non-target distractors) are reflected in V4 (Bichot et al., 2005; Chelazzi et al., 2001; Haenny et al., 1988; Kosai, El-Shamayleh, Fyall, & Pasupathy, 2014; Maunsell et al., 1991), IT (E.N. Eskandar et al., 1992; Gibson & Maunsell, 1997; Leuschow et al., 1994; Liu & Richmond, 2000; Meunier, Bachevalier, Mishkin, & Murray, 1993; Pagan et al., 2013), and perirhinal cortex (Miller & Desimone, 1994; Pagan et al., 2013). However, exactly where and how the comparison of visual and target identity is performed is not well understood.

Here we present two idealized proposals for how top-down target modulation might be integrated within the ventral visual pathway during visual target search. In the first proposal (Fig 1a), these signals emerge gradually and increase in strength along the visual hierarchy, as a result of multiple stages in which top-down target modulation is combined with feed-forward visual information. In a second proposal (Fig 1b), a single brain area (e.g. V4) serves as the locus of the combination of visual and target information, and higher brain areas receive information about target identity by way of inheriting this information from earlier stages. While the gradient proposal (Fig 1a) is broadly-assumed to be true, published evidence is more supportive of V4 as a single locus (Fig 1b) for receiving top-down signals during visual target search. For example, under the most comparable conditions published to date, average target modulations in V4 and IT were reported to be large and similar in magnitude (63% versus 70% of the visually-evoked response; Chelazzi et al. (1998); Chelazzi et al. (2001).) Other measures of target modulation magnitudes in V4 (Bichot et al., 2005; Haenny et al., 1988; Maunsell et al., 1991) and IT (Chelazzi, Miller, Duncan, & Desimone, 1993; Miller & Desimone, 1994) report comparable values. Additionally, a comparison between two higher stages of the pathway (IT and its projection area, perirhinal cortex) reported that differences between these two brain areas were reflected as differences in the format of target match information format, as opposed to overall amounts, consistent with a feed-forward process (Pagan et al., 2013). In contrast, reports of modulation magnitudes in brain areas that lie earlier in the pathway (e.g. V1 and V2) are consistently smaller than those reported for V4 (Luck et al., 1997; McAdams & Maunsell, 1999; Moran & Desimone, 1985). Together, these results suggest that V4 may indeed act as a locus in the ventral visual pathway for receiving top-down, context-specific signals.

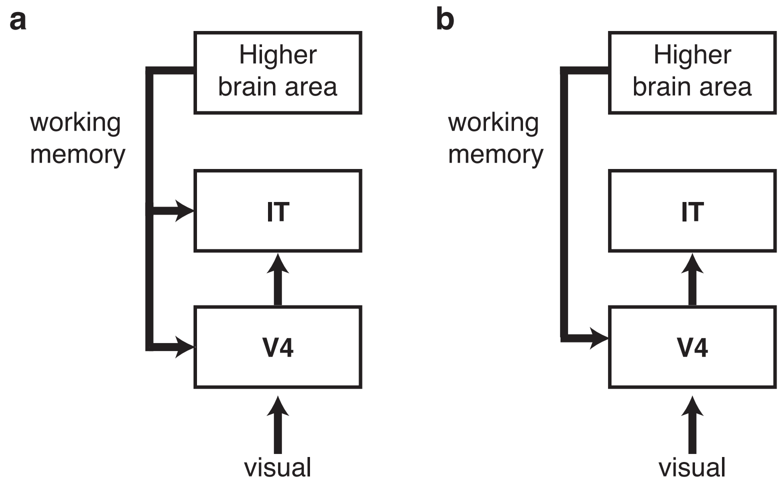

Figure 2-1. *Where is target information combined with visual information along the ventral visual pathway?* Discriminating between classes of models of where the target match signal is computed **a)** Schematic showing target information being fed back to V4 and combined there. In this model, this combined information (target match signal) is inherited by IT cortex. **b)** Schematic showing target information being fed back to multiple stages of the ventral visual pathway (V4, IT cortex). In this model, the target match signal is larger in IT than in V4.

While compelling, one mystery associated with accounts that V4 might act as a singular locus for integrating visual and task-dependent signals relates to the question of how the brain might achieve this. Flexibly finding different sought objects requires differential responses to the same visual inputs based on task context (as relayed by top-down contextual signals). Given that V4 receptive fields are small and retinotopically organized and consequently, that V4 lacks an explicit representation of object identity, how could the brain determine which subsets of neurons to target with these top-down contextual signals (Maunsell & Treue, 2006)? Notably, only one study has examined the real-world problem of visual object search in the context of an object that can appear under different identity-preserving transformations (Leuschow et al., 1994) and only in IT. Moreover, no study has systematically compared signals in V4 and IT during object search in the same region of the visual field, using the same images, in the same monkeys, performing the same task. This is thus what we set out to do.

RESULTS

The invariant delayed-match-to-sample task (IDMS)

To compare the amount and format of target match information between V4 and IT, we trained two monkeys to perform an "invariant delayed-match-to-sample" (IDMS) task that required them to report when target objects appeared across variation in the objects' positions, sizes and background contexts. In this task, the target object was held fixed for short blocks of trials (~3 minutes on average) and each block began with a cue trial indicating the target for that block (Fig 2a, "Cue trial"). Subsequent test trials always began with the presentation of a distractor and on most trials this was followed by 0-5 additional distractors (for a total of 1-6 distractor images) and then an image containing the target match (Fig 2a, "Test trial"). The monkeys' task required them to fixate during the presentation of distractors and make a saccade to a response dot on the screen following target match onset to receive a reward. To minimize the predictability of the match appearing as a trial progressed, on a small subset of the trials the match did not appear and the monkey was rewarded for maintaining fixation. Our experimental design differs from other classic DMS tasks (Chelazzi et al., 1993; Eskandar, Optican, &

Richmond, 1992; Leuschow et al., 1994; Miller & Desimone, 1994; Pagan et al., 2013) in that it does not incorporate a cue at the beginning of each test trial, to better mimic real-world object search conditions in which target matches are not repeats of the same image presented shortly before.

Our experiment included a fixed set of 20 images, broken down into 4 objects presented at each of 5 transformations (Fig 2b). Our goal in selecting these specific images was to make the task of classifying object identity challenging for the IT population and these specific transformations were built on findings from our previous work (Rust & DiCarlo, 2010). In any given block (e.g. a squirrel target block), a subset of 5 of the images would be considered target matches and the remaining 15 would be distractors (Fig 2b). Our full experimental design amounted to 20 images (4 objects presented at 5 identity-preserving transformations), all viewed in the context of each of the 4 objects as a target, resulting in 80 experimental conditions (Fig 2c). In this design, "target matches" fall along the diagonals of each looking at / looking for matrix slice (where "slice" refers to a fixed transformation; Fig 2c, gray). For each condition, we collected at least 10 repeats on correct trials. Monkeys generally performed well on this task (Fig 2d). Their mean reaction times (computed as the time their eyes left the fixation window relative to the target match stimulus onset) were 364 ms and 324 ms (Fig 2e).

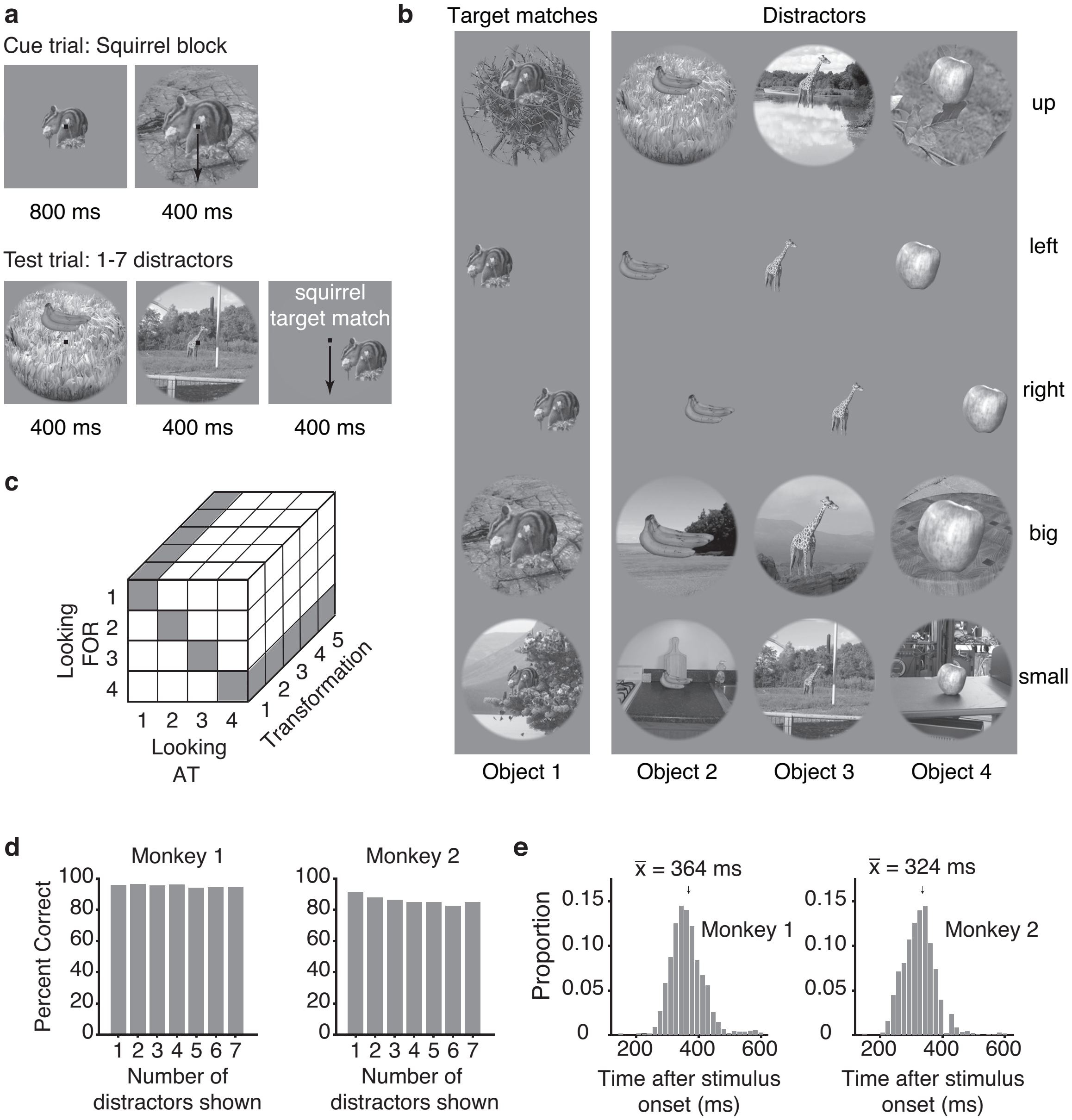

Figure 2-2. *The invariant delayed-match-to-sample task.* **a)** Monkeys performed an invariant delayed-match-to-sample task. Each block (~3 minutes in duration) began with a cue trial indicating the target object for that block. On subsequent trials, monkeys initiated a trial by fixating on a small dot. After a 250 ms delay, a random number (1-7) of distractors were presented, and on most trials, this was followed by the target match. Monkeys were required to maintain fixation throughout the distractors and make a

saccade to a response dot within a window 75 - 600 ms following the onset of the target match to receive a reward. In cases where the target match was presented for 400 ms and the monkey had still not broken fixation, a distractor stimulus was immediately presented. **b)** The experiment included 4 objects presented at each of 5 identity-preserving transformations (“up”, “left”, “right”, “big”, “small”), for 20 images in total. In any given block, 5 of the images were presented as target matches and 15 were distractors. **c)** The complete experimental design included looking “at” each of 4 objects, each presented at 5 identity-preserving transformations (for 20 images in total), viewed in the context of looking “for” each object as a target. In this design, target matches (highlighted in gray) fall along the diagonal of each “looking at” / “looking for” transformation slice. **d)** Percent correct for each monkey, calculated based on both misses and false alarms (but disregarding fixation breaks). Percent correct is plotted as a function of the number of distractors shown. **e)** Histograms of reaction times during correct trials (ms after stimulus onset) during the IDMS task for each monkey, with means indicated by arrows and labeled.

To systematically compare the responses of V4 and IT during this task, we applied a population-based approach in which we fixed the images and their placement in the visual field across all the units that we studied, and we sampled from representative units whose receptive fields overlapped the stimuli we presented. Specifically, we presented images at the center of gaze, with a diameter of 5 degrees. Neurons in IT typically have receptive fields that extend beyond 5 degrees and extend into all four quadrants (Fig 3a top; Op De Beeck and Vogels, 2000). In contrast, V4 receptive fields are smaller, retinotopically organized, and confined to the contralateral hemifield (Fig 3a bottom; Desimone and Schein, 1987, Gattass et al. 1988). To compare these two brain areas, we applied extensions of approaches developed in our earlier work (Rust & DiCarlo, 2010) in which we compared the responses of a set of randomly sampled IT units with a population of V4 units whose receptive fields tiled the image (Fig 3b). This required sampling V4 units with receptive fields in both upper and lower visual fields, which we achieved through recording at different positions within and around the inferior occipital sulcus. This also required measuring units with receptive fields on both sides of the vertical meridian, which we approximated by isolating our recordings to one hemisphere but reflecting the images along the vertical axis in approximately half the sessions (see Methods).

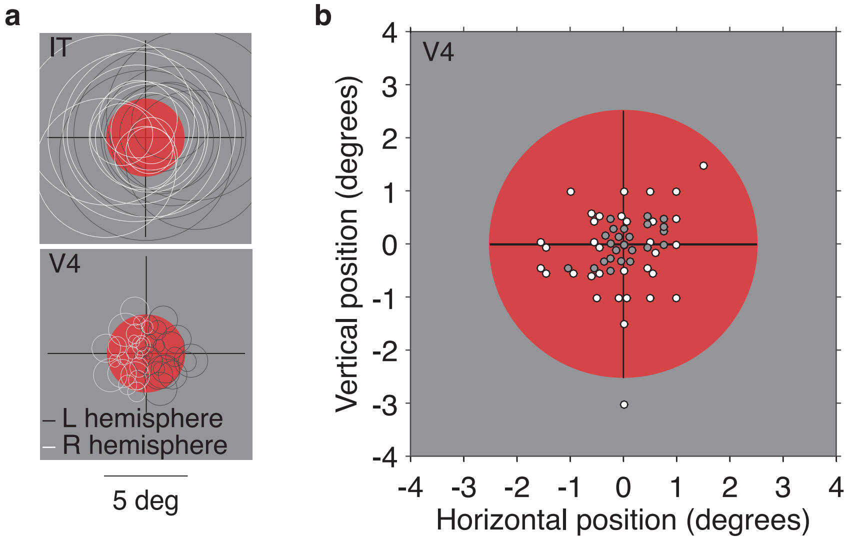

Figure 2-3. *Experimental design: V4 - IT comparisons* **a)** Images were displayed at the center of gaze and were 5 degrees in diameter (red circle indicates location and size of images.) Expected receptive field locations and sizes for neurons in V4 (top; Desimone and Schein, 1987; Gattass et al., 1988), and IT (bottom; Op De Beeck and Vogels, 2000). **b)** We targeted V4 neurons such that their receptive fields tiled the image. The receptive field locations of V4 neurons recorded for each session. If a session included

more than one receptive field location, all are included. Dots illustrate the center of each receptive field (gray, Monkey 1; white, Monkey 2).

Because V4 receptive fields in the region of the field we recorded are small, one potential issue of concern is the replicability of retinal image placement across trials. We quantified the stability of monkeys' eye positions across repeated trials as the percent of eye positions that were within windows corresponding to V4 receptive field sizes at the range of eccentricities we recorded (Gattass et al., 1988). We found that 89% of eye positions were reliably within windows corresponding to the average RF sizes at the fovea (0.56 degrees), and 98% of eye positions were within windows corresponding to RF sizes at an eccentricity of 2.5 degrees (1.4 degrees). To achieve this in Monkey 2, fixational control was improved by aligning the images closer to the center of gaze at stimulus onset (see Methods). These approaches were effective in producing similar distributions of trial-by-trial variability between V4 and IT, as measured by the mean and standard deviation of Fano factor across units (mean +/- std, V4 = 1.41+/-0.3; IT = 1.35 +/- 0.33).

As two monkeys performed this task, we recorded neural activity from small populations using 24-channel probes that were acutely lowered into V4 or IT before each session. In all of our analyses, we counted spikes in a 170 ms window (V4: 40-210 ms; IT: 80-250 ms following stimulus onset), which always preceded the monkeys' reaction times and thus corresponded to periods of fixation. The data reported here were extracted from trials with correct responses. To create comparable populations in V4 and IT, we first screened for units based on their stability, isolation, and task modulation. By design, we recorded more units in V4 than IT, and to compare them, we randomly subsampled neurons in V4 (see Methods). Most of our analyses are applied to pseudopopulations that are matched in size (n = 193 units in each area; Monkey 1, n = 98 and Monkey 2 n = 95).

Equating the recorded V4 and IT populations for total visual information

There are important factors to consider when making a systematic comparison between V4 and IT. For example, should V4 and IT be compared one-to-one with equal numbers of units? How does one know if the samples from two brain areas accurately reflect differences between them? As an illustrative example, imagine a scenario in which the V4 neurons sampled all had small, overlapping receptive fields confined to the same small region of the visual field whereas IT neurons, by virtual of their large receptive fields, had access to much more of the visual field. From this data we might erroneously conclude that total target match information is higher in IT than V4 by way of sampling, whereas in reality the two brain areas might actually contain matched amounts of target match information.

In addition to overall differences in receptive field size, we considered several factors when systematically comparing V4 and IT. First, we note that all the visual information in IT is thought to arrive there after first traveling through V4 and total visual information in IT thus cannot exceed information in V4. As such, one reasonable benchmark for assessing whether two recorded populations are comparable is by an assessment of whether the two populations have matched amounts of total information about visual identity. Second, it is also the case that the format of visual information is known to differ between V4 and IT insofar as information about object identity (across changes in object position, size and background context) is more accessible to a linear read-out whereas in V4 that information is more nonlinear. One way to circumvent this issue is to make comparisons of the amounts of visual information at each transformation separately. Finally, under this line of thinking, the question about how many units to include in the V4 as compared to the IT population is somewhat of an empirical one, as the right answer is determined by the number of units required to equate total visual information in the two populations; in our previous work, we found that the two populations could be equated with approximately equal numbers of units (Rust & DiCarlo, 2010).

To assess the degree to which total visual information in our recorded V4 and IT populations was matched, we quantified the ability of each population to discriminate between the 4 images computed separately for each of the 5 transformations (Fig 4a, right; see Methods). For 3 out of the 5 transformations, visual information was well-matched between V4 and IT in each monkey when equal numbers of units were considered, both when averaged across all the transformations (Fig 4d) as well as when each transformation was considered individually (Fig 4e). For 2 of the transformations ("left" and "right"), the V4 population had significantly lower performance than V4 for the other 3 transformations ("big", "small", "up"), and V4 performance on the visual identification task was considerably lower than IT (not shown). This is consistent with the absence of more peripheral receptive fields (Fig 3b) in our data. We thus focused further analysis on the 3 of 5 transformations in which we were confident total visual information was equated in our samples of V4 and IT.

Target match information is lower in V4 than IT

To compare V4 and IT, we asked each population to solve the task that the monkey had to solve: a two-way classification between the same images presented as target matches (Fig 4c, gray) versus as distractors (Figure 4c, black). To compare total amounts of target match information in V4 and IT, we measured cross-validated performance of a maximum likelihood classifier to perform this 2-way classification at each transformation separately and then averaged over transformations (see Methods). We found that the cross-validated population performance was higher than chance in V4, but was even higher in IT (Fig 4e; pooled data: p<0.005), and this result was confirmed in each monkey individually (Figure 4f filled points, monkey 1 p<0.005; monkey 2, p=0.007) These results suggest that IT target match information is not exclusively inherited from V4, and they are consistent with descriptions in which top-down, task-relevant signals are integrated in IT (as well as V4; Fig 1a).

The maximum likelihood classifier is designed to measure total target match information regardless of its format (e.g. linear or nonlinear). To determine how much of this total information was formatted in a manner accessible to a linear population read-out, we also computed the performance of a linear classifier (Fig 4d; a Fisher Linear Discriminant, see Methods). Like total information, this measure of linearly separable information was higher in IT than V4 (Fig 4f white dots, monkey 1, p<0.005; monkey 2, p<0.005). In summary, both when assessed by the performance of a maximum likelihood or linear classifier, IT performance at differentiating between target matches versus distractors was larger than that in V4.

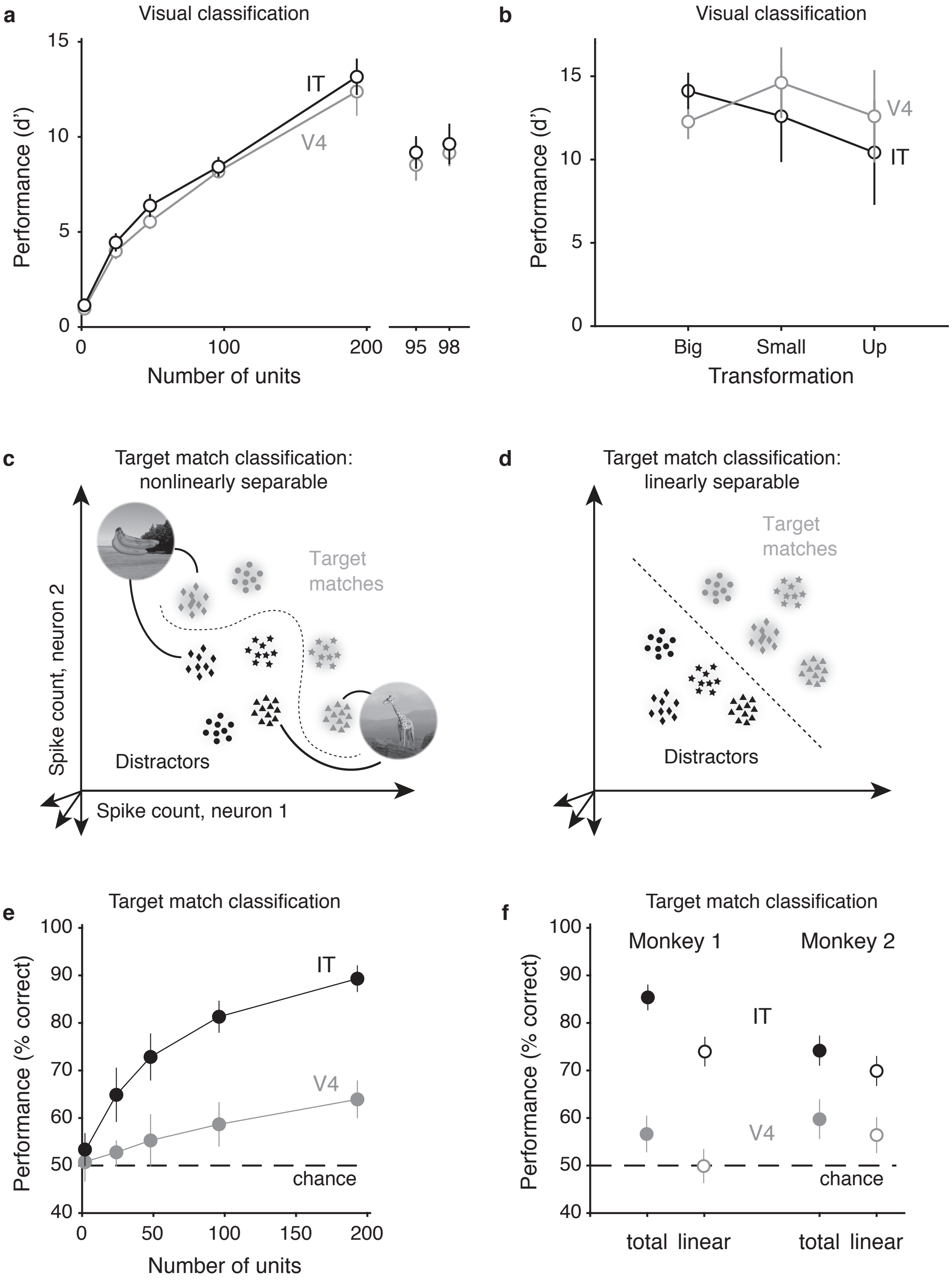
a
Visual classification
Performance (d')
15
10
5
0
IT
V4
0
100
200
95
98
Number of units
b
Visual classification
Performance (d')
15
10
5
0
V4
IT
Big
Small
Up
Transformation
c
Target match classification:
nonlinearly separable
Target
matches
Distractors
Spike count, neuron 2
Spike count, neuron 1
d
Target match classification:
linearly separable
Target
matches
Distractors
e
Target match classification
Performance (% correct)
100
90
80
70
60
50
40
IT
V4
chance
0
50
100
150
200
Number of units
f
Target match classification
Performance (% correct)
100
90
80
70
60
50
40
Monkey 1
Monkey 2
IT
V4
chance
total
linear
total
linear

Figure 2-4. *A comparison of target match information in V4 and IT* **a)** Visual discrimination in V4 and IT was matched. Shown is performance as a function of the number of neurons for the V4 and IT populations assessed by a linear readout of object identity. The performances are averaged across a subset of the images used in the experiment (3 of the 5 transformations which elicited high discriminability in V4). This performance is shown for pseudopopulation across both monkeys, n = 193 units, left, and for each monkey individually, right. Error bars (standard error) reflect the variability that can be attributed to the specific subset of trials chosen for training and testing, and, for subsets of neurons smaller than the full population, the specific subset of neurons chosen. **b)** Visual discrimination in V4 and IT, shown for each transformation individually at the total number of units (n = 193). **c)** The target search task can be envisioned as a two-way classification of the same images presented as target matches versus as distractors. Shown are cartoon depictions where each point depicts a hypothetical population response for a population of two neurons on a single trial, and clusters of points depict the dispersion of responses across repeated trials for the same condition. Included are responses to the same images presented as target matches and as distractors. The dotted line depicts a hypothetical decision boundary. **d)** Same as in (c), but dotted line depicts a hypothetical linear decision boundary. In this schematic, the target matches versus distractors are linearly separable. **e)** Target match information is higher in IT than in V4. Total amount of target match information, as assessed by the performance of an ideal observer trained to classify between whether an object was a match or a distractor, invariant of the object's identity and transformation. Total information was higher in IT than V4 (p <0.005) in a pseudopopulation across both monkeys, n = 193 units. Error bars (standard error) reflect the variability that can be attributed to the specific subset of trials chosen for training and testing, and, for subsets of neurons smaller than the full population, the specific subset of neurons chosen. Dashed line indicates chance performance. **f)** Total target match information was higher in IT than V4 in each monkey individually (filled points). Total target match information could be mostly accounted for by linearly separable target match information in both V4 (compare gray open points and gray filled points) and IT (compare black open points and black filled points) in each monkey individually.

Single units in V4 and IT differ in both their amount and format of context-specific modulations

Because the approach presented thus-far is from the perspective of population coding and from the somewhat abstract perspective of total information, we were interested in also arriving at more intuitive, single-unit descriptions of the types of firing rate modulations that give rise to differences between V4 and IT. To do so, we return to the experimental design of the IDMS task (Fig 5a). We first consider the responses of a neuron to different conditions within one slice of this matrix (corresponding to one transformation; Fig 5a), where each slice corresponds to viewing each of four objects ('Looking AT') in the context of each of four target objects ('Looking FOR'). Different types of task modulation produce distinct structure in these response matrices: visual modulation translates to vertical structure (Fig 5a, 'visual'), target identity modulation translates to horizontal structure (Fig 5a, 'target identity') and nonlinear combinations of these visual and target identity signals translate to diagonal structure (Fig 5a, 'target match' or equivalently, differential responses to the same images presented as target matches versus as distractors. We note that target match modulation corresponds to a nonlinear combination of visual and target identity, and can be instantiated as units that report whether one particular object is a target match or a distractor (Fig 5a, ‘selective target match detector’) or units that report whether an object is a target match, invariant to the identity of the object in view (Fig 5a, ‘four-object target match detector’).

To quantify the amounts of different types of task-relevant modulations, we applied a bias-corrected procedure that quantifies different types of modulation in terms of the number of standard deviations around each unit’s grand mean spike count (Pagan & Rust, 2014b). Modulation types were grouped into intuitive sets as described above (e.g. visual, target, and target match modulation) as well as “residual” modulations attributed to nonlinear interactions between the visual stimulus and target that were not captured by target match modulation (e.g. specific distractor conditions). Figures 5b,e illustrate these modulations when computed as a function of time relative to stimulus onset and averaged across units. In line with our population results, we found that across units in V4 and IT, visual modulation (compare Fig 5b and e, red) was of comparable size. Furthermore, target match modulation (compare Fig 5a and e, dashed gray) was large in IT but small in V4. In both V4 and IT, we found a signal reflecting information about the target identity (Fig 5b and e, solid gray). This signal appears before stimulus onset, suggesting that it reflects persistent working memory information about the target identity on each trial. Note that because the IDMS task cued monkeys to the identity of the target at the beginning of each block, we expect target identity information to be present before the onset of each presented stimulus. Lastly, we found that in both V4 and IT, residual modulation was small (Fig 5b and e, dotted gray.)

To more directly compare these measures with our population results, we quantified the modulation amounts in the spike count window used for population analysis (Fig 5b,e, gray rectangle). To gain insight into the total amount of context-specific modulation of visual signals, we also plot total cognitive modulation, comprised of all non-visual modulation types. We found that total cognitive modulation was smaller in V4 than in IT in each monkey (Fig5 c-d, f-g, dark gray; 0.26x versus 0.71x the visual modulation in Monkey 1, $p<0.001$; 0.53x versus 0.78x the visual modulation in Monkey 2, $p=0.013$). We next parsed total cognitive modulation into different types (Fig5 c-d, f-g, light gray). In V4, the cognitive modulation was comprised of mostly target identity modulation and low target match modulation (target identity modulation was 0.29x the visual modulation versus target match modulation was 0.07x the visual modulation in Monkey 1; target identity modulation was 0.57x the visual modulation versus target match modulation was 0.22x the visual modulation in Monkey 2). In contrast, in IT, cognitive modulation was comprised of similar amounts of target identity modulation as in V4, but this target identity modulation was roughly matched to large target match modulation (target identity modulation was 0.55x the visual modulation and target match modulation was 0.44x the visual modulation in Monkey 1; target identity modulation was 0.43x the visual modulation and target match modulation was 0.57x the visual modulation in Monkey 2.). In sum, while both V4 and IT units reflect cognitive modulations, they were larger in IT. In V4, cognitive modulations were comprised of pure target identity signals, while in IT they were mixtures of target identity and target match modulations.

How do these single unit modulations relate to the population-based measures presented in Figure 4? Figure 4 displays measures of total information for this task, which requires both visual and cognitive signals. Because visual signals are larger than cognitive signals (Fig 5b-g), cognitive modulations act as the informational bottleneck for the ability of these populations to contribute to total task-relevant information. Thus we can think of average total cognitive modulations as a proxy for the population total information. In contrast, population linear classifier performance is constrained by the amount of explicit information differentiating between target matches and distractors. In fact, there is an analytical mapping between the single unit target match modulations and the population linear classifier performance (Pagan & Rust, 2014b; discussed in more detail in Chapter 3). Stated differently, a population with large amounts of target match modulations (Fig 5a, diagonal structure) will contain linearly separable information, whereas a population with all of its cognitive information formatted as target identity modulations (Fig 5a, horizontal structure) will reflect all of its information in a nonlinearly separable format. To illustrate this link between single unit quantifications and our population results, we replotted the modulation breakdown shown in Fig 5c-d, 5f-g to compare with the population based measures. We found that indeed, for each monkey and each brain area, these modulation amounts align with the differences we find at the level of the population (compare Fig 4f and 5h).

On average across monkeys, we found that while V4 contains cognitive modulations (on average of size 35% of its visual signal), IT contains more (on average of size 72% of its visual signal). While V4 and IT have similar amounts of modulation by target identity (i.e. horizontal structure in the response matrix, see Fig 5a modulation subtypes, center; V4: 39% versus IT: 52% of the visual signal), the discrepancy in total cognitive modulation between V4 and IT comes from a difference in target match information (i.e. diagonal structure in the response matrix see Fig 5a modulation subtypes, right). V4 contains much less target match information (12% of its visual signal), while IT contains large target match information (48% of its visual signal).

To summarize these results, we found that at the single unit level, V4 and IT differ in their total amounts of cognitive modulation. While V4 contains some cognitive modulation, IT contains more. Furthermore, the cognitive modulations that V4 contains are reflected purely as target identity, or working memory signals, while in IT cognitive modulations were comprised of both target identity modulations and target match signals. Therefore, while IT might inherit some of its information from the large V4 target identity modulations, there remain large cognitive modulations in IT that are not present in V4. These results thus suggest that top-down cognitive information is integrated into the ventral visual pathway at multiple stages.

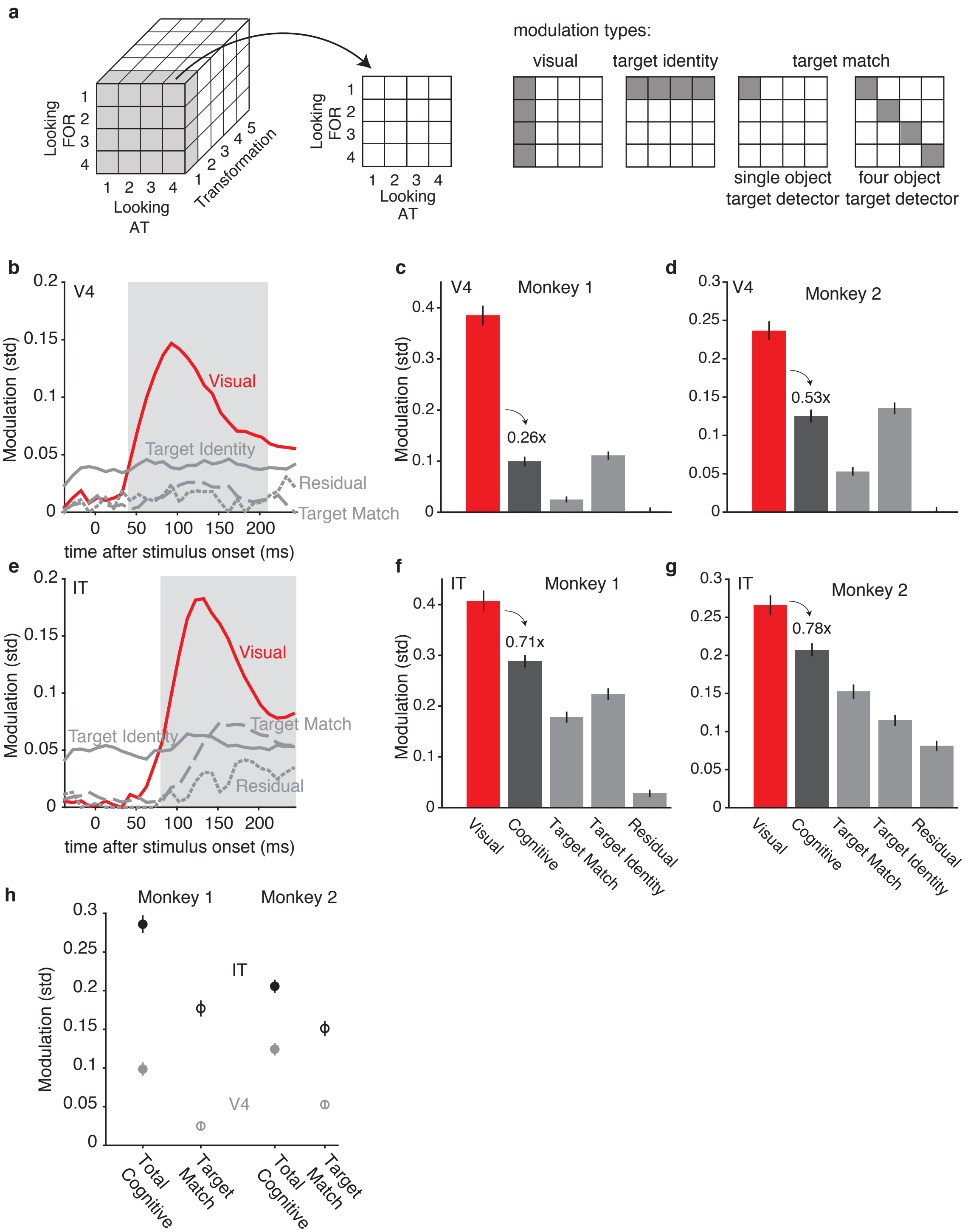

Figure 2-5. *Average single unit modulations in V4 and IT cortex.* Modulations were computed for each type of experimental parameter, in units of the standard deviations

around each unit’s grand mean spike count (see Results). **a)** The IDMS experimental design (see Figure 2c) shown for one particular transformation (left) and 3 different possible response modulation types (right). Shown are visual modulations, which differentiate between different objects in view (vertical structure), target identity modulations, which differentiate between different target objects (horizontal structure), and target match modulations, which differentiate between whether an object (single object target detector) or all objects (four object target detector) appear as a target match versus a distractor (diagonal structure). **b)** Average modulation magnitudes across units in V4 (n=193) shown as a function of time (ms after stimulus onset). Each unit’s firing rate responses are parsed into visual modulation (red) target identity modulation (gray solid), target match modulation (gray dashed), and residual modulation (gray dotted). Spike counting window used for analyses is indicated by the gray rectangle. **c)** Modulations in (a), for Monkey 1 only, as computed during the spike counting window. **d)** Same as in (b), for Monkey 2 only. **e-g)** Same as in (b-d), for the IT population. **h)** Average summed modulation magnitudes across units in V4 (gray) and IT (black) for individual monkeys, replotted from “cognitive” and “target match” modulations in (c-d, f-g). “Total cognitive” is defined as the combination of “target match”, “target identity” and “residual” modulation (for each unit, this was computed as the square root of the sum of the squares of target identity, target match and residual modulation, and then averaged across units as for all modulation types), and corresponds to the total information for the target search task (Figure 4b, left); Target match corresponds to the linearly separable information for the target search task (Figure 4b, right).

Cognitive modulations in V4 are consistent with previous studies

The existence of large cognitive modulation in V4 reflected as a persistent working memory that exists before image onset was surprising in light of other types of attentional modulations (e.g. spatial attention) that are reported to be largely reflected in V4 as multiplicative modulations on top of the image-evoked response (e.g. McAdams & Maunsell, 1999). We thus applied the same measures reported in a number of the most comparable reports. Maunsell et al. (1991) performed an experiment where monkeys were cued to their target orientation via a tactile stimulus (by feeling the orientation of a grooved plate, and were then shown a series of visual gratings. They were required to report when a visual grating matching the sample orientation appeared. To quantify the amount of modulation by the identity of the target, the authors computed a target preference index as the difference in mean firing rate to the preferred target compared to the least preferred target, divided by their sum. They reported this index for all units in their recorded population that were significantly modulated by the identity of the target via a 2-way ANOVA. We applied the same screen and computed the target preference index for our data, and found a median index value of 0.37 (Fig 6a), compared to a 0.31 in their study. Notably, this large median value in V4 can be explained by the fact that this index purely measures modulation by target identity. That is, a high value can be explained by units whose response matrices have purely horizontal structure (Fig 5a, 'target identity'), and does not require target match information (Fig 5a, 'target match'). Haenny et al. (1988) performed the same experiment as described above, and computed a different modulation index which quantifies the differences in firing rates to target matches compared to distractors. Specifically, this index was computed as the difference in mean firing rate to target matches (averaged across preferred and least preferred image) compared to the mean firing rate to distractors (averaged across preferred and least preferred image), divided by their sum. In our data, the absolute modulation preference index was 0.16 (Fig 6b) compared to a 0.26 in their study. , and in both our and their data, there exist mixtures of units that increased their firing rates to matches and those that decreased their firing rates to matches (though this index was shifted towards enhanced units in both our data, p = 0.009, and their data). Notably, this index could be explained by units whose responses reflect linear combinations of visual modulation and target identity modulation (i.e. Fig 5a 'visual', 'target identity'). That is, a high value of this index does not necessarily imply that a unit's responses reflect nonlinear combinations of these inputs (i.e. Fig 5a, 'target match'). While these two studies did not record in IT, for comparison, we also computed these values for our IT data (Figure 6d-e). In both V4 and IT, we found mixes of target match enhanced and suppressed units (Fig 6 b,e), but both populations showed mostly target matched enhanced units as assessed by a significant rightward shift from zero in the match enhancement index (V4 p = 0.009, IT $p < 10^{-5}$). Importantly, for both measures, the distributions of indices were shifted rightward in IT compared to those in V4, showing increased firing rate modulations in IT compared to V4 (the target preference index was greater in IT than V4, p = 0.024; the match enhancement index was greater in IT than V4, $p < 10^{-5}$).

We next compared modulations in our V4 and IT data to those found by Chelazzi et al. (1998); Chelazzi et al. (2001). This series of studies trained monkeys to perform a visual search task, where monkeys were first cued to their target stimulus, which would subsequently appear at one of two possible locations within the receptive field of the neuron. The receptive field would always contain two stimuli: one that generated a relatively strong response and one that generated a relatively weak response. The authors computed a target effect index as the difference between the mean firing rate to each unit's preferred image as a target and the mean firing rate to the unit's least preferred image as a target, divided by their sum. In our data, the average target effect index was 0.31 in V4 and 0.42 in IT ($p < 10^{-5}$; Fig 6c, f; compared to 0.24 in V4 and 0.26 in IT in their study). Notably, positive values of this index can be instantiated by both target identity and target match signals, thus explaining the large index found in V4, and an even larger value in our IT data.

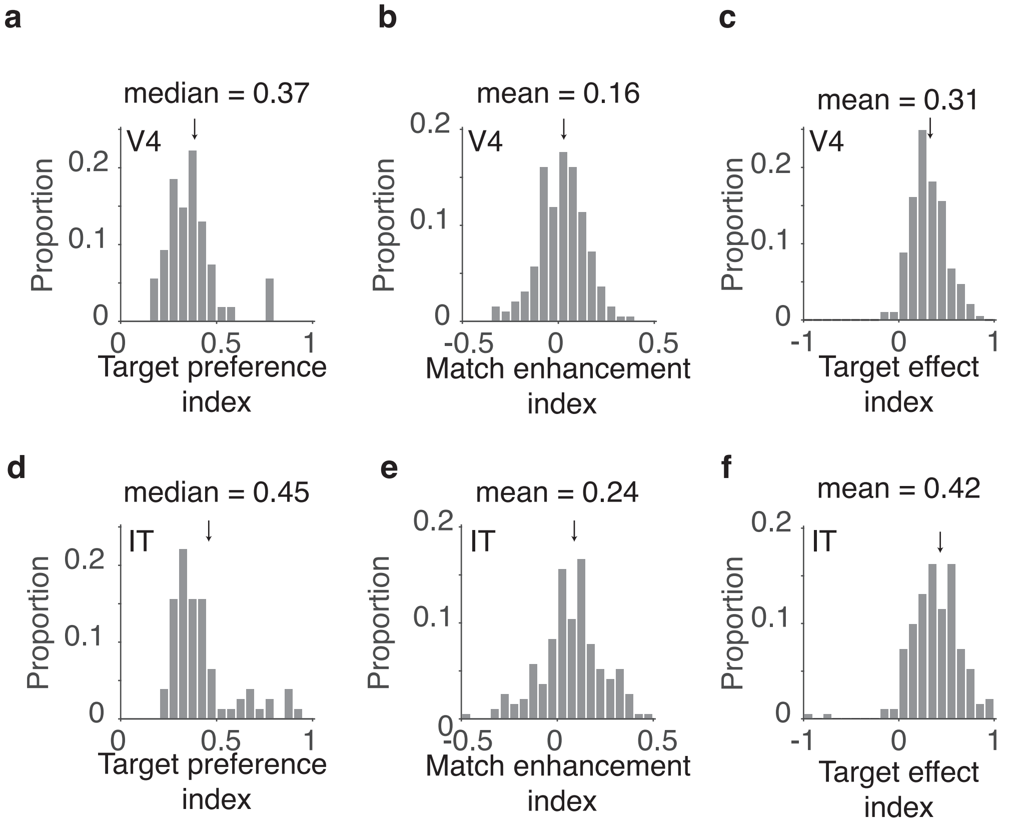

Figure 2-6. *Single neuron match enhancement and target signals in V4 and IT.* **a)** Target effect index as calculated in Maunsell et al. (1991). An index of the target effect was computed for each of the 193 units recorded in V4 that passed a 2-way ANOVA screen for target modulation, p<0.05. This index was (P-N)/(P+N), where P was the

average rate of firing during trials of any condition where the preferred target was the target; N was the average firing rate during trials of any condition where the least-preferred target was the target. Mean enhancement to the preferred target (0.37) is indicated by the arrow. **b)** Modulation index as calculated by Haenny et al. (1988). An index of match enhancement was computed for each of the 193 units recorded in V4. This index was (M-D)/(M+D) where M was the average rate of firing across trials where the preferred object was both in view and the target and trials where the non-preferred object was both in view and the target; D was the same for non match conditions of preferred and non-preferred objects. Average deviations from zero (absolute value of modulation) was 0.16, and the mean of the distribution was significantly shifted rightward from zero, $p=0.009$). The mean of the distribution is indicated by the arrow. **c)** Firing rate index as calculated by Chelazzi et al. (1998). The firing rate index was calculated as (FRp-FRn)/(FRp+FRn), where FRp represented the mean firing rate when the preferred image was in view and was the target object; FRn was the mean firing rate when the least preferred image was in view and was the target object. The mean of this distribution (0.31) is indicated by the arrow; this distribution was significantly shifted rightward from zero, $p<10^{-5}$. **d)** Same as in (a), for the population of IT units that passed a 2-way ANOVA screen for target modulation, $p<0.05$ ($n=193$). Mean enhancement to the preferred target (0.45) is indicated by the arrow. **e)** Same as in (b), for IT. Average deviations from zero (absolute value of modulation) was 0.24, and the mean of the distribution was significantly shifted rightward from zero, $p<10^{-5}$). **f)** Same as in (c), for IT. The mean of this distribution (0.42) is indicated by the arrow; this distribution was significantly shifted rightward from zero, $p<10^{-5}$.

In sum, the magnitudes of context-specific modulation we measured in V4 are largely consistent with previous reports. We also find that the same measures, applied to IT, are consistently larger. Our finding of larger magnitude context signals in IT is thus unlikely to follow from a discrepancy between our V4 data and that of previous studies. Rather, our data suggest that top-down, context-specific signals are combined with feed-forward visual information at multiple stages along the ventral visual pathway during the IDMS task.

DISCUSSION

Finding sought objects requires the brain to compare visual information about the objects in view with information about the currently sought target to compute a signal that reports when a target match has been found. In this study, we sought to differentiate between two scenarios of how this target match signal might be computed: one in which top-down, context-specific signals are introduced at multiple stages of the ventral visual pathway, and another in which V4 is the single locus for that combination. We found multiple lines of evidence supporting the hypothesis that context-specific signals are introduced at multiple stages of the ventral visual pathway. First, we found that the V4 population contains less total (and linearly separable) information for this task than the IT population does, suggesting that IT does not inherit all of its information from V4. Second, we found that V4 single units reflect information about target identity but not information that explicitly differentiates between target matches and distractors, while IT units reflect both of these types of information. Lastly, we found that while our measures of V4 single unit context modulation are largely consistent with previous reports, the same measures applied to IT reveal significantly more context modulation. Together, these results suggest that top-down context-specific signals are inserted at multiple stages along the ventral visual pathway.

Our study was motivated in large part by previous results that compared responses between V4 and IT during visual target search task and reported roughly matched amounts of task-relevant modulation (Chelazzi et al., 1998; Chelazzi et al., 2001). However, we found the opposite result. Differences between our results and theirs cannot be attributed to differences in the measures applied to the data (Fig 6), but other differences may account for them. Their experimental design included two images within each receptive field, one of which matched the cue stimulus, and the monkey was required to make a saccade to the target match stimulus location. Thus their target effect can be explained, at least in part, by modulations of spatial attention, as can be viewed within the context of the biased competition model. Furthermore, because all target images matched the cue images, stimulus repetition as described by an earlier report from the same lab (Miller & Desimone, 1994) could have played a role in the modulation of their neurons (under the assumption that this type of adaptation can transfer across different spatial locations). Lastly, the differences could arise from the fact that their study required matching the same images whereas the IDMS task required matching an object that could appear under different identity preserving transformations. Because object invariance is stronger in IT as compared to V4, (Rust & DiCarlo, 2010), it could be the case that invariant object search more strongly recruits IT cortex.

The results of our study support a scenario in which IT contains more task-relevant signal than V4 does during invariant object search. One of the central reasons this scenario is attractive is that when a task requires finding an object that can vary in its background context, the combination of object and target information would facilitated by an underlying visual representation in which information about object identity can be easily accessed, and that this type of representation gradually emerges across the ventral visual pathway. Stated differently, flexibly finding different objects requires differential responses to the same visual stimuli based on context. However, V4 receptive fields are small and retinotopically organized and consequently, that V4 lacks an explicit representation of object identity Therefore, it might be impossible for the brain to determine which subsets of neurons to target with top-down contextual signals. In contrast, it might be easier for the brain to integrate top-down cognitive information to late stages in the ventral visual pathway (e.g. IT) where visual representations are tolerant to identity-preserving transformations. Notably, one earlier study also explored the responses of IT neurons in the context of a DMS task in which, like ours, the objects could appear at different identity-preserving transformations (Leuschow et al., 1994), but this study did not compare signals to those in V4. Our study provides the first systematic comparison between these two areas within the context of a task that incorporates the real-world challenge of searching for objects can appear at different positions, sizes and background contexts.

We designed our experiment such that a cue wasn't presented immediately before the presentation of the sequence of distractors and the target match. This experimental design was motivated by Miller and Desimone (1994), who found some neurons that were enhanced and others that were suppressed by target matches compared to distractors. In this study, the authors suggested that the match suppressed responses might have arisen as the result of passive, stimulus repetition of the target match following the cue, while the match enhanced neurons alone carry behaviorally-relevant target match information. In our study, in both V4 and IT, we found that the majority of modulations found were firing rate enhancements. One previous study in V4 (Kosai et al., 2014) found equal amounts of enhancement and suppression. In IT, previous studies have consistently found both increases and decreases, in many cases finding on average suppression. In these previous studies, target match signals have been investigated via a classic version of the delayed-match-to-sample (DMS) paradigm where each trial begins with a visual cue indicating the identity of the target object, and this cue is often the same image as the target match. Our results reveal that when target matches do not follow the presentation of the same visual image at a short time before (as is the case for natural object search), match suppression is weaker than match enhancement (Fig 6), in line with the model that match enhanced neurons carry behaviorally-relevant information, while match suppressed neurons reflect adaptation to repeated stimuli. Two other studies in V4 (Bichot et al., 2005; Haenny et al., 1988) did not use classic DMS tasks, and they, like our study, found mostly enhancements.

In our study, we found that equal sized populations of units were matched for visual discriminability, consistent with previous results (Rust & DiCarlo, 2010) (DiCarlo et al., 2012). Notably, this need not be the case: it could have been the case that we found we needed to record from larger numbers of neuron in one area to make fair comparisons (e.g. convergence or divergence ratios different than 1). Furthermore, our results describe that both when you limit a classifier to the format of the information (e.g. linear classifier) and when you include the possibility of information being separable but nonlinearly formatted (e.g. maximum likelihood classifier), IT contained more information than V4. However, because there was still some small amount of task-relevant information in V4, one could imagine a readout rule that could give strong weights to a small subpopulation V4 neurons with the most information, and via a different convergence rule (e.g. 3x times more V4 neurons than IT neurons), match the amount of task-relevant information in IT. Such connections could be wired via a reinforcement learning algorithm (such as Law and Gold (2009)) during the natural experience of searching for target objects. Our results cannot rule this possibility out.

Similarly, it could have been the case that the total information in V4 (which was largely formatted in a nonlinearly separable way, as seen by larger total than linear information and corresponding to the single unit target identity modulations) was computed upon (untangled) by IT and thus matched the amount of linearly separable information in IT. This was not the case, i.e., the amount of linearly separable information in IT was significantly larger than the amount of total information in V4. It is however likely that what information V4 does have is inherited by IT, and IT simply receives more information from a top down source.

METHODS

Experiments were performed on two adult male rhesus macaque monkeys (*Macaca mulatta)* with implanted head posts and recording chambers. All procedures were performed in accordance with the guidelines of the University of Pennsylvania Institutional Animal Care and Use Committee.

The invariant delayed-match-to-sample (IDMS) task

All behavioral training and testing was performed using standard operant conditioning (juice reward), head stabilization, and high-accuracy, infrared video eye tracking. Stimuli were presented on an LCD monitor with an 85 Hz refresh rate using customized software (http://mworks-project.org).

As an overview, the monkeys' task required an eye movement response to a specific location when a target object appeared within a sequence of distractor images (Fig 2a). Objects were presented across variation in the objects' position, size and background context (Fig 2b). Monkeys viewed a fixed set of 20 images across switches in the identity of 4 target objects, each presented at 5 identity-preserving transformations (Fig 2c). We ran the task in short blocks (~3 min) with a fixed target before another target was pseudorandomly selected. Our design included two types of trials: cue trials and test trials (Fig 2a). Only test trials were analyzed for this report.

Trials were initiated by the monkey fixating on a red dot (0.15°) in the center of a gray screen, within a square window of ±1.5°, followed by a 250 ms delay before a stimulus appeared. Cue trials, which indicated the current target object, were presented at the beginning of each block and after three subsequent trials with incorrect responses. To minimize confusion, cue trials were designed to be distinct from test trials and began with the presentation of an image of each object that was distinct from the images used on test trials (a large version of the object presented at the center of gaze on a gray background; Fig 2a). Test trials, which are the focus of this report, always began with a distractor image, and neural responses to this image were discarded to minimize non-stationarities such as stimulus onset effects. Unless otherwise noted (see below), all images were presented at the center of gaze, in a circular aperture that blended into a gray background (Fig 2b). Distractors were drawn randomly from a pool of 15 possible images within each block without replacement until each distractor was presented once on a correct trial, and the images were then re-randomized. On most trials, a random number of 1-6 distractors were presented, followed by a target match (Fig 2a). On a small fraction of trials, 7 distractors were shown, and the monkey was rewarded for fixating through all distractors. Each stimulus was presented for 400 ms (or until the monkeys' eyes left the fixation window) and was immediately followed by the presentation of the next stimulus. Following the onset of a target match image, monkeys were rewarded for making a saccade to a response target within a window of 75 – 600 ms to receive a juice reward. In monkey 1 this target was positioned 10 degrees below fixation; in monkey 2 it was 10 degrees above fixation. If 400 ms following target onset had elapsed and the monkey had not moved its eyes, a distractor stimulus was immediately presented. If the monkey continued fixating beyond the required reaction time, the trial was considered a "miss". False alarms were differentiated from fixation breaks via a comparison of the monkeys' eye movements with the characteristic pattern of eye movements on correct trials: false alarms were characterized by the eyes leaving the fixation window via its top (monkey 1) or bottom (monkey 2) outside the allowable correct response period and traveling more than 0.5 degrees whereas fixation breaks were characterized by the eyes leaving the fixation window in any other way. Within each block, 4 repeated presentations of the 20 images were collected, and a new target object was then pseudorandomly selected. Following the presentation of all 4 objects as targets, the targets were re-randomized. At least 10 repeats of each condition were collected. Overall, monkeys performed this task with high accuracy. Disregarding fixation breaks (monkey 1: 8% of trials, monkey 2: 11% of trials), percent correct on the remaining trials was as follows: monkey 1: 87% correct, 3% false alarms, and 10% misses; monkey 2: 96% correct, 1% false alarms, and 3% misses.

V4 receptive fields in region of the visual field in which we presented stimuli are small, on average they have radii of .56 degrees at the fovea, extending to radii of 1.4 at an eccentricity of 2.5 degrees (which was the largest eccentricity of interest for our study, as our stimuli were 5 degrees in width; Desimone & Schein, 1987; Gattass et al., 1988). It was thus important to ensure that monkeys had fixational control such that the same region of an image fell on each V4 receptive field across repeated presentations. In one monkey, fixational control was good (on average 85 and 97% of presentations occurred within a radius of 0.56 and 1.4 degrees respectively). In a second monkey, adequate fixational control could not be achieved naively. We thus applied a procedure in which we shifted each image at stimulus 25% toward the center of gaze (e.g. if the eyes were displaced 0.5 degrees to the left, the image was repositioned such that the center of the image fell 0.125 degrees to the left and 0.375 degrees from fixation). Image position then remained fixed until the onset of the next stimulus. This deviation was measured relative to the mean position across the 10 trials per stimulus condition, and we found that in Monkey 2 this deviation was thus relatively small: on average, 95, and 99% of presentations occurred within windows with a radius of 0.56 and 1.4 degrees, respectively.

For both monkeys, a V4 recording chamber was implanted on the left hemisphere, and an IT recording chamber was implanted on the right hemisphere. While IT receptive fields span the vertical meridian, thus allowing us to access the visual representation of both sides with a single chamber, V4 receptive fields are confined to the contralateral hemifield. To simulate V4 coverage of the ipsilateral visual field, on roughly half of the V4 recording sessions, (n = 7/15 sessions in Monkey 1, n = 11/20 sessions in Monkey 2), we presented the images reflected across the vertical axis. We then treated all V4 neurons recorded during these sessions as if they were in the left hemisphere (and thus, whose receptive fields were in the right visual field.). In both monkeys, IT chamber implantation and recording preceded V4. Behavioral performance was similar across the sessions (V4 percent correct overall = 96.5%; IT percent correct overall = 91.4%).

Neural recording

The activity of neurons in each V4 and IT was recorded via a single recording chamber in each monkey, for a total of four recording chambers across our experiments. Chamber placement for the IT chambers was guided by anatomical magnetic resonance images in both monkeys, and in one monkey, Brainsight neuronavigation (https://www.rogue-research.com/); both V4 chambers were guided by Brainsight neuronavigation. . The region of IT recorded was located on the ventral surface of the brain, over an area that spanned 4 mm lateral to the anterior middle temporal sulcus and 15-19 mm anterior to the ear canals.

Both V4 chambers were centered 1 mm posterior to the ear canals and 29 mm lateral to the midline, positioned at a 30 degree angle. V4 recording sites were confirmed by a combination of receptive field location and position in the chamber, corresponding to results reported previously (Gattass et al., 1988). Specifically, we recorded from units within and around the inferior occipital sulcus, between the lunate sulcus and superior temporal sulcus. V4 units in lower visual field were confirmed as having receptive field centers that traversed from the vertical to horizontal meridian across posterior to anterior recording. Units with receptive fields at the fovea and near the upper visual field were found lateral to those in the lower visual field, and were confirmed by having receptive field centers that traversed from the horizontal meridian to the vertical meridian across median to lateral recordings at increasing depths. Aside from their receptive field locations, units in the upper visual field did not have any obvious, distinguishable properties from those in the lower visual field.

Neural activity was largely recorded with 24-channel U probes (Plexon, Inc) with linearly arranged recording sites spaced with 100 mm intervals, with a handful of units recorded with single electrodes (Alpha Omega, glass-coated tungsten). Continuous, wideband neural signals were amplified, digitized at 40 kHz and stored using the OmniPlex Data Acquisition System (Plexon). Spike sorting was done manually offline (Plexon Offline Sorter). At least one candidate unit was identified on each recording channel, and 2-3 units were occasionally identified on the same channel. Spike sorting was performed blind to any experimental conditions to avoid bias. A multi-channel recording session was included in the analysis if the animal performed the task until the completion of at least 10 correct trials per stimulus condition, there was no external noise source confounding the detection of spike waveforms, and the session included a threshold number of task-modulated units (>4 on 24 channels). The sample size (number of units recorded) was chosen to approximately match our previous work (Pagan & Rust, 2014a; Pagan et al., 2013).

For all the analyses presented in this chapter, we measured neural responses by counting spikes in a window that began, in V4, 40 ms after stimulus onset, and in IT, 80 ms after stimulus onset. We then counted spikes in a 170 ms window in both areas, such that the spike counting windows were of equal length across the two compared areas and always preceded the monkeys' reaction times On 1.9% of all correct target match presentations, the monkeys had reaction times faster than 250 ms, and those instances were excluded from analysis such that spikes were only counted during periods of fixation.

In IT, we recorded neural responses across 20 experimental sessions (Monkey 1: 10 sessions, and Monkey 2: 10 sessions). In V4, we recorded neural responses across 35 experimental sessions (Monkey 1: 15 sessions, and Monkey 2: 20 sessions). When combining the units recorded across sessions into a larger pseudopopulation, we began by screening for units that met three criteria. First, units had to be modulated by our task, as quantified by a one-way ANOVA applied to our neural responses (80 conditions * 10 repeats) with p < 0.01. Second, we applied a loose criterion on recording stability, as quantified by calculating the variance-to-mean ratio (Fano factor) for each unit (computed by fitting the relationship between the mean and variance of spike count across the 80 conditions), and eliminating units with a Fano factor > 2.5. Finally, we applied a loose criterion on unit recording isolation, quantified by calculating the signal-to-noise ratio (SNR) of the waveform (as the difference between the maximum and minimum points of the average waveform, divided by twice the standard deviation across the differences between each waveform and the mean waveform), and excluding (multi)units with an SNR < 2. In IT, this yielded a pseudopopulation of 193 units (of 563 possible units), including 98 units from monkey 1 and 95 units from monkey 2.

In V4, this yielded a pseudopopulation of 598 units (of 970 possible units), including 345 units from monkey 1 and 253 units from monkey 2. We found that for these population sizes of V4 units, V4 visual discriminability exceeded that of IT. We thus randomly subselected units from each monkey to match visual discriminability. We found that for matched numbers of units, V4 and IT had matched visual discriminability both across the transformations used (Figure 4a) and for each transformation individually (Figure 4b). Our results were qualitatively unchanged for different subselections of V4 units given matched numbers of units between V4 and IT. Therefore, for the analyses shown, our final population size in V4 was thus 98 units from monkey 1 and 95 units from monkey 2, yielding a total pseudopopulation in V4 of 193 units.

Because we were unable to obtain receptive field coverage of the entire visual field, we observed that for 2 of the transformations (namely, “Left” and “Right”, Figure 2b), we did not get equal visual discriminability in V4 both compared to IT and compared to the visual discriminability for the other 3 transformations in V4. Therefore we restricted our analysis to the subset of transformations with matched discriminability: “Big”, “Up” and “Small”.

On half of the V4 sessions (Monkey 1: 7 out of 15 sessions, monkey 2: 11 out of

21 sessions), we presented identical images that were reflected around their vertical axis. For these sessions, we make the assumption that V4 response properties are equivalent across hemispheres, and for all subsequent analyses, we treat these neurons as if their receptive field centers are in the exact location on the other visual field.

V4 receptive field mapping

To measure the location and extent of V4 receptive fields, bars were presented, each for 500 ms, one per trial, centered on a 5 x 5 invisible grid. Bar orientation, length, and width as well as the grid center and extent were adjusted for each recording session based on preliminary hand mapping. On each trial, the monkey was required to maintain fixation on a small response dot (0.125°) to receive a reward. The responses to at least five repeats were collected at each position for each recording session. Only those units that produced clear visually evoked responses at a minimum of one position were considered for receptive field position analysis. The center of the receptive field was estimated by the maximum of the response across the 5x5 grid of oriented bar stimuli and confirmed by visual inspection.

Population target match performance

To determine both the performance of the neural populations at classifying target matches versus distractors we applied two types of decoders: a Fisher Linear Discriminant (a linear decoder) and Maximum Likelihood decoder (a nonlinear decoder) using approaches that are described previously in detail (Pagan et al., 2013) and are summarized here.

All decoders were cross-validated with the same resampling procedure. On each iteration of the resampling, we randomly shuffled the trials for each condition and for each unit, and (for numbers of units less than the full population size) randomly selected units. On each iteration, 9 trials from each condition were used for training the decoder and 1 trial from each condition was used for cross-validated measurement of performance. In separate data (see Chapter 3 Methods), we determined a value for regularization of the classifiers, and this optimal value was used for these studies.

Classifier analyses were done per transformation, for three transformations ("Big", "Up" and "Small"). For each transformation, there were thus 16 conditions (4 objects the monkey could be looking at under 4 different target contexts).

To ensure that decoder performance was not biased by unequal numbers of target matches and distractors, on each iteration of the resampling we included 4 target match conditions and 4 (of 12 possible) distractor conditions. Each set of 4 distractors was selected to span all possible combinations of mismatched object and target identities (e.g. objects 1, 2, 3, 4 paired with targets 4, 3, 2, 1), of which there are 9 possible sets. To compute proportion correct a mean performance value was computed on each resampling iteration by averaging binary performance outcomes across the 9 possible sets of target matches and distractors, each which contained 16 test trials, and across the three transformations used. Mean and standard error of performance was computed as the mean and standard deviation of performance across 2000 resampling iterations.

Fisher Linear Discriminant:

The general form of a linear decoding axis is:

(1) $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$,

where **w** is an N-dimensional vector (where N is the number of units) containing the linear weights applied to each unit, and b is a scalar value. We fit these parameters using a Fisher Linear Discriminant (FLD), where the vector of linear weights was calculated as:

(2) $\boldsymbol{w} = \Sigma^{-\boldsymbol{1}}(\mu_1 - \mu_2)$

and b was calculated as:

(3) $b = \boldsymbol{w} \cdot \frac{1}{2}(\mu_1 + \mu_2) = \frac{1}{2}\mu_1{}^T\Sigma^{-\boldsymbol{1}}\mu_1 - \frac{1}{2}\mu_2{}^T\Sigma^{-\boldsymbol{1}}\mu_2$

Here μ_1 and μ_2 are the means of the two classes (target matches and distractors, respectively) and the mean covariance matrix is calculated as:

(4) $\Sigma = \frac{\Sigma_1 + \Sigma_2}{2}$

where Σ_1 and Σ_2 are the regularized covariance matrices of the two classes. These covariance matrices were computed using a regularized estimate equal to a linear combination of the sample covariance and the identity matrix I (Pagan & Rust, 2014a):

(5) $\Sigma_i = \gamma\, \Sigma_i + (1 - \gamma) \cdot I$

We determined γ by exploring a range of values from 0.01 to 0.99, and we selected the value that maximized average performance across all iterations, measured with the cross-validation "regularization" trials set aside for this purpose (see above). We then computed performance for that value of γ with separately measured "test" trials, to ensure a fully cross-validated measure. Because this calculation of the FLD parameters incorporates the off-diagonal terms of the covariance matrix, FLD weights are optimized for both the information conveyed by individual units as well as their pairwise interactions.

We computed two measures of performance: proportion correct (Fig 3b-c), and population d' (Fig 6a). Each calculation began by computing the dot product of the test data and the linear weights **w**, adjusted by *b* (Eq. 1). Proportion correct was computed as the fraction of test trials that were correctly assigned as target matches and distractors, according to their true labels. Population d' was computed for the distributions of these values across the 4 different objects presented as target matches versus as distractors:

$$(6)\ d' = \frac{|\mu_{Match} - \mu_{Distractor}|}{\sigma_{pooled}},$$

where μ_{Match} and $\mu_{Distractor}$ correspond to the mean across the set of matches and distractors, $\sigma_{pooled} = \sqrt{\frac{\sigma_{Match}^2 + \sigma_{Distractor}^2}{2}}$, and σ_{Match} and $\sigma_{Distractor}$ correspond to the standard deviation across the set of matches and distractors, respectively.

Maximum likelihood decoder:

As a measure of total target match information (combined linear and nonlinear), we implemented a maximum likelihood decoder (Fig 3a-b). We began by using the set of training trials to compute the average response r_{uc} of each unit u to each of the 20 conditions c. We then computed the likelihood that a test response k was generated from a particular condition as a Poisson-distributed variable:

$$(7)\ lik_{u,c}(k) = \frac{(r_{uc})^{k} \cdot e^{-r_{uc}}}{k!}$$

The likelihood that a population response vector was generated in response to each condition was then computed as the product of the likelihoods of the individual units. Next, we computed the likelihood that each test vector arose from the category target match as compared to the category distractor as the product of the likelihoods across the conditions within each category. We assigned the population response to the category with the maximum likelihood, and we computed performance as the fraction of trials in which the classification was correct based on the true labels of the test data.

<u>Population performance (visual discriminability)</u>

To determine how well a population of neurons could classify object identity, we applied a fisher linear discriminant, as described above, in the following way.

We used a standard “one-versus-rest” training and testing classification scheme (Rust and Dicarlo 2010; Hung et al., 2005; Li et al., 2009). Specifically, one linear classifier was determined for each image ; To determine the population “decision” about which image was presented, a response vector **x**, corresponding to the population response of one image, was then applied to each of the classifiers, and the classifier with the largest output [the classifier with the largest, positive *f*(**x**)] was taken as the choice of the population.

To compute performance as population d’ (Fig 4c-d), d’ was computed on each resampling iteration for the 4 target match conditions and 4distractor conditions, separately for each set of 9 match/distractor combinations, and then averaged across the 9 sets. When computing d’ as a function of the number of units (Fig4c), this value was also averaged across the three transformations used. Mean and standard error of population d’ was computed as the mean and standard deviation of d’ across 2000 resampling iterations. Standard error thus reflected the variability due to the specific trials assigned to training and testing and, for populations smaller than the full size, the specific units chosen.

Quantifying single-unit modulations

To compare the degree to which the firing rates of individual units were modulated by target search, we compared firing rate modulations, computed as three different indices as in previous studies (Chelazzi et al., 1998; Haenny et al., 1988; Maunsell et al., 1991). Each of these indices was computed for each unit in V4 and IT. First, we calculated the target effect index as calculated in Maunsell et al. (1991). To compare to their results, this index was computed only for neurons which were significantly modulated by the identity of the target (via a 2-way ANOVA, $p<.05$). For units that passed this screen, an index was computed as (P-N)/(P+N), where P was the average rate of firing during trials of any condition where the preferred target was the target; N was the average firing rate during trials of any condition where the least preferred target was the target. Next, we calculated the modulation index as calculated by Haenny et al. (1988). This index was computed as (M-D)/(M+D) where M was the average rate of firing across trials where the preferred object was both in view and the target and trials where the non-preferred object was both in view and the target; D was the same for non match conditions of preferred and non-preferred objects. To compute average deviations from zero, we took the average across the absolute value of each unit's modulation index. Lastly, we calculated a target effect index as calculated by Chelazzi et al. (1998). This index was computed as (FRp-FRn)/(FRp+FRn), where FRp represented the mean firing rate when the preferred image was in view and was the target object; FRn was the mean firing rate when the least preferred image was in view and was the target object.

To quantify the degree to which individual units were modulated by different types of task parameters, we applied a bias-corrected, ANOVA-like procedure described in detail by (Pagan & Rust, 2014b) and summarized here. As an overview, this procedure considers the total variance in the spike count responses for each unit across conditions (n=16 for each transformation) and trials for each condition (m=10), and parses this total variance into the variance that can be attributed to each type of experimental parameter and variance attributed to trial variability. Similar to an ANOVA, the procedure is designed to parse response variance, including the variance that can be attributed to changes in the identity of the visual image, the identity of the target object and whether each condition was a target match or a distractor. These variances are converted into measures of spike count modulation (i.e. standard deviation around each unit's grand mean spike count) via a procedure that includes bias correction for over-estimates in modulation due to noise.

The procedure begins by developing an orthonormal basis of 80 vectors designed to capture all types of modulation with intuitive groupings. The number of each type is imposed by the experimental design. This basis $\boldsymbol{b}$ included vectors $\boldsymbol{b}_i$ that reflected 1) the grand mean spike count across all conditions ($\boldsymbol{b}_1$, 1 dimension), 2) whether the object in view was a target or a distractor ($\boldsymbol{b}_2$, 1 dimension), 3) visual image identity ($\boldsymbol{b}_3 - \boldsymbol{b}_5$, 3 dimensions), 4) target object identity ($\boldsymbol{b}_6 - \boldsymbol{b}_8$, 3 dimensions), and 5) "residual", nonlinear interactions between target and object identity not captured by target match modulation ($\boldsymbol{b}_9 - \boldsymbol{b}_{16}$, 8 dimensions). A Gram-Schmidt process was used to convert an initially designed set of vectors into an orthonormal basis.

Because this basis spans the space of all possible responses for our task, each trial-averaged vector of spike count responses to the 16 experimental conditions for each transformation used; $\boldsymbol{R}$ can be re-expressed as a weighted sum of these basis vectors. To quantify the amounts of each type of modulation reflected by each unit, we began by computing the squared projection of each basis vector $\boldsymbol{b}_i$ and $\boldsymbol{R}$. An analytical bias correction, described and verified in (Pagan & Rust, 2014b), was then subtracted from this value:

$$(8)\ w_i^2 = (\boldsymbol{R} \cdot \boldsymbol{b}_i^T)^2 - \frac{\sigma_t^2 \ \cdot (\boldsymbol{b}_i^T)^2}{m}$$

where σ_t^2 indicates the trial variance, averaged across conditions (n=16), and where m indicates the number of trials (m=10). When more than one dimension existed for a type of modulation, we summed values of the same type. Next, we applied a normalization factor (1/(n-1) where n=16) to convert these summed values into variances. Finally, we computed the square root of these quantities to convert them into modulation measures that reflected the number of spike count standard deviations around each unit's grand mean spike count. Target match modulation was thus computed as:

(9) $\sigma_{TM} = \sqrt{\frac{1}{n-1} \cdot w_2^2}$

and nuisance modulation was computed as:

(10) $\sigma_{Nui} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=3}^{80} w_i^2}$

Similarly, to compute the different subtypes of nuisance modulation, we replaced the weights w_i^2 in Eq. 10 with the weights that corresponded to the orthonormal basis vectors corresponding to each subtype, including visual modulation ($i = 3\ to\ 5$), target modulation ($i = 6\ to\ 8$), and 3) residual modulation ($i = 9\ to\ 16$), as described above.

We computed the trial variability for each unit ($\sigma_{Trial},$) in an comparable manner as the square root of the average (across conditions) variance across trials:

(11) $\sigma_{Trial} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{m-1} \cdot \sum_{t=1}^{m} (s_{it} - s_i)^2}$

where the spike count response for a particular trial t of condition i was s_{it}, and the mean spike count response across all trials of condition i was s_i.

When estimating modulation for individual units, (Fig 4a), the bias-corrected squared values were rectified for each unit before taking the square root. When estimating modulation population means (Fig 4b, 5b), the bias-corrected squared values were averaged across units before taking the square root. Because these measures were not normally distributed, standard error about the mean was computed via a bootstrap procedure. On each iteration of the bootstrap (across 1000 iterations), we randomly sampled values from the modulation values for each unit in the population, with replacement. Standard error was computed as the standard deviation across the means of these newly created populations.

Statistical tests

When comparing population decoding measures (Fig 3b), we reported *P* values as an evaluation of the probability that differences were due to chance. We calculated these *P* values as the fraction of resampling iterations on which the difference was flipped in sign relative to the actual difference between the means of the full data set (for example, if the mean of decoding measure 1 was larger than the mean of decoding measure 2, the fraction of iterations in which the mean of measure 2 was larger than the mean of measure 1).

When evaluating whether the single neuron indices (Fig 5) were significantly different from zero, we reported p values as computed by a Wilcoxon sign rank test.

CHAPTER 3

Large nuisance modulation has little impact on IT target match performance

ABSTRACT

Many everyday tasks require us to extract a specific type of information from our environment while ignoring other things. When the neurons in our brains that carry task-relevant signals are also modulated by task-irrelevant "nuisance" information, nuisance modulation is expected to act as performance-limiting noise. To investigate the impact of nuisance modulation on neural task performance, we recorded responses in inferotemporal cortex (IT) as monkeys performed a task in which they were rewarded for indicating when a target object appeared amid considerable nuisance variation. Within IT, we found a robust, behaviorally-relevant target match signal that was mixed with large nuisance modulations in individual neurons. Unexpectedly, we also found that these nuisance modulations had little impact on performance, either within individual IT neurons or across the IT population. We demonstrate how these results follow from fast processing in IT, which placed IT in a low spike count regime where the impact of nuisance variability was blunted by Poisson-like trial variability. These results demonstrate that some basic intuitions about neural coding are misguided in the context of a fast-processing, low spike count regime.

INTRODUCTION

Task performance is determined not only by the amount of task-relevant signal present in our brains, but also by the presence of noise, which can arise from multiple sources. Internal noise, or “trial variability” manifests as trial-by-trial variations in neural responses under seemingly identical conditions (Fig 1a). External factors can also translate into noise, particularly when a task requires extracting a particular type of information from our environment amid changes in other task-irrelevant, nuisance parameters (Fig 1b; Haefner & Bethge, 2010; Kim et al., 2016). Stated differently, for any given task, neurons in a brain area may be modulated by multiple experimental variables, but when viewed from the perspective of task performance, one type of modulation reflects the task-relevant signal, whereas other types of modulation act as noise.

Despite notions that mixing different types of signals within the responses of individual neurons should be detrimental for task performance (Fig 1b), growing evidence suggests that the brain does often mix them, both at the locus at which task-relevant solutions are computed as well as downstream (Freedman & Assad, 2009; Kobak et al., 2016; Mante et al., 2013; Meister et al., 2013; Raposo et al., 2014; Rigotti et al., 2013; Rishel et al., 2013; Zoccolan et al., 2007). One example is visual target search, which requires the brain to compare incoming visual information with a remembered representation of a target to create a signal that reports when a target match is in view. When considered across changes in target identity (e.g. looking for your car keys and then your wallet), target search can be envisioned as differentiating the same images presented as target matches versus as distractors (e.g. when looking for your car keys, your wallet is a distractor; when looking for your wallet, your car keys are a distractor and your wallet is a target match). Consequently, other types of modulation, such as visual modulation (e.g. signals that differentiate wallets and car keys regardless of what you are searching for), act as noise. A number of lines of evidence suggest that target match information emerges in the ventral visual pathway as early as V4 (Kosai et al., 2014; Maunsell et al., 1991) and inferotemporal cortex (IT, Chelazzi et al., 1993; E.N. Eskandar et al., 1992; Leuschow et al., 1994; Miller & Desimone, 1994; Pagan et al., 2013), where nuisance modulation, including visual modulation, is expected to be large. This suggests that nuisance modulation may place strong limitations on neural target match performance in these ventral visual pathway brain areas.

Understanding how nuisance modulation affects neural task performance requires considering its impact in individual neurons as well as across the population. Investigations, focused in part on view-invariant object recognition, have demonstrated the means by which individual neurons can multiplex different types of signals such that each type of signal can be extracted from the population with a simple linear decoder (DiCarlo & Cox, 2007; Hong, Yamins, Majaj, & DiCarlo, 2016; Li, Cox, Zoccolan, & Dicarlo, 2009; Ohki, Chung, Ch'ng, Kara, & Reid, 2005). But little attention has been directed toward understanding how signal mixing impacts population performance within the context of these linearly separable representations. Some insight into these issues can be gained from work focused on how correlated interactions between neurons impacts population performance within a linear decoding scheme (reviewed by Averbeck & Lee, 2006; Cohen & Kohn, 2011; Kohn, Coen-Cagli, Kanitscheider, & Pouget, 2016). However, this work has focused nearly exclusively on correlated trial (as opposed to nuisance) variability (but see Kim et al., 2016). Understanding how nuisance modulation impacts neural task performance will thus require extending these population-based approaches to incorporate considerations about nuisance modulation.

To investigate the impact of nuisance modulation on IT target match performance, we recorded neural signals in IT as monkeys performed a modified delayed-match-to-sample task in which they were rewarded for indicating when a target object appeared across changes in the objects' position, size and background context.

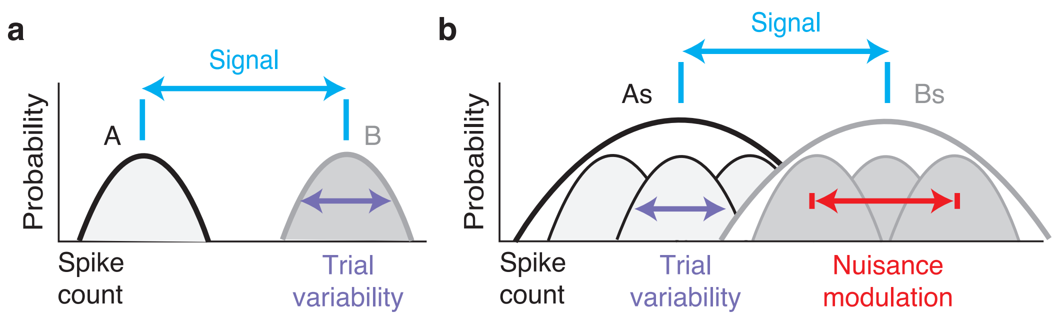

Figure 3-1. *Nuisance modulation limits task performance.* **a)** Schematic of single unit task performance (d') for a classic, two-way discrimination task in which a subject is asked to label different conditions as "A" or "B" across repeated trials. Shown are hypothetical distributions of spike count responses for the two conditions. d', is measured as the separation of the two spike count distributions in units of the number of standard deviations separating their means. d' is proportional to the amount of signal, which determines the separation between the means of the distributions (cyan), and d' is inversely proportional to spread within each distribution, which arises as a result of variability across repeated trials within each condition ("trial variability"; purple). **b)** Schematic of single unit task performance (d') for the same discrimination task, but extended to require grouping multiple conditions into each of two sets, "As" and "Bs" (e.g. an object identification task where two objects are presented in multiple background contexts). In this case, "nuisance" modulations (e.g. firing modulations by the background context), increase the spread of the responses within each condition and thus lower d'.

RESULTS

The invariant delayed-match-to-sample task (IDMS)

To investigate the degree to which nuisance modulation impacts neural task performance, we trained two monkeys to perform an “invariant delayed-match-to-sample” (IDMS) task that required them to report when target objects appeared across variation in the objects’ positions, sizes and background contexts. In this task, the target object was held fixed for short blocks of trials (~3 minutes on average) and each block began with a cue trial indicating the target for that block (Fig 2a, “Cue trial”). Subsequent test trials always began with the presentation of a distractor and on most trials this was followed by 0-5 additional distractors (for a total of 1-6 distractor images) and then an image containing the target match (Fig 2a, “Test trial”). The monkeys’ task required them to fixate during the presentation of distractors and make a saccade to a response dot on the screen following target match onset to receive a reward. To minimize the predictability of the match appearing as a trial progressed, on a small subset of the trials the match did not appear and the monkey was rewarded for maintaining fixation. Our experimental design differs from other classic DMS tasks (Chelazzi et al., 1993; E.N. Eskandar et al., 1992; Leuschow et al., 1994; Miller & Desimone, 1994; Pagan et al., 2013) in that it does not incorporate a cue at the beginning of each test trial, to better mimic real-world object search conditions in which target matches are not repeats of the same image presented shortly before.

Our experiment included a fixed set of 20 images, broken down into 4 objects presented at each of 5 transformations (Fig 2b). Our goal in selecting these specific images was to make the task of classifying object identity challenging for the IT population and these specific transformations were built on findings from our previous work (Rust & DiCarlo, 2010). In any given block (e.g. a squirrel target block), a subset of 5 of the images would be considered target matches and the remaining 15 would be distractors (Fig 2b). Our full experimental design amounted to 20 images (4 objects presented at 5 identity-preserving transformations), all viewed in the context of each of the 4 objects as a target, resulting in 80 experimental conditions (Fig 2c). In this design, “target matches” fall along the diagonals of each looking at / looking for matrix slice (where “slice” refers to a fixed transformation; Fig 2c, gray). For each condition, we collected at least 20 repeats on correct trials. Monkeys generally performed well on this task (Fig 2d). Their mean reaction times (computed as the time their eyes left the fixation window relative to the target match stimulus onset) were 364 ms and 332 ms (Fig 2e).

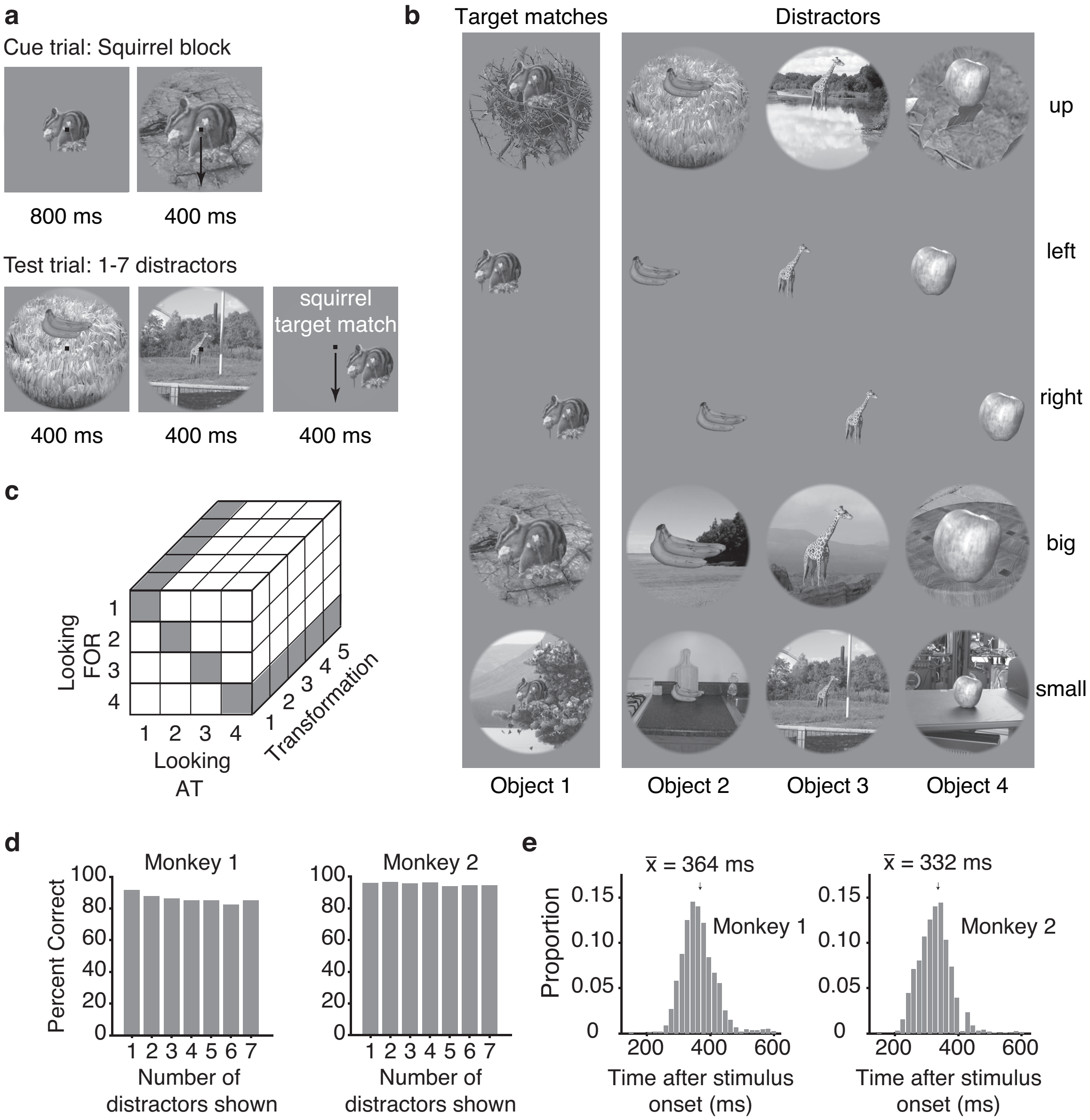

Figure 3-2. *The invariant delayed-match-to-sample task.* **a)** Monkeys performed an invariant delayed-match-to-sample task. Each block (~3 minutes in duration) began with a cue trial indicating the target object for that block. On subsequent trials, monkeys initiated a trial by fixating on a small dot. After a 250 ms delay, a random number (1-7) of distractors were presented, and on most trials, this was followed by the target match. Monkeys were required to maintain fixation throughout the distractors and make a

saccade to a response dot within a window 75 - 600 ms following the onset of the target match to receive a reward. In cases where the target match was presented for 400 ms and the monkey had still not broken fixation, a distractor stimulus was immediately presented. **b)** The experiment included 4 objects presented at each of 5 identity-preserving transformations (“up”, “left”, “right”, “big”, “small”), for 20 images in total. In any given block, 5 of the images were presented as target matches and 15 were distractors. **c)** The complete experimental design included looking “at” each of 4 objects, each presented at 5 identity-preserving transformations (for 20 images in total), viewed in the context of looking “for” each object as a target. In this design, target matches (highlighted in gray) fall along the diagonal of each “looking at” / “looking for” transformation slice. **d)** Percent correct for each monkey, calculated based on both misses and false alarms (but disregarding fixation breaks). Percent correct is plotted as a function of the number of distractors shown. **e)** Histograms of reaction times during correct trials (ms after stimulus onset) during the IDMS task for each monkey, with means indicated by arrows and labeled.

As two monkeys performed this task, we recorded neural activity from small populations in IT using 24-channel probes. We performed two types of analyses on these data. The first type of analysis was performed on the data recorded simultaneously across units within a single recording session (n=20 sessions). The second type of analysis was performed on data that was concatenated across different sessions to create a pseudopopulation after screening for units based on their stability, isolation, and task modulation (see Methods; n=204 units). For all but one of our analyses (Fig 4d), we counted spikes in a window that started 80 ms following stimulus onset (to allow stimulus-evoked responses time to reach IT) and ended at 250 ms, which was always before the monkeys' reaction times on these trials. For all but one of our analyses (Fig 3c), the data are extracted from trials with correct responses.

IT reflects behaviorally-relevant target match information

The primary focus of this report is the impact of mixing signal and nuisance modulation on neural task performance. Before exploring the consequences of nuisance modulation, we begin by demonstrating that behaviorally-relevant target match information is in fact reflected in IT during the IDMS task.

The IDMS task required monkeys to determine whether each condition (an image viewed in the context of a particular target) was a target match or a distractor. This task ultimately maps all the target match conditions onto one behavioral response (a saccade) and all the distractor conditions onto another (maintain fixation), and as such, this task can be envisioned as a two-way classification that must be performed invariant to changes in other nuisance parameters, including changes in target and image identity

(Fig 3a). To quantify the amount and format of target match information within IT, we began by quantifying cross-validated performance of this two-way classification with a linear population decoder (a Fisher Linear Discriminant, FLD). Linear decoder performance began near chance and grew as a function of population size, consistent with a robust IT target match representation (Fig 3b, black). To determine the degree to which a component of IT target match information might be present in a nonlinear format that could not be accessed by a linear decoder, we measured the performance of a maximum likelihood decoder designed to extract target match information regardless of its format (combined linear and nonlinear, Pagan et al., 2013, see Methods). Performance of this nonlinear decoder (Fig 3b, gray) was slightly higher and significantly better than linear decoder performance (p = 0.022), suggesting that while the majority of IT target match information is reflected in a linearly separable format, a smaller nonlinear component exists as well.

Upon establishing the format of target match information on correct trials, we were interested in determining the degree to which behavioral confusions were reflected in the IT neural data. To measure this, we focused on the data recorded simultaneously across multiple units within each session, where all units observed the same errors. With this data, we trained the linear decoder to perform the same target match versus distractor classification described for Fig 3b using data from correct trials, and we measured cross-validated performance on pairs of condition-matched trials: one for which the monkey answered correctly, and the other for which the monkey made an error. On correct trials, target match decoder performance grew with population size and reached above chance levels in populations of 24 units (Fig 3c, black). On error trials, decoder performance fell below chance, and these results replicated across each monkey individually (Fig 3c, white). These results establish that IT reflects behaviorally-relevant target match information insofar as this measure co-varies with the monkeys' behavior.

We were also interested in understanding how target match modulation was reflected in individual units. Target match modulation, by definition, requires a differential response to the same images presented as matches versus as distractors - to what degree is this modulation reflected by firing rate increases versus decreases? To measure this, we computed a target match modulation index for each unit as the average difference between the responses to the same images presented as target matches versus as distractors, divided by the sum of those two quantities. This index (Fig 3d) was shifted toward target match preferring units, with a mean value of 0.067 (monkey 1 = 0.071; monkey 2 = 0.063). These results are consistent with a target match signal that is largely reflected in most IT units via increased responses to target matches as compared to distractors.

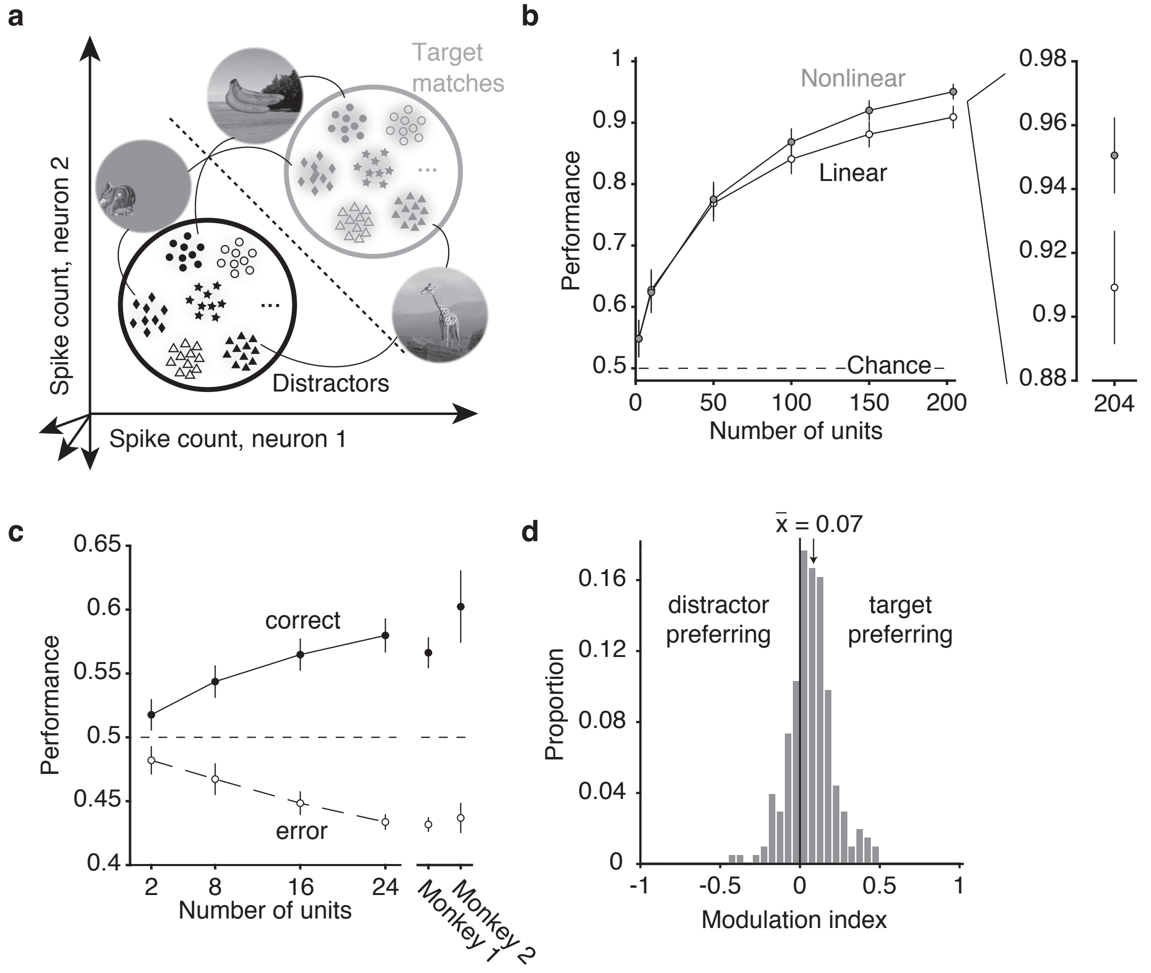

Figure 3-3. *IT reflects behaviorally-relevant target match information during the IDMS task.* **a)** The target search task can be envisioned as a two-way classification of the same images presented as target matches versus as distractors. Shown are cartoon depictions where each point depicts a hypothetical population response for a population of two neurons on a single trial, and clusters of points depict the dispersion of responses across repeated trials for the same condition. Included are responses to the same images presented as target matches and as distractors - here only 6 images are depicted but 20 images were used in the actual analysis. The dotted line depicts a hypothetical linear decision boundary. **b)** Linear (FLD) and nonlinear (maximum

likelihood) decoder performance as a function of population size for a pseudopopulation of 204 units. Error bars (SEM) reflect the variability that can be attributed to the random selection of units (for populations smaller than the full dataset) and the random assignment of training and testing trials in cross-validation. **c)** Linear decoder performance, applied to the simultaneously recorded data for each session, after training on correct trials and cross-validating on pairs of correct and error trials matched for condition. n=20 sessions. Error bars (SEM) reflect the variability that can be attributed to the random selection of units (for populations smaller than the full dataset) and the random assignment of training and testing trials in cross-validation. **d)** A match modulation index, computed for each unit by calculating the mean spike count response to target matches and to distractors, and computing the ratio of the difference and the sum of these two values. Arrow indicates the distribution mean.

During the IDMS task, nuisance modulation is prominent

As described above, we were interested in understanding whether and how nuisance modulation impacted IT target match performance. As a first step toward addressing this question, we wanted to quantify the relative amounts of target match and nuisance modulation present within individual units. To quantify the different types of modulation reflected in IT, we applied a bias-corrected procedure that quantified different types of modulation in terms of the number of standard deviations around each unit's grand mean spike count (Pagan & Rust, 2014b). Modulation types were grouped into intuitive sets, including modulation that could be attributed to whether each condition was a target match or a distractor (the "target match" signal), modulation due to changes in the identity of the visual stimulus ("visual"), modulation due to changes in the identity of the target ("target id."), and "residual" modulations attributed to nonlinear interactions between the visual stimulus and target that were not captured by target match modulation (e.g. specific distractor conditions). We also combined all the different types of "nuisance" modulation into one measure for each unit.

Our measure of modulation is similar to a multi-way ANOVA, with important extensions. Specifically, a two-way ANOVA applied to a unit's responses (configured into a matrix of 4 targets * 20 images * 20 trials for each condition) would parse the total response variance into two linear terms, a nonlinear interaction term, and an error term. We make 3 extensions to the ANOVA analysis. First, an ANOVA returns measures of variance (in units of spike counts squared) whereas we compute measures of standard deviation (in units of spike count) such that our measures of modulation are intuitive (e.g., doubling firing rates causes signals to double as opposed to quadruple). Second, while the linear terms of the ANOVA map onto our "visual" and "target id." modulations (after squaring), we split the ANOVA nonlinear interaction term into two terms, including target match modulation (i.e. Fig 2c gray versus white) and all other nonlinear “residual” modulation. This parsing is essential, as target match modulation corresponds to the signal for the IDMS task whereas residual modulation acts as noise (described in more detail below, Fig 4b). Finally, raw ANOVA values are biased by trial-by-trial variability (which the ANOVA addresses by computing the probability that each term is higher than chance given this noise) whereas our measures of modulation are bias-corrected to provide an unbiased estimate of modulation magnitude (see Methods).

Across the 204 IT units, we found that total nuisance modulation was larger than target match modulation in most cases (Fig 4a), and that average nuisance modulation was 2.8x the average target match signal (Fig 4b). A more detailed parsing of the total nuisance modulation into different subtypes revealed that the largest type of nuisance modulation could be attributed to “visual” modulations (on average 2.6x the target match signal; Fig 4b). Other types of modulation were also prominent, including “target id.” modulations (on average 0.8x the target match signal; Fig 4b), and “residual” modulation (on average 0.6x the target match modulation; Fig 4b). These results reveal that within IT, nuisance modulations are prominent and they are mixed with the target match signal in individual units.

In sum, the results presented thus far verify the existence of a robust, behaviorally-relevant target match signal in IT, and they confirm our predictions that IT target match signals are mixed with large nuisance modulations within individual IT units. Together, these results support assertions that the activity of IT units during visual target search should be an effective test of the impact that nuisance modulation has on neural task performance.

Unexpectedly, the impact of nuisance modulation on single-unit performance is modest

Ultimately, understanding the impact of nuisance modulation on linearly decoded task performance requires considering both the responses of individual units as well as their population interactions. Here we begin by quantifying the impact of nuisance modulation on individual units, the results of which were quite unexpected.

As a measure of linearly decoded target match performance for individual units, we focus on single-unit d' (Fig 1b). Single-unit d' is determined by the separation between the spike count responses of a unit to the set of all images presented as target matches versus the same images presented as distractors, and is quantified as the ratio between the distance between the means over the average standard deviation of the two distributions (Fig 1b). Single-unit d' is thus proportional to the amount of "target match signal", equivalent to the distance between the means of the responses to target matches and to distractors (Fig 1b, cyan). Conversely, single-unit d' is inversely proportional to the spread within each distribution, where spread is determined by two factors. The first contributor to this spread is the variability in the spike count responses across repeated trials of the same condition, or "trial variability" (Fig 1b, purple). The second contributor to this spread is the dispersion between different conditions within each set, equivalent to all types of modulation that are not the target match signal ("nuisance" modulation; Fig 1b, red). This is why signal mixing is predicted to be detrimental to single-unit task performance – because any nuisance modulation that exists within a unit is predicted to increase the overlap between target matches and distractors.

In a previous report, we formalized these intuitions into a mathematical relationship between the single-unit modulation magnitudes as measured in Fig 4a-b and single-unit d' (Pagan & Rust, 2014b). This derivation can be applied here with minor extensions. To summarize that approach, d' is a measure of the ratio between signal and noise, where signal is proportional to the amount of target match modulation (Fig 4b, cyan) and noise is parsed into one component proportional to total nuisance modulation (Fig 4b, red) and another component proportional to trial variability (Fig 4b, purple):

$$|d'| = \sqrt{\frac{k_1 * target\ match\ modulation^2}{k_2 * nuisance\ modulation^2 + trial\ variability^2}}$$

where k_1 and k_2 are constants (see Methods). With this formulation, the impact of nuisance modulation on d' can be determined by considering the increase in d' when nuisance modulation is incorporated into the calculation (i.e. for the intact data) compared to when it is not (i.e. a hypothetical scenario in which nuisance modulation does not exist, analogous to the increase in d' in Fig 1a relative to 1b). Fig 4c shows the result of this analysis, which reveals that removing nuisance only results in a modest increase in d' across units, with an average increase of 10.1%. Focusing on the most informative units (i.e. those with the highest d'), did not change the qualitative nature of the result (average impact for the top 25%, 15%, 10% of units = 10.1%, 9.6% and 9.8% respectively).

This modest increase was surprising in light of the fact that nuisance modulations were 2.8x the target match signal (Fig 4b, compare cyan and dark red bars), coupled with the intuition that large nuisance modulation should be highly detrimental to task performance (Fig 1b). However, this result can be understood by examining the trial variability component of the noise, which was 5.2x larger than the target match signal and nearly 2x larger than the nuisance component (Fig 4b, purple bar) and as a result, dominated the denominator of the d' derivation. As an illustrative example, compare ratios of the numbers 5/(10+100)=0.045 versus 5/(0+100)=0.05; while the first component of the denominator, 10, is 2-fold the size of the numerator (5), including versus excluding it only leads to a change in the total ratio of 10% because the denominator is dominated by the second entry, 100. Consequently, although the amount of nuisance modulation is large relative to the size of the target match signal, its impact is blunted by the existence of trial variability, which is even larger. Stated differently, while IT nuisance modulations are larger than the IT target match signal, both are small relative to the size of trial variability. Because trial variability is so much larger than nuisance variability, the existence of nuisance modulation has little consequence for d'.

Large trial variability in IT is a consequence of fast processing

Why is trial variability so much larger than nuisance modulation (and signal modulation) in our data? During the IDMS task, spike count windows were short, as a consequence of terminating the count window before the monkeys' reaction times, which were fast (Fig 2e; total counting window duration 170 ms, 80-250 ms following stimulus onset). Within these short spike count windows, the average grand mean spike count was 0.94 spike per condition per trial, and the average peak spike count across the 80 conditions was 2.63 spikes (which translates into mean and peak firing rates of 5.5 spikes/sec and 15.5 spikes/sec, respectively). We also found that, consistent with earlier reports, IT trial variability was approximately Poisson (average variance-to-mean ratio across units = 1.20, relative to the Poisson benchmark of 1.0). Simple simulations confirm that within a low spike count, Poisson regime, trial variability is much larger than signal modulation. Large trial variability in IT thus does not arrive from exotic mechanisms, rather, it is a natural consequence of the low spike counts that follow from fast processing, coupled with Poisson-like trial variability.

To illustrate how the impact of nuisance modulation depends on overall spike count, we recalculated the impact of nuisance modulation as a function of increasing window size. In this analysis, we always started the spike count window for each unit at 80 ms following stimulus onset, and we ended the count window at different times up to 170 ms total duration (equivalent to the count window for the analyses presented in Fig 4b-c). These results illustrate a systematic increase in the impact of nuisance modulation on task performance as a function of spike count window duration (Fig 4d), consistent with the interpretation that the impact of nuisance modulation is inversely proportional to the overall spike count.

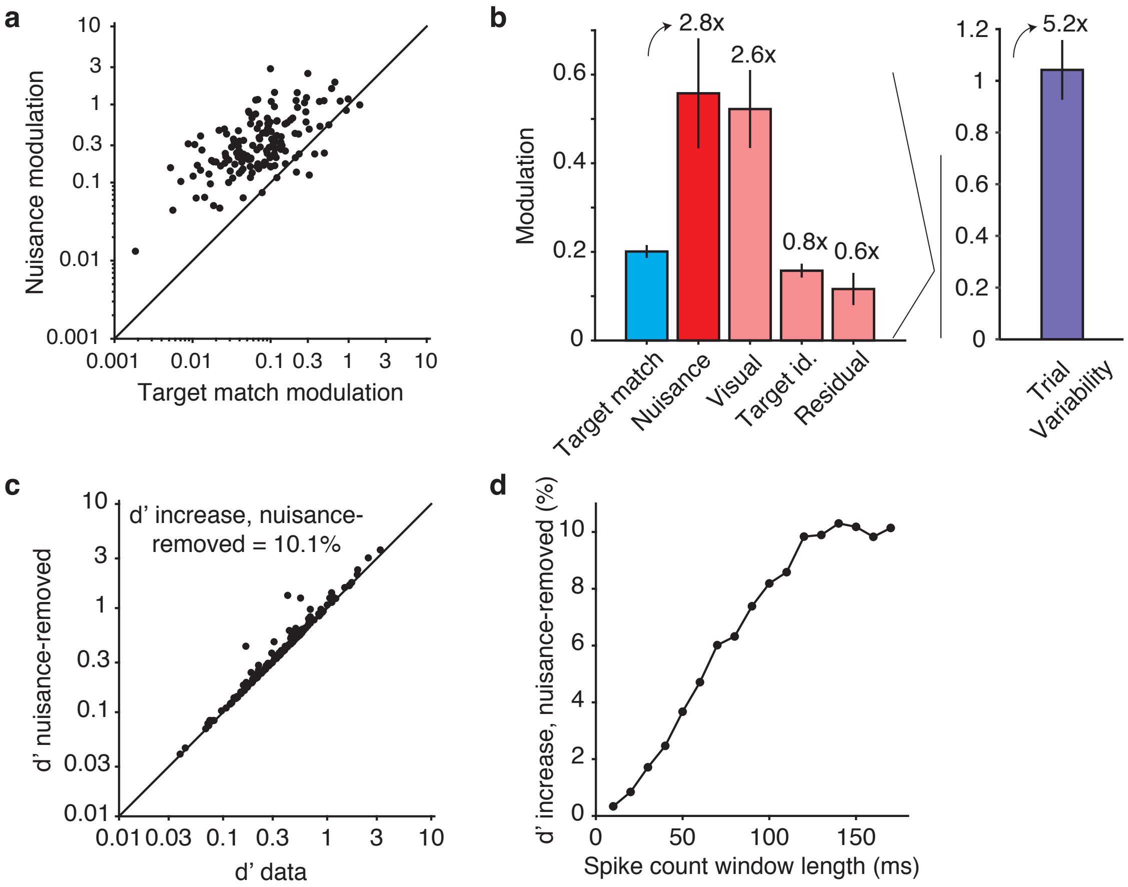

Figure 3-4. *The impact of nuisance modulation on single-unit d'.* Modulations were computed for each type of experimental parameter, in units of the standard deviations around each unit's grand mean spike count (see Results). **a**) Total nuisance modulation plotted against target match modulation for each unit. **b**) Average modulation magnitudes across units, parsed into target match modulation (cyan), combined nuisance modulation (dark red), and different nuisance modulation subtypes (light red) including visual, target identity, and residual. The right subpanel indicates the size of trial variability, computed in a comparable way. Error bars represent standard error across units. Numbers above each type of nuisance modulation indicate its size relative to the target match signal. **c)** Single-unit d' computed on the intact data and with the

nuisance-term set to zero. The average proportional impact of nuisance was computed as the average proportional increase in performance when nuisance was removed. **d)** The average impact of nuisance modulation on single-unit d’ (computed as described in panel c), applied to data using spike count windows of increasing size.

To illustrate that the amount of signal mixing we observed would have impacted task performance at higher spike counts than we recorded in our data (e.g. if counting windows were longer and/or firing rates were higher), we performed a simulation in which we rescaled the responses for each unit in our data (after noise correction, see Methods). Specifically, we kept the proportions and types of signal and nuisance modulation for each unit intact, but rescaled the trial-averaged spike count responses for each unit by different factors of N, followed by the reintroduction of Poisson trial variability. We then recomputed the impact of nuisance modulation on single-unit d’ as described for Fig 4c-d. We found that the impact of nuisance on d’ grew substantially with rescaling (Fig 5a). For example, with a 6-fold rescaling, which roughly translates into a 1 second counting window (under the assumption that the response properties are constant with time), eliminating nuisance resulted in a 53.0% increase in d’ (as compared to the 12.1% increase in simulation with no rescaling; Fig 5a). The increased impact of nuisance with rescaling cannot not be attributed to changes in the relative amounts of signal and nuisance modulation, as these remained fixed with rescaling (compare Fig 4b and 5b, cyan, red). Rather, the increased impact of nuisance with rescaling is due to a decrease in magnitude of trial variability relative to the magnitudes of signal and nuisance modulation (compare Fig 4b and 5b, purple).

Together, these results indicate that mixing signals in a fast processing regime (where spike counts are low) has the unexpected consequence that nuisance modulation is largely inconsequential for task performance. In contrast, our simulations reveal that mixing signals in the same proportions but in regime where spike counts are high (e.g. with long integration windows and/or higher firing rates) would be highly detrimental. These results thus suggest that within IT during the IDMS task, the potentially deleterious impact of nuisance modulation is blunted by virtue of a fast processing, low spike count regime.

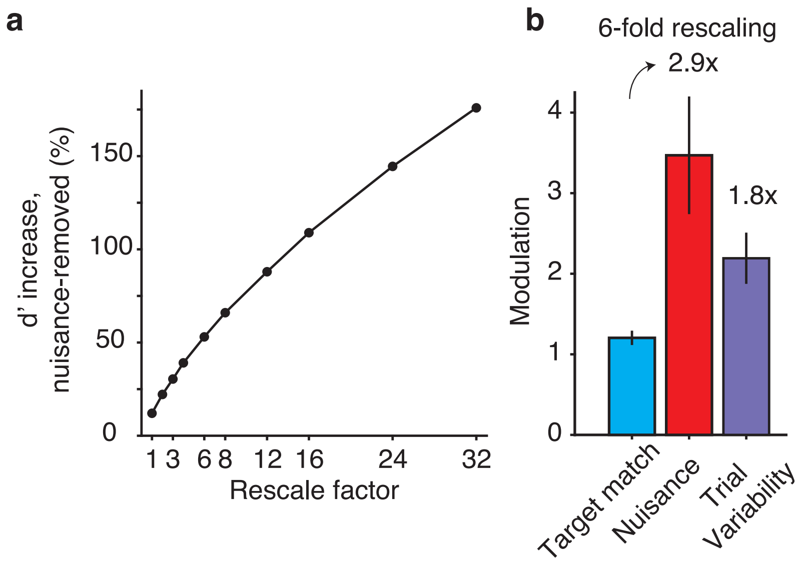

Figure 3-5. *Nuisance modulation is predicted to be detrimental for higher spike counts.* **a)** The simulated impact of nuisance modulation on single-unit d' as a function of rescaling the spike counts for each unit. **b)** Average modulation magnitudes across simulated units, for the 6-fold spike count rescaling data point in subpanel a.

The impact of nuisance modulation on population performance is also modest

In the previous section, we examined the impact of nuisance modulation as it applies to single unit performance. Next, we were interested in the impact of nuisance modulation on the performance of the neural population. Specifically, any particular population decoding scheme defines an axis in population space, and of interest is whether or not the single neuron intuitions established above hold when nuisance modulations are projected along this population axis. The simplest possible assumption for such a population decoding scheme is that every IT unit receives an equal weight of one. In such a decoding scheme, the brain simply counts the spikes of all of the units in the population to determine whether or not an image is a match or a distractor (i.e. this decoding scheme is equivalent to the performance of a spike count classifier on the population responses). The impact of the projected nuisance modulation along this axis is equivalent to the average impact of nuisance modulation across single units, which we have shown to be modest (Fig 4c.) The next simplest population decoding scheme, producing a different axis in population space, is a more traditional weighted linear readout. In this type of readout, IT units are weighted proportional to the amount of task-relevant information that they carry and interactions between units are taken into consideration. In particular, we have shown that one such linear readout, the Fisher Linear Discriminant (FLD), is behaviorally relevant insofar as it reflects misclassifications on trials in which the monkeys make errors (Fig 3c). While this axis might not be exactly the one that the monkey is using to distinguish matches from distractors, it does in fact captures information relevant to that discrimination. For the following is, we thus assume that this is in fact the axis that the monkeys are using to make this discrimination, and we examine the impact of nuisance modulations when projected along this axis.

As we demonstrate in this section, the impact of nuisance modulation on IT performance described above for single units (Fig 4c), remains modest even when population factors are considered. To address population considerations, we begin with a data-based “pseudosimulation” approach that allows us to compute important benchmarks for our results. However, because these simulations require assumptions about the data, we also verify our results with analyses applied directly to neural data.

To estimate the impact of nuisance modulation on IT population performance, we applied an approach similar in concept to the single-unit analysis presented in Fig 4c, where we estimated the impact of nuisance by comparing the intact data with a hypothetical version of our data with nuisance removed. However, in the case of the population, we did not have an analytical solution and we thus performed pseudosimulations to determine it. To perform this analysis, we simulated the responses of two versions of each unit: an intact version with the same number and types of signals as well as the same grand mean spike count (after noise correction, see Methods), and a version in which the nuisance modulation was removed. In both cases, we simulated trial variability for each unit with an independent, Poisson process. Cross-validated linear decoder performance, measured in units of population d’, grew with increasing population size for the intact and nuisance-removed populations with an approximately fixed ratio (Fig 6a). The proportional impact of nuisance modulation as a function of population size saturated at ~18% with larger sized populations (Fig 6b). These results suggest that the modest impact of nuisance modulation measured in individual units remains modest across the population (under the assumption that trial variability is Poisson and is independent between units).

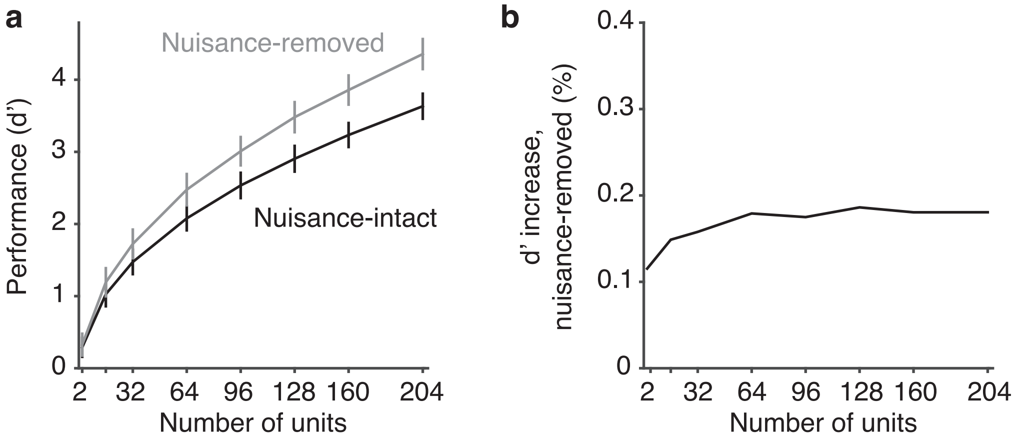

Figure 3-6. *Estimating the impact of nuisance modulation on population performance.* **a)** Linear decoder performance, shown in units of population d', as a function of population size for two simulated populations: "Nuisance-Intact": a version of our data in which the responses of each unit are replicated (after noise-correction), coupled with independent, Poisson trial variability; "Nuisance-removed": a similar version of our data, but with the nuisance modulations for each unit set to zero (see Methods). Error bars (SEM) reflect the variability that can be attributed to the random selection of units (for populations smaller than the full dataset) and the random assignment of training and testing trials in cross-validation. **b)** The proportional impact of nuisance (computed as the proportional increase in performance when nuisance was removed), plotted as a function of population size, computed for the data shown in panel a.

Our simulation-based approach allowed us to estimate the impact of nuisance modulation on population performance relative to a benchmark of the same population but without nuisance. However, our pseudosimulations incorporate the assumption that trial variability is independent (i.e. uncorrelated) between units, whereas we do in fact expect it to be weakly correlated (e.g. Cohen & Maunsell, 2009). How might the existence of weakly correlated variability impact our results? To summarize the well-established framework for thinking about correlated trial variability (reviewed by Averbeck & Lee, 2006; Cohen & Kohn, 2011; Kohn et al., 2016), when the component of trial variability that falls along a linear decoding axis is uncorrelated between neurons, it will average away as a function of population size. Relative to this benchmark, correlated trial variability has the potential to either be beneficial or detrimental to performance (Fig 7a). We have determined that nuisance modulation is similar insofar as the component of nuisance modulation that falls along a linear decoding axis that is uncorrelated between neurons will average away as a function of population size. Relative to this benchmark, interactions between neurons can configure nuisance modulation to have beneficial or detrimental consequences (Fig 7b).

When a task does not include nuisance variability (e.g. a two-way discrimination between exactly two conditions), the impact of correlated trial variability on population performance can be measured by comparing performance for the simultaneously recorded, intact data with performance when the trials are independently shuffled for each unit to destroy correlations (Averbeck & Lee, 2006). Increases in performance with shuffling indicate that noise correlations are detrimental (Fig 7a, left) whereas decreases in performance indicate that noise correlations are beneficial (Fig 7a, right). This shuffling procedure can be extended for tasks that incorporate a nuisance component by comparing population performance for the intact data with performance when the experimental conditions are shuffled independently for each unit within each class (i.e. shuffling conditions within the set of target matches and within the set of distractors).

To assess the impact of both correlated trial and nuisance variability on IT population performance, we analyzed the raw, simultaneously recorded data within each session. Here we present the results only for populations of size 24 (to simplify the data, given the number of comparisons of interest). Relative to the intact data, shuffling trial variability resulted in a small increase in performance (Fig 7c, “Intact” versus “Shuffle TV”; proportional increase with shuffling = 8%), indicating that correlated trial variability is aligned along the target match decoding axis in a manner that is weakly detrimental. Next we computed performance when both trial and nuisance variability were shuffled, and found that it was slightly higher than shuffling trial variability alone (Fig 7c, “Shuffle TV&NV”; proportional increase = 7%). This suggests that like trial variability, nuisance variability is correlated in a manner weakly detrimental to performance.

How does the existence of weakly detrimental correlated trial and nuisance variability impact the results presented in Fig 6? First, note that the analysis presented in Fig 6 is not impacted by the existence of correlated trial variability (because any correlations that existed were destroyed in the pseudosimulation process). Second, note that Fig 6 presents an estimate of the “total” impact of nuisance variability that captures contributions arising from both the existence of nuisance modulations as well as any detrimental correlations that fall along the decoding axis. To parse their relative contributions, we returned to the pseudosimulation and applied the nuisance shuffling procedure. Shuffling nuisance variability led to a small proportional increase (relative to shuffling trial variability alone; Fig 7d; 8%) that was similar to the value measured for the intact data (7%, as described above). The remaining proportional impact of nuisance modulation, calculated as the increase between shuffling nuisance and removing it altogether, was 10% (Fig 7d; “Shuffle TV&NV” vs. “Shuffle TV, remove NV”).

To summarize these results, we measured the impact of nuisance modulation on population performance in simulation by comparing performance of an intact population (with independent trial variability) with a simulation of the same population with nuisance variability removed. In our data, the impact of nuisance modulation was modest (~18%) and approximately flat as a function of population size. An analysis targeted at understanding how correlated trial and nuisance variability between units impacts task performance revealed that their contributions to task performance were also measurable but modest, and did not change the interpretation that while nuisance modulation is large in IT, its impact on task performance (both for single units and for the population) is small.

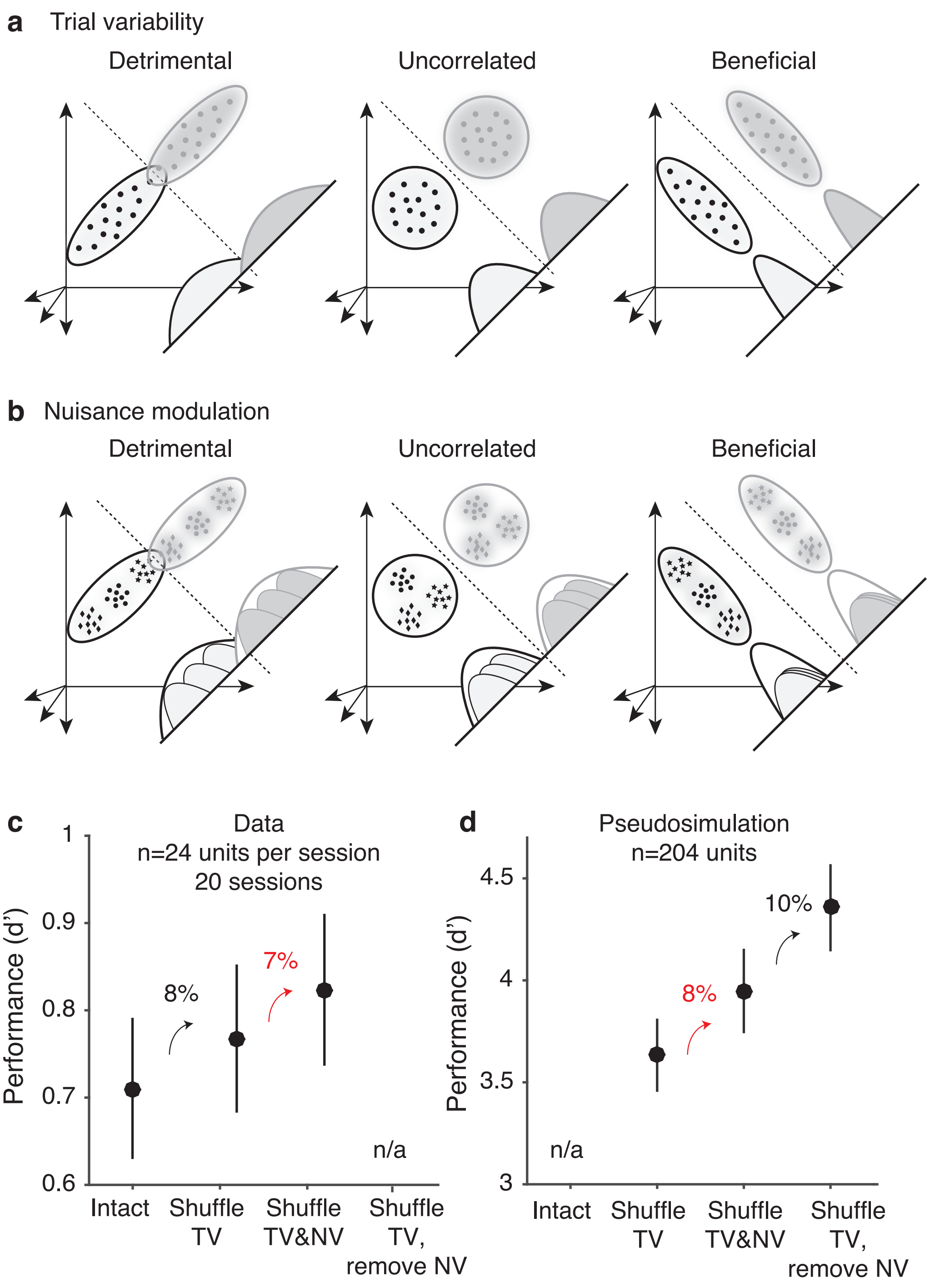
a
Trial variability
Detrimental
Uncorrelated
Beneficial
b
Nuisance modulation
Detrimental
Uncorrelated
Beneficial
c
Data
n=24 units per session
20 sessions
Performance (d')
1
0.9
0.8
0.7
0.6
8%
7%
n/a
Intact
Shuffle TV
Shuffle TV&NV
Shuffle TV, remove NV
d
Pseudosimulation
n=204 units
Performance (d')
4.5
4
3.5
3
10%
8%
n/a
Intact
Shuffle TV
Shuffle TV&NV
Shuffle TV, remove NV

Figure 3-7. *Understanding how correlated trial variability and correlated nuisance modulations impact task performance.* **a)** Shown are cartoon depictions of the “beneficial” and “detrimental” impact that correlated trial variability can have on task performance relative to the “uncorrelated” benchmark. Each point depicts a hypothetical population response for a population of two neurons on a single trial, and clusters of points depict the dispersion of responses across repeated trials. Dotted lines depict the linear decision boundary optimized for a two-way classification. Population performance is determined by projecting each class onto an axis perpendicular to the decision boundary. Correlated trial variability between units can be configured to increase or decrease the variance of the projected population response relative to benchmark of uncorrelated trial variability, and thus have a detrimental or beneficial impact on performance. **b)** Same as in a, but expanded to incorporate correlated nuisance variability. Included are 3 experimental conditions within each set (clusters of points). Like trial variability, correlated nuisance variability between units can be configured to increase or decrease the variance of the projected population response, relative to benchmark of uncorrelated nuisance variability. **c)** To assess the impact that correlated trial and nuisance variability between units has on population performance, we applied shuffling procedures to the raw data recorded within each session (across 20 sessions). Shown is linearly decoded population performance (d’) for populations of size 24 for: “Intact” – without shuffling; “Shuffle TV” – shuffled trial variability while maintaining nuisance variability correlations intact; and “Shuffle TV&NV” – shuffling both trial and nuisance variability. This analysis cannot be performed in a manner that determines what happens when nuisance variability is removed, indicated by the placeholder “n/a” for comparison with subpanel d. **d)** The same pseudosimulation data presented in Fig 6

(n = 204). Because that data is simulated as independent between units, the “Intact” condition cannot be computed, as indicated by the placeholder n/a for comparison with panel c. Shown is linearly decoded population performance for: “Shuffle TV” – shuffled trial variability while maintaining nuisance variability correlations intact; “Shuffle TV&NV” – shuffling both trial and nuisance variability; “Shuffle TV, remove NV”: shuffling trial variability and removing nuisance variability. In both c and d, numbers above the arrows indicate the proportional increase in d’. Error bars (SEM) reflect the variability that can be attributed to the random assignment of training and testing trials in cross-validation.

DISCUSSION

In many everyday situations, we are faced with the challenge of extracting one type of information from our environment while ignoring many other things that are going on around us. This study was inspired by a very simple intuition: when the neurons involved in computing the solutions for these tasks are modulated by both task-relevant signals as well as task-irrelevant nuisance information, nuisance modulation should be a source of noise that limits our ability to perform these tasks. Unexpectedly, we found that this simple intuition was largely wrong in IT. During a visual target search task, we found that nuisance modulations in IT were indeed large and that they were mixed with task-relevant signals in the responses of individual units, however, their consequences for task performance were modest. This result could be explained by the existence of another noise source, trial variability, which was larger than nuisance variability and blunted its impact on performance. Large trial variability in IT could, in turn, be accounted for by fast processing (implied by fast reaction times), which positioned IT within a low spike count regime, coupled with trial variability that was approximately Poisson. We found that these results applied not only to individual units but also to the performance of the IT population. Our results thus reveal that when the brain operates in a regime where signals are small relative to the size of trial variability, nuisance modulations are of very little consequence to task performance.

Many of our intuitions about neural coding have been developed within the context of a high spike count regime, largely following on foundational work in early and mid-level visual brain areas in primates (e.g. V1, MT) where firing rates are high. Notably, recent work has called into question whether even in those brain areas, high spike counts do in fact translate into a high signal-to-noise ratio, due to supra-Poisson trial variability that begins to dominate when spike counts are large (Goris, Movshon, & Simoncelli, 2014). Moreover, the low spike count regime that we present here is likely to be representative of the operating regime in many brain areas during many real-world tasks. The unexpected nature of our results highlights the fact that in this low spike count regime, some of the basic intuitions that we have constructed about neural coding may not hold.

Our results shed insight into why the brain might continue to “mix” modulations for different task-relevant parameters within individual neurons, even at the highest stages. Growing evidence suggests that the brain does not seek to produce neurons with increasingly “pure selectivity” at higher stages of processing, but rather that the brain continues to mix modulations for different task-relevant parameters within individual neurons, both at the locus at which task-relevant solutions are computed, as well as downstream (Freedman & Assad, 2009; Kobak et al., 2016; Mante et al., 2013; Meister et al., 2013; Raposo et al., 2014; Rigotti et al., 2013; Rishel et al., 2013; Zoccolan et al., 2007). A number of explanations have been proposed to account for mixed selectivity. Some studies have documented situations in which signal mixing is an inevitable consequence of the computations required for certain tasks, such as identifying objects invariant to the view in which they appear (Zoccolan et al., 2007). Others have suggested that mixed selectivity may be an essential component of the substrate required to maintain a representation that can rapidly and flexibly switch with changing task demands (Raposo et al., 2014; Rigotti et al., 2013). Still others have maintained that broad tuning across different types of parameters is important for learning new associations (Rigotti et al., 2013). When viewed from the perspective that signal mixing introduces noise in the form of nuisance modulation, one might suspect that one or more of these benefits outweigh the performance costs associated with mixed selectivity. However as we demonstrate here, within the fast processing, low spike count regime that most of these high-level brain areas are likely to operate in, large nuisance modulations are expected to have only a modest impact on task performance.

The framework with which we explore how nuisance interactions between different neurons impact population performance builds on foundational work focused on correlated trial variability between units, or “noise correlations” (Averbeck & Lee, 2006; Cohen & Kohn, 2011; Kohn et al., 2016). Recent work has emphasized the importance of not just measuring the degree to which neurons are correlated, but how those correlations align with a decoding axis and thus how they impact performance (Moreno-Bote, 2014). In the visual search task we present here, we found that correlations between units in both trial and nuisance variability had a small, detrimental impact on performance. In other tasks, nuisance interactions along a decoding axis may be much more impactful – such as in the case of dissociating self versus object motion (Kim et al., 2016), and in those cases, other decoding schemes may be required to disambiguate signal from nuisance modulation.

Our results support the existence of a robust target match representation in IT during this task that reflects confusions on trials in which the monkeys make errors (Fig 3c); this result has not been reported previously. One earlier study also explored the responses of IT neurons in the context of a DMS task in which, like ours, the objects could appear at different identity-preserving transformations (Leuschow et al., 1994), but this study did not sort neural responses based on behavior. Target match signals have been investigated most extensively in IT via a classic version of the delayed-match-to-sample (DMS) paradigm where each trial begins with a visual cue indicating the identity of the target object, and this cue is often the same image as the target match. In this paradigm, approximately half of all IT neurons that differentiate target matches from distractors do so with enhanced responses to matches whereas the other half are match suppressed (Miller & Desimone, 1994; Pagan et al., 2013). Because match suppressed responses are thought to arise as the result of passive, stimulus repetition of the target match following the cue, some have speculated that the match enhanced neurons alone carry behaviorally-relevant target match information (Miller & Desimone, 1994). Conversely, others have argued that a representation comprised exclusively of match enhanced neurons would likely confuse the presence of a match with nuisance modulations that evoke changes in overall firing rate, such as changes in stimulus contrast (Engel & Wang, 2011). Additionally, these authors have proposed that matched suppressed neurons could be used in these cases to disambiguate target match versus nuisance modulation. Our results reveal that when target matches do not follow the presentation of the same visual image at a short time before (as is the case for natural object search), match suppression is very weak (Fig 3e), and consequently, in these cases, this specific disambiguation strategy cannot be employed. Our results also suggest that for the types of nuisance modulation that we have investigated here (changes in position, size and background context), its impact is modest and in these cases, such a strategy is not necessary.

In this report, we showed that the impact of nuisance modulation was modest, both while using a spike count classifier (which reads out target matches versus distractors by giving each unit the same weight; Fig 4c) and when using a FLD (which weighs each unit by to its ability differentiate target matches versus distractors; Fig 6). While the performance of the FLD correlated with the behavioral confusions of the monkeys (Fig 3c), it is possible that the brain uses a different decoding scheme to read out target matches versus distractors and drive behavior. However, it is likely that if such an alternate decoding scheme more optimally reads out target match information than the FLD does, the impact of nuisance modulation on decoding performance would remain small.

In a previous series of reports (Pagan & Rust, 2014a; Pagan, Simoncelli, & Rust, 2016; Pagan et al., 2013), we investigated target match signals in the context of the classic DMS design in which target matches were repeats of cues presented earlier in the trial and each object was presented on a gray background. One of our main findings from that work was that the IT target match representation was reflected in a partially nonlinearly separable format, whereas an IT downstream projection area, perirhinal cortex, contained the same amount of target match information but in a format that was largely linearly separable. In the data we present here, we also found evidence for a nonlinear component of the IT target match representation, reflected by higher performance of a maximum likelihood as compared to linear decoder (Fig 3b). However, in this study, a larger proportion of the IT target match representation was linear as compared to our previous DMS results. The source of these quantitative differences is unclear. They could arise from the fact that the IDMS task requires an “invariant” visual representation of object identity, which first emerges in a linearly separable format in the brain area that we are recording from (IT; Rust & DiCarlo, 2010), whereas the DMS task could rely on the visual representation at an earlier stage. Alternatively, these differences could arise from the fact that during IDMS, images are not repeated within a trial, and the stronger nonlinear component revealed in DMS may be produced by stimulus repetition. Our current data cannot distinguish between these alternatives.

METHODS

Experiments were performed on two adult male rhesus macaque monkeys (*Macaca mulatta)* with implanted head posts and recording chambers. All procedures were performed in accordance with the guidelines of the University of Pennsylvania Institutional Animal Care and Use Committee.

The invariant delayed-match-to-sample (IDMS) task:

All behavioral training and testing was performed using standard operant conditioning (juice reward), head stabilization, and high-accuracy, infrared video eye tracking. Stimuli were presented on an LCD monitor with an 85 Hz refresh rate using customized software (http://mworks-project.org).

As an overview, the monkeys' task required an eye movement response to a specific location when a target object appeared within a sequence of distractor images (Fig 2a). Objects were presented across variation in the objects' position, size and background context (Fig 2b). Monkeys viewed a fixed set of 20 images across switches in the identity of 4 target objects, each presented at 5 identity-preserving transformations (Fig 2c). We ran the task in short blocks (~3 min) with a fixed target before another target was pseudorandomly selected. Our design included two types of trials: cue trials and test trials (Fig 2a). Only test trials were analyzed for this report.

Trials were initiated by the monkey fixating on a red dot (0.15°) in the center of a gray screen, within a square window of ±1.5°, followed by a 250 ms delay before a stimulus appeared. Cue trials, which indicated the current target object, were presented at the beginning of each block and after three subsequent trials with incorrect responses. To minimize confusion, cue trials were designed to be distinct from test trials and began with the presentation of an image of each object that was distinct from the images used on test trials (a large version of the object presented at the center of gaze on a gray background; Fig 2a). Test trials, which are the focus of this report, always began with a distractor image, and neural responses to this image were discarded to minimize non-stationarities such as stimulus onset effects. Distractors were drawn randomly from a pool of 15 possible images within each block without replacement until each distractor was presented once on a correct trial, and the images were then re-randomized. On most trials, a random number of 1-6 distractors were presented, followed by a target match (Fig 2a). On a small fraction of trials, 7 distractors were shown, and the monkey was rewarded for fixating through all distractors. Each stimulus was presented for 400 ms (or until the monkeys' eyes left the fixation window) and was immediately followed by the presentation of the next stimulus. Following the onset of a target match image, monkeys were rewarded for making a saccade to a response target within a window of 75 – 600 ms to receive a juice reward. In monkey 1 this target was positioned 10 degrees below fixation; in monkey 2 it was 10 degrees above fixation. If 400 ms following target onset had elapsed and the monkey had not moved its eyes, a distractor stimulus was immediately presented. If the monkey continued fixating beyond the required reaction time, the trial was considered a "miss". False alarms were differentiated from fixation breaks via a comparison of the monkeys' eye movements with the characteristic pattern of eye movements on correct trials: false alarms were characterized by the eyes leaving the fixation window via its bottom (monkey 1) or top (monkey 2) outside the allowable correct response period and traveling more than 0.5 degrees whereas fixation breaks were characterized by the eyes leaving the fixation window in any other way. Within each block, 4 repeated presentations of the 20 images were collected, and a new target object was then pseudorandomly selected. Following the presentation of all 4 objects as targets, the targets were re-randomized. At least 20 repeats of each condition were collected. Overall, monkeys performed this task with high accuracy. Disregarding fixation breaks (monkey 1: 8% of trials, monkey 2: 11% of trials), percent correct on the remaining trials was as follows: monkey 1: 87% correct, 3% false alarms, and 10% misses; monkey 2: 96% correct, 1% false alarms, and 3% misses.

Neural recording

The activity of neurons in IT was recorded via a single recording chamber in each monkey. Chamber placement was guided by anatomical magnetic resonance images in both monkeys, and in one monkey, Brainsight neuronavigation (https://www.rogue-research.com/). The region of IT recorded was located on the ventral surface of the brain, over an area that spanned 4 mm lateral to the anterior middle temporal sulcus and 15-19 mm anterior to the ear canals. Neural activity was largely recorded with 24-channel U probes (Plexon, Inc) with linearly arranged recording sites spaced with 100 mm intervals, with a handful of units recorded with single electrodes (Alpha Omega, glass-coated tungsten). Continuous, wideband neural signals were amplified, digitized at 40 kHz and stored using the OmniPlex Data Acquisition System (Plexon). Spike sorting was done manually offline (Plexon Offline Sorter). At least one candidate unit was identified on each recording channel, and 2-3 units were occasionally identified on the same channel. Spike sorting was performed blind to any experimental conditions to avoid bias. A multi-channel recording session was included in the analysis if the animal performed the task until the completion of 20 correct trials per stimulus condition, there was no external noise source confounding the detection of spike waveforms, and the session included a threshold number of task modulated units (>4 on 24 channels). The sample size (number of units recorded) was chosen to approximately match our previous work (Pagan & Rust, 2014a; Pagan et al., 2016; Pagan et al., 2013).

For all the analyses presented in this chapter, we measured neural responses by counting spikes in a window that began 80 ms after stimulus onset. For all analyses but Fig 4d, the spike count window ended at 250 ms. On 1.9% of all correct target match presentations, the monkeys had reaction times faster than 250 ms, and those instances were excluded from analysis such that spikes were only counted during periods of fixation. When combining the units recorded across sessions into a larger pseudopopulation, we screened for units that met three criteria. First, units had to be modulated by our task, as quantified by a one-way ANOVA applied to our neural responses (80 conditions * 20 repeats) with $p < 0.01$. Second, we applied a loose criterion on recording stability, as quantified by calculating the variance-to-mean for each unit (computed by fitting the relationship between the mean and variance of spike count across the 80 conditions), and eliminating units with a variance-to-mean ratio > 5. Finally, we applied a loose criterion on unit recording isolation, quantified by calculating the signal-to-noise ratio (SNR) of the waveform (as the difference between the maximum and minimum points of the average waveform, divided by twice the standard deviation across the differences between each waveform and the mean waveform), and excluding (multi)units with an SNR < 2. This yielded a pseudopopulation of 204 units (of 563 possible units), including 96 units from monkey 1 and 108 units from monkey 2.

Population performance

To determine the performance of the IT population at classifying target matches versus distractors, we applied two types of decoders: a Fisher Linear Discriminant (a linear decoder) and Maximum Likelihood decoder (a nonlinear decoder) using approaches that are described previously in detail (Pagan et al., 2013) and are summarized here.

When applied to the pseudopopulation data (Fig 3b, Fig 6a, Fig 7d), all decoders were cross-validated with the same resampling procedure. On each iteration of the resampling, we randomly shuffled the trials for each condition and for each unit, and (for numbers of units less than the full population size) randomly selected units. On each iteration, 18 trials from each condition were used for training the decoder, 1 trial was used to determine a value for regularization, and 1 trial from each condition was used for cross-validated measurement of performance.

To ensure that decoder performance was not biased by unequal numbers of target matches and distractors, on each iteration of the resampling we included 20 target match conditions and 20 (of 60 possible) distractor conditions. Each set of 20 distractors was selected to span all possible combinations of mismatched object and target identities (e.g. objects 1, 2, 3, 4 paired with targets 4, 3, 2, 1), of which there are 9 possible sets. When computing proportion correct (Fig 3b), a mean performance value was computed on each resampling iteration by averaging binary performance outcomes across the 9 possible sets of target matches and distractors, each which contained 40 test trials. Mean and standard error of performance was computed as the mean and standard deviation of performance across 2000 resampling iterations. When computing population d' (Fig 6a, Fig 7d), d' was computed on each resampling iteration for the 20 target match conditions and 20 distractor conditions, separately for each set of 9 match/distractor combinations, and then averaged across the 9 sets. Mean and standard error of population d' was computed as the mean and standard deviation of d' across 2000 resampling iterations. For both measures, standard error thus reflected the variability due to the specific trials assigned to training and testing and, for populations smaller than the full size, the specific units chosen.

Fisher Linear Discriminant:

The general form of a linear decoding axis is:

(1) $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b,$

where **w** is an N-dimensional vector (where N is the number of units) containing the linear weights applied to each unit, and b is a scalar value. We fit these parameters using a Fisher Linear Discriminant (FLD), where the vector of linear weights was calculated as:

(2) $\boldsymbol{w} = \Sigma^{-\mathbf{1}}(\mu_1 - \mu_2)$

and b was calculated as:

(3) $b = \boldsymbol{w} \cdot \frac{1}{2}(\mu_1 + \mu_2) = \frac{1}{2}{\mu_1}^T\Sigma^{-\mathbf{1}}\mu_1 - \frac{1}{2}{\mu_2}^T\Sigma^{-\mathbf{1}}\mu_2$

Here $\mu_1\ and\ \mu_2$ are the means of the two classes (target matches and distractors, respectively) and the mean covariance matrix is calculated as:

(4) $\Sigma = \frac{\Sigma_1+\Sigma_2}{2}$

where Σ_1 and Σ_2 are the regularized covariance matrices of the two classes. These covariance matrices were computed using a regularized estimate equal to a linear combination of the sample covariance and the identity matrix I (Pagan et al., 2016):

(5) $\Sigma_i = \ \gamma\, \Sigma_i + (1-\gamma) \cdot I$

We determined γ by exploring a range of values from 0.01 to 0.99, and we selected the value that maximized average performance across all iterations, measured with the cross-validation “regularization” trials set aside for this purpose (see above). We then computed performance for that value of γ with separately measured “test” trials, to ensure a fully cross-validated measure. Because this calculation of the FLD parameters incorporates the off-diagonal terms of the covariance matrix, FLD weights are optimized for both the information conveyed by individual units as well as their pairwise interactions.

We computed two measures of performance: proportion correct (Fig 3b-c), and population d’ (Fig 6a). Each calculation began by computing the dot product of the test data and the linear weights **w**, adjusted by *b* (Eq. 1). Proportion correct was computed as the fraction of test trials that were correctly assigned as target matches and distractors, according to their true labels. Population d' was computed for the distributions of these values across the 20 different images presented as target matches versus as distractors:

(6) $d' = \frac{|\mu_{Match} - \mu_{Distractor}|}{\sigma_{pooled}}$,

where μ_{Match} and $\mu_{Distractor}$ correspond to the mean across the set of matches and distractors, $\sigma_{pooled} = \sqrt{\frac{\sigma^2_{Match} + \sigma^2_{Distractor}}{2}}$, and σ_{Match} and $\sigma_{Distractor}$ correspond to the standard deviation across the set of matches and distractors, respectively.

To compare FLD performance on correct versus error trials (Fig 3c), we used the same methods described above with the following modifications. First, the analysis was applied to the simultaneously recorded data within each session, and the correlation structure on each trial was kept intact on each resampling iteration. Second, when more than 24 units were available, a subset of 24 units were selected as those with the most task modulation, quantified via the p-value of a one-way ANOVA applied to each unit's responses (80 conditions * 20 repeats). Finally, on each resampling iteration, each error trial was randomly paired with a correct trial of the same condition and cross-validated performance was performed exclusively for these pairs of correct and error responses. As was the case for the pseudopopulation analysis, training was performed exclusively on correct trials. A mean performance value was computed on each resampling iteration by averaging binary performance outcomes across all possible error trials and their condition-matched correct trial pairs, and averaging across different recording sessions. Mean and standard error of performance was computed as the mean and standard deviation of performance across 2000 resampling iterations. Standard error thus reflected error in a manner similar to the pseudopopulation analysis - the variability due to the specific trials assigned to training and testing and, for populations smaller than the full size, the specific units chosen.

To determine the impact of correlated trial and nuisance variability on IT population performance (Fig 7c), we compared the FLD applied to the simultaneously recorded data as described above where the correlation structure on each trial was kept intact on each resampling iteration (Fig 7c, “intact”), with two different shuffling procedures. In the first, we randomly shuffled the trials within each condition, for each unit, on each iteration of the bootstrap (Fig 7c, “Shuffle TV”). In the second, we randomly shuffled both trial variability as well as the assignment of image identity for each the 20 distractor conditions and 20 target match conditions on each bootstrap iteration (Fig 7c, “Shuffle TV & NV”). The analysis to determine the impact of correlated nuisance variability on the pseudosimulation (Fig 7d) was performed in the same manner, but applied to the pseudosimulated data.

Maximum likelihood decoder:

As a measure of total IT target match information (combined linear and nonlinear), we implemented a maximum likelihood decoder (Fig 3b). We began by using the set of training trials to compute the average response r_{uc} of each unit u to each of the 40 conditions c. We then computed the likelihood that a test response k was generated from a particular condition as a Poisson-distributed variable:

(7) $lik_{u,c}(k) = \frac{(r_{uc})^{k} \cdot e^{-r_{uc}}}{k!}$

The likelihood that a population response vector was generated in response to each condition was then computed as the product of the likelihoods of the individual units. Next, we computed the likelihood that each test vector arose from the category target match as compared to the category distractor as the product of the likelihoods across the conditions within each category. We assigned the population response to the category with the maximum likelihood, and we computed performance as the fraction of trials in which the classification was correct based on the true labels of the test data.

Quantifying single-unit modulation magnitudes

To quantify the degree to which the firing rates of individual units were modulated by whether an image was presented as a target match versus as a distractor (Fig 3d), we calculated a target match modulation index for each unit by computing its mean spike count response to target matches and to distractors, and computing the ratio of their difference and their sum.

To quantify the degree to which individual units were modulated by different types of task parameters, we applied a bias-corrected, ANOVA-like procedure described in detail by (Pagan & Rust, 2014b) and summarized here. As an overview, this procedure considers the total variance in the spike count responses for each unit across conditions (n=80) and trials for each condition (m=20), and parses this total variance into the variance that can be attributed to each type of experimental parameter and variance attributed to trial variability. Similar to an ANOVA, the procedure is designed to parse response variance, including the variance that can be attributed to changes in the identity of the visual image, the identity of the target object and whether each condition was a target match or a distractor. These variances are converted into measures of spike count modulation (i.e. standard deviation around each unit's grand mean spike count) via a procedure that includes bias correction for over-estimates in modulation due to noise.

The procedure begins by developing an orthonormal basis of 80 vectors designed to capture all types of modulation with intuitive groupings. The number of each type is imposed by the experimental design. This basis $\boldsymbol{b}$ included vectors $\boldsymbol{b}_i$ that reflected 1) the grand mean spike count across all conditions ($\boldsymbol{b}_1,$ 1 dimension), 2) whether the object in view was a target or a distractor ($\boldsymbol{b}_2,$ 1 dimension), 3) visual image identity ($\boldsymbol{b}_3 - \boldsymbol{b}_{21},$ 19 dimensions), 4) target object identity ($\boldsymbol{b}_{22} - \boldsymbol{b}_{24},$ 3 dimensions), and 5) "residual", nonlinear interactions between target and object identity not captured by target match modulation ($\boldsymbol{b}_{25} - \boldsymbol{b}_{80},$ 56 dimensions). A Gram-Schmidt process was used to convert an initially designed set of vectors into an orthonormal

basis.

Because this basis spans the space of all possible responses for our task, each trial-averaged vector of spike count responses to the 80 experimental conditions $\boldsymbol{R}$ can be re-expressed as a weighted sum of these basis vectors. To quantify the amounts of each type of modulation reflected by each unit, we began by computing the squared projection of each basis vector $\boldsymbol{b}_i$ and $\boldsymbol{R}$. An analytical bias correction, described and verified in (Pagan & Rust, 2014b), was then subtracted from this value:

(8) $w_i^2 = (\boldsymbol{R} \cdot \boldsymbol{b}_i^T)^2 - \frac{\sigma_t^2 \cdot (\boldsymbol{b}_i^T)^2}{m}$

where σ_t^2 indicates the trial variance, averaged across conditions (n=80), and where m indicates the number of trials (m=20). When more than one dimension existed for a type of modulation, we summed values of the same type. Next, we applied a normalization factor (1/(n-1) where n=80) to convert these summed values into variances. Finally, we computed the square root of these quantities to convert them into modulation measures that reflected the number of spike count standard deviations around each unit's grand mean spike count. Target match modulation was thus computed as:

(9) $\sigma_{TM} = \sqrt{\frac{1}{n-1} \cdot w_2^2}$

and nuisance modulation was computed as:

(10) $\sigma_{Nui} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=3}^{80} w_i^2}$

Similarly, to compute the different subtypes of nuisance modulation, we replaced the weights w_i^2 in Eq. 10 with the weights that corresponded to the orthonormal basis vectors corresponding to each subtype, including visual modulation ($i = 3\ to\ 21$), target modulation ($i = 22\ to\ 24$), and 3) residual modulation ($i = 25\ to\ 80$), as described above.

We computed the trial variability for each unit ($\sigma_{Trial},$) in an comparable manner as the square root of the average (across conditions) variance across trials:

$$(11)\ \sigma_{Trial} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{m-1} \cdot \sum_{t=1}^{m} (s_{it} - s_i)}^{2}$$

where the spike count response for a particular trial t of condition i was s_{it}, and the mean spike count response across all trials of condition i was s_i.

When estimating modulation for individual units, (Fig 4a), the bias-corrected squared values were rectified for each unit before taking the square root. When estimating modulation population means (Fig 4b, 5b), the bias-corrected squared values were averaged across units before taking the square root. Because these measures were not normally distributed, standard error about the mean was computed via a bootstrap procedure. On each iteration of the bootstrap (across 1000 iterations), we randomly sampled values from the modulation values for each unit in the population, with replacement. Standard error was computed as the standard deviation across the means of these newly created populations.

Relating modulation magnitudes and single unit performance (d’):

To determine the impact of nuisance modulation on single unit task performance (Fig 4c-d, Fig 5a) we re-expressed d’ (Eq. 6) as a function of the different types of signal modulations described above (Eqs. 8-10):

(12) $d' = \frac{|\mu_{Match} - \mu_{Distractor}|}{\sigma_{pooled}} = \sqrt{\frac{a \cdot \sigma_{TM}^2}{b \cdot \sigma_{Nui}^2 + \sigma_{Trial}^2}}$ where $a = \frac{n-1}{3}$, and $b = \frac{n-1}{n}$

This derivation is described in detail in (Pagan & Rust, 2014b).

To quantify the impact of nuisance modulation on single unit performance (d’), we compared each unit’s d’ in the presence of nuisance modulation (Eq. 12) versus d’ when the nuisance modulation term σ_{Nui} was set to zero (d’$_{NoNui}$). We then calculated the impact of nuisance modulation as the percent increase in d’ without nuisance:

(13) $Impact = \left(\frac{d'_{NoNui}'}{d'} - 1\right) \cdot 100\%$

Simulations

To better understand our results, we performed a number of data-inspired simulations. Each simulation began by computing the bias-corrected weights for each unit as described above (Eq. 8).

To explore how rescaling the spike counts by different factors of N influenced the impact of nuisance modulation (Fig 5), we rectified bias-corrected modulations that fell below zero, recomputed the noise-corrected mean spike count responses for each condition, rescaled the mean spike counts by N, and generated trial variability with an independent Poisson process.

To estimate the impact of nuisance modulation on population performance, we simulated two versions of each of our recorded units (Fig 6a compare “Nuisance-intact” to “Nuisance-removed”; Fig 7d compare “Shuffle TV” and “Shuffle TV & NV” to “Shuffle TV, remove NV”). In the “Intact” version, we computed each unit’s responses as described for the rescaling simulation but with a rescale factor N = 1. In the “Nuisance removed” version, we used a similar procedure but set the modulations corresponding to all nuisance dimensions to zero. The responses were thus computed based on the grand mean spike count response as well as the target match modulation alone.

Statistical tests

When comparing population decoding measures (Fig 3b), we reported *P* values as an evaluation of the probability that differences were due to chance. We calculated these *P* values as the fraction of resampling iterations on which the difference was flipped in sign relative to the actual difference between the means of the full data set (for example, if the mean of decoding measure 1 was larger than the mean of decoding measure 2, the fraction of iterations in which the mean of measure 2 was larger than the mean of measure 1).

CHAPTER 4

General Conclusions

In this dissertation, we examined the responses of populations of neurons recorded in V4 and IT as monkeys performed an invariant delayed match to sample object search task. Our results showed that information about whether the currently viewed stimulus matches a sought target is reflected by populations of neurons in both V4 and IT, but these signals are larger in IT. These results suggest that top-down context-specific modulations are integrated into the ventral visual pathway at multiple stages. Next, we focused on responses recorded from IT from a neural coding perspective. We found that while modulations in IT that were expected to act as noise (nuisance modulations) were large, they unexpectedly had little impact on neural task performance. In this chapter, we discuss the implications of our results and some possible future directions.

The role of V4 and IT in visual search

In Chapter 2, we compared neural responses in V4 and IT while monkeys performed invariant object search. In this study, we sought to differentiate between two scenarios of how the solution to this task might be computed: one in which top-down, context-specific signals are introduced at multiple stages of the ventral visual pathway, and another in which V4 is the single locus for that combination. We found multiple lines of evidence supporting the hypothesis that context-specific signals are introduced at multiple stages of the ventral visual pathway. First, we found that the V4 population contains less total (and linearly separable) information for this task than the IT population does, suggesting that IT does not inherit all of its information from V4. Second, we found that V4 single units reflect information about target identity but not information that explicitly differentiates between target matches and distractors, while IT units reflect both of these types of information. Lastly, we found that while our measures of V4 single unit context modulation are largely consistent with previous reports, the same measures applied to IT reveal significantly more context-specific modulation.

The format of target-specific signals in V4 and IT

A large body of literature supports the idea that attentional modulation can affect the baseline firing rate, gain, or contrast sensitivity, with little effect on feature selectivity (Luck et al., 1997; McAdams & Maunsell, 2000; McAdams & Maunsell, 1999; Motter, 1994; Reynolds, Pasternak, & Desimone, 2000; Treue & Martinez Trujillo, 1999). In contrast, a recent series of studies suggests that attentional modulation can cause shifts in neural tuning both in V4 single units (David, Hayden, Mazer, & Gallant, 2008) and across the human brain (Cukur, Nishimoto, Huth, & Gallant, 2013). The V4 responses in our study do not seem to align with these results, as the context-specific modulations we found in V4 are primarily linear in format. However, it might be the case that tuning shifts do exist in IT. In particular, an idealized neuron whose responses are formatted as nonlinear combinations of visual and target signals which show the full solution to the task (full diagonal structure, i.e. Fig 2-5a, target match modulation) would reflect a full 'tuning shift': under each different target context, the visual tuning of the neuron is completely different. The extent to which we see tuning shifts in our data warrants further investigation in our data. Specifically, future work will include using both similar methodologies to those in (David et al., 2008), as well as extensions of our methods which parse modulation magnitudes in single unit responses (Pagan & Rust, 2014b).

In both V4 and IT, we found information that reflects the identity of the target match. Due to the design of our experiment, monkeys knew the identity of the target from the beginning of each block of trials. Consistent with this, the target identity information in both V4 and IT appears before stimulus onset, reinforcing the idea that it is a true working memory signal. However, the exact format of these signals in V4 and IT has not been fully investigated. In particular, do different subpopulations of neurons signal the identity of the target at different times across the stimulus presentation interval? Furthermore, do these working memory signals decrease in strength as a function of time after the presentation of the cue at the beginning of a block of trials, and do these signals correlate with behavior? These questions remain untested in our data. To address them, we are currently investigating the dynamics of the working memory signal, both within each stimulus presentation as well as across blocks of trials.

While previous studies of IT and PRH which studied classic DMS tasks (which include a sample stimulus at the beginning of each trial, as opposed to the beginning of a block) have mostly found mixes of match enhanced and match suppressed neurons, our results revealed mostly match enhanced neurons. This results is in line with a theory put forth by Miller and Desimone (1994), wherein match suppressed responses are thought to arise as the result of passive, stimulus repetition of the target match following the cue, while match enhanced neurons alone carry behaviorally-relevant target match information. However, Engel and Wang (2011) argued that a representation comprised exclusively of match enhanced neurons would likely confuse the presence of a match with nuisance modulations that evoke changes in overall firing rate, such as changes in stimulus contrast. In our experiment, the types of nuisance modulation that we investigated (changes in object identity, position, size, and background context) did not have a large impact on task performance, suggesting that this confusion does not occur. However this theory poses an interesting question that cannot be addressed by our data. How might changes in experimental parameters such as stimulus contrast, which are thought to be represented via increases in overall firing rate, impact the representation of the target match signal? Future experiments designed to test this would require a task in which objects are presented under a wider range of visual transformations that are expected to change stimulus contrast.

The transformation of target match information along the ventral visual pathway

In our study, we found that visual discriminability between V4 and IT was matched for equal sized populations of units, consistent with previous results (Rust & DiCarlo, 2010) (DiCarlo et al., 2012). Importantly, it could have been the case that we found we needed to record from larger numbers of neuron in one area to make fair comparisons (e.g. convergence or divergence ratios different than 1). Our results describe that the IT population contained more information than the V4 population, but there was still some small amount of total target match information in V4. Thus, there might exist a readout rule that could give preferential weights to a small subpopulation V4 neurons with the most target match information, and via a different convergence rule (e.g. 3x times more V4 neurons than IT neurons), match the amount of task-relevant information in IT. To test this, we plan to compare different computational models with the goal to model our IT responses based on V4 inputs. Notably, similar analyses from previous work in our lab (Pagan et al., 2016) have been successful at describing transformation between IT and PRH, though these areas were matched in their amounts of total target match information. Regardless, these analyses will lend insight to potential connectivity rules between V4 and IT.

We demonstrated that within our IT population, the representation of target match information was behaviorally relevant, insofar as it co-varied with the monkeys' behavior. While this result is consistent with IT playing a role in the generation of behavior, we did not directly establish a causal link. Lastly, a previous study from our lab established that a downstream area, PRH, contains the same amount of total target match information as IT, but it is formatted in a more linearly separable way (Pagan et al., 2013). Both causal and descriptive studies further elucidating the remaining components of the circuit responsible for invariant search are thus needed.

The role of signal and noise in determining task performance

In Chapter 3, we focused on the recorded responses in IT from a neural coding perspective. In this study, we sought to understand the role that signal and noise play in determining task performance. This is particularly important for performance in complex visual tasks such as invariant object search, where our brains must combine multiple types of information to arrive at the task solution. We expected that modulations differentiating between whether an image was a target match versus a distractor (target match modulations) would act as signal for the invariant object search task, and all other response modulations (nuisance modulations, such as responses that differentiate between the visual identity of different objects regardless of whether they are a target match) would act as performance-limiting noise. We found that, surprisingly, while nuisance modulation was large in IT, it had little impact on neural task performance.

These results follow from a low spike count regime, which were a result of both short spike counting windows (constrained by the monkeys' fast reaction times) and relatively low firing rates in IT cortex. In fact, our results suggest that at longer spike integration windows (with larger spike counts) or earlier brain areas with larger firing rates, the impact of nuisance variability on task performance will be greater. Specifically, within our data, we found an increase in the impact of nuisance modulation on task performance as a function of spike count window duration. However, this increase seems to begin to saturate at the end of our spike counting window (170 ms in length). Recent work has highlighted that the structure of trial variability deviates from a Poisson model, specifically, that trial variability is higher than the mean spike count, particularly for large spike counts (Goris et al., 2014). Together, these results imply that at longer spike integration windows, variability may deviate further from the Poisson model towards the end of a spike integration window. Thus, the ratio of trial variability to signal modulation might be expected to increase at these later time points. Since the small impact of nuisance on task performance in our data follows from large trial variability, what impact would such higher, supra-Poisson, trial variability have as spike counting windows are extended? As our monkeys' response times were quite short, we were unable to test this within the context of our data. Testing this in the context of a fixed duration task with a longer integration window may reveal more insights about the impact of noise on task performance.

These results follow from a low spike count regime, which were a result of both short spike counting windows (constrained by the monkeys' fast reaction times) and relatively low firing rates in IT cortex. In fact, our results suggest that at longer spike integration windows (with larger spike counts) or earlier brain areas with larger firing rates, the impact of nuisance variability on task performance will be greater. Specifically, within our data, we found an increase in the impact of nuisance modulation on task performance as a function of spike count window duration. However, this increase seems to begin to saturate at the end of our spike counting window (170 ms in length). Recent work has highlighted that the structure of trial variability deviates from a Poisson model, specifically, that trial variability is higher than the mean spike count, particularly for large spike counts (Goris et al., 2014). Together, these results imply that at longer spike integration windows, variability may deviate further from the Poisson model and be larger and larger towards the end of a spike integration window. Since the small impact of nuisance on task performance in our data follows from large trial variability, what impact would even higher, supra-Poisson, trial variability have as spike counting windows are extended? As our monkeys' response times were quite short, we were unable to test this in the context of our data. Testing this in the context of a fixed duration task with a longer integration window may reveal more insights about the impact of noise on task performance.

References:

Averbeck, B. B., & Lee, D. (2006). Effects of noise correlations on information encoding and decoding. *J Neurophysiol, 95*(6), 3633-3644. doi: 10.1152/jn.00919.2005

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science, 308*(5721), 529-534.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiology, 80*, 2918-2940.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363*, 345-347.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb Cortex, 11*(8), 761-772.

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat Neurosci, 14*(7), 811-819. doi: 10.1038/nn.2842

Cohen, M. R., & Maunsell, J. H. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci, 12*(12), 1594-1600. doi: 10.1038/nn.2439

Cukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci, 16*(6), 763-770. doi: 10.1038/nn.3381

David, S. V., Hayden, B. Y., Mazer, J. A., & Gallant, J. L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron, 59*(3), 509-521. doi: 10.1016/j.neuron.2008.07.001

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol, 57*(3), 835-868.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn Sci, 11*(8), 333-341.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415-434. doi: S0896-6273(12)00092-X [pii]

10.1016/j.neuron.2012.01.010

Engel, T. A., & Wang, X. J. (2011). Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *J Neurosci, 31*(19), 6982-6996. doi: 10.1523/JNEUROSCI.6150-10.2011

Eskandar, E. N., Optican, L. M., & Richmond, B. J. (1992). Role of inferior temporal neurons in visual memory. II. Multiplying temporal waveforms related to vision and memory. *J Neurophysiol, 68*(4), 1296-1306.

Eskandar, E. N., Richmond, B. J., & Optican, L. M. (1992). Role of inferior temporal neurons in visual memory I. Temporal encoding of information about visual images, recalled images, and behavioral context. *Journal of Neurophysiology, 68*, 1277-1295.

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

Freedman, D. J., & Assad, J. A. (2009). Distinct encoding of spatial and nonspatial visual information in parietal cortex. *J Neurosci, 29*(17), 5671-5680. doi: 10.1523/JNEUROSCI.2878-08.2009

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol, 61*(2), 331-349.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science, 173*(3997), 652-654.

Gattass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci, 8*(6), 1831-1845.

Gibson, J. R., & Maunsell, J. H. R. (1997). The sensory modality specificity of neural activity related to memory in visual cortex. *Journal of Neurophysiology, 78*, 1263-1275.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philos Trans R Soc Lond B Biol Sci, 351*(1346), 1445-1453. doi: 10.1098/rstb.1996.0129

Goris, R. L., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nat Neurosci, 17*(6), 858-865. doi: 10.1038/nn.3711

Haefner, R. M., & Bethge, M. (2010). Evaluating neural codes for inference using Fisher Information. . *Paper Presented at: Advances in Information Processing Systems*.

Haenny, P. E., Maunsell, J. H., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Exp Brain Res, 69*(2), 245-259.

Herrero, J. L., Gieselmann, M. A., Sanayei, M., & Thiele, A. (2013). Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron, 78*(4), 729-739. doi: 10.1016/j.neuron.2013.03.029

Holmes, E. J., & Gross, C. G. (1984). Stimulus equivalence after inferior temporal lesions in monkeys. *Behav Neurosci, 98*(5), 898-901.

Hong, H., Yamins, D. L., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat Neurosci, 19*(4), 613-622. doi: 10.1038/nn.4247

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*(5749), 863-866.

Huxlin, K. R., Saunders, R. C., Marchionini, D., Pham, H. A., & Merigan, W. H. (2000). Perceptual deficits after lesions of inferotemporal cortex in macaques [In Process Citation]. *Cereb Cortex, 10*(7), 671-683.

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology, 73*, 218-226.

Kim, H. R., Pitkow, X., Angelaki, D. E., & DeAngelis, G. C. (2016). A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons. *J Neurophysiol, 116*(3), 1449-1467. doi: 10.1152/jn.00005.2016

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., . . . Machens, C. K. (2016). Demixed principal component analysis of neural population data. *Elife, 5*. doi: 10.7554/eLife.10989

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol, 71*(3), 856-867.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., & Pouget, A. (2016). Correlations and Neuronal Population Information. *Annu Rev Neurosci, 39*, 237-256. doi: 10.1146/annurev-neuro-070815-013851

Kosai, Y., El-Shamayleh, Y., Fyall, A. M., & Pasupathy, A. (2014). The role of visual area V4 in the discrimination of partially occluded shapes. *J Neurosci, 34*(25), 8570-8584. doi: 10.1523/JNEUROSCI.1375-14.2014

Kubota, K., & Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J Neurophysiol, 34*(3), 337-347.

Law, C. T., & Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat Neurosci, 12*(5), 655-663. doi: 10.1038/nn.2304

Leuschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex, 5*, 523-531.

Li, N., Cox, D. D., Zoccolan, D., & Dicarlo, J. J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? *J Neurophysiol*.

Liu, Z., & Richmond, B. J. (2000). Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. *Journal of Neurophysiology, 83*(3), 1677-1692.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77*, 24-42.

Machens, C. K., Romo, R., & Brody, C. D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science, 307*(5712), 1121-1124. doi: 10.1126/science.1104171

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

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., . . . Kennedy, H. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb Cortex, 24*(1), 17-36. doi: 10.1093/cercor/bhs270

Maunsell, J. H., Sclar, G., Nealey, T. A., & DePriest, D. D. (1991). Extraretinal representations in area V4 in the macaque monkey. *Vis Neurosci, 7*(6), 561-573.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci, 29*(6), 317-322. doi: S0166-2236(06)00087-7 [pii]

10.1016/j.tins.2006.04.001

Maunsell, J. H. R., & Cook, E. P. (2002). The role of attention in visual processing. *Philos Trans R Soc Lond B Biol Sci, 357*(1424), 1063-1072.

McAdams, C. J., & Maunsell, J. H. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. *J Neurophysiol, 83*(3), 1751-1755.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neuroscience, 19*(431-441).

Meister, M. L., Hennig, J. A., & Huk, A. C. (2013). Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *J Neurosci, 33*(6), 2254-2267. doi: 10.1523/JNEUROSCI.2984-12.2013

Merigan, W. H., Nealey, T. A., & Maunsell, J. H. (1993). Visual effects of lesions of cortical area V2 in macaques. *J Neurosci, 13*(7), 3180-3191.

Meunier, M., Bachevalier, J., Mishkin, M., & Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *J Neurosci, 13*(12), 5418-5432.

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263*(5146), 520-522.

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

Mishkin, M., Prockop, E. S., & Rosvold, H. E. (1962). One-trial objectdiscrimination learning in monkeys with frontal lesions. *J Comp Physiol Psychol, 55*, 178-181.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron, 55*(1), 131-141. doi: 10.1016/j.neuron.2007.06.018

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron, 63*(6), 879-888. doi: 10.1016/j.neuron.2009.09.013

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science, 319*(5869), 1543-1546. doi: 10.1126/science.1150769

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782-784.

Moreno-Bote, R. (2014). Poisson-like spiking in circuits with probabilistic synapses. *PLoS Comput Biol, 10*(7), e1003522. doi: 10.1371/journal.pcbi.1003522

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol, 70*(3), 909-919.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J Neurosci, 14*(4), 2178-2189.

Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., & Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature, 433*(7026), 597-603.

Pagan, M., & Rust, N. C. (2014a). Dynamic target match signals in perirhinal cortex can be explained by instantaneous computations that act on dynamic input from inferotemporal cortex. *J Neurosci, 34*(33), 11067-11084. doi: 10.1523/JNEUROSCI.4040-13.2014

Pagan, M., & Rust, N. C. (2014b). Quantifying the signals contained in heterogeneous neural responses and determining their relationships with task performance. *J Neurophysiol, 112*(6), 1584-1598. doi: 10.1152/jn.00260.2014

Pagan, M., Simoncelli, E. P., & Rust, N. C. (2016). Neural Quadratic Discriminant Analysis: Nonlinear Decoding with V1-Like Computation. *Neural Comput*, 1-29. doi: 10.1162/NECO_a_00890

Pagan, M., Urban, L. S., Wohl, M. P., & Rust, N. C. (2013). Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. *Nat Neurosci, 16*(8), 1132-1139. doi: 10.1038/nn.3433

Raposo, D., Kaufman, M. T., & Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nat Neurosci, 17*(12), 1784-1792. doi: 10.1038/nn.3865

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron, 26*(3), 703-714.

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature, 497*(7451), 585-590. doi: 10.1038/nature12160

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

Romo, R., Brody, C. D., Hernandez, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature, 399*(6735), 470-473.

Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J Neurosci, 30*(39), 12978-12995. doi: 30/39/12978 [pii]

10.1523/JNEUROSCI.0179-10.2010

Schiller, P. H. (1995). Effect of lesion in visual cortical area V4 on the recognition of transformed objects. *Nature, 376*, 342-344.

Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain, 120 (Pt 3)*, 535-559.

Sugase-Miyamoto, Y., Liu, Z., Wiener, M. C., Optican, L. M., & Richmond, B. J. (2008). Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Comput Biol, 4*(5), e1000073. doi: 10.1371/journal.pcbi.1000073

Sundberg, K. A., Mitchell, J. F., Gawne, T. J., & Reynolds, J. H. (2012). Attention influences single unit and local field potential response latencies in visual cortical area V4. *J Neurosci, 32*(45), 16040-16050. doi: 10.1523/JNEUROSCI.0489-12.2012

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval [In Process Citation]. *Nature, 401*(6754), 699-703.

Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399*, 575-579.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-585). Cambridge, MA: M.I.T. Press.

Yaginuma, S., Niihara, T., & Iwai, E. (1982). Further evidence on elevated discrimination limens for reduced patterns in monkeys with inferotemporal lesions. *Neuropsychologia, 20*(1), 21-32.

Zenon, A., & Krauzlis, R. J. (2012). Attention deficits without cortical neuronal deficits. *Nature, 489*(7416), 434-437. doi: 10.1038/nature11497

Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci, 27*(45), 12292-12307.