UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

The effects of cognitive state on visual encoding properties in human cortex

Permalink

https://escholarship.org/uc/item/9rq1d71w

Author

Vo, Vy Ai

Publication Date

2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

The effects of cognitive state on visual encoding properties in human cortex

A dissertation submitted in partial satisfaction
of the requirements for the Doctor of Philosophy

in

Neurosciences with a Specialization in Computational Neurosciences

by

Vy Ai Vo

Committee in charge:

Professor John T. Serences, Chair
Professor Edward M. Callaway
Professor Timothy Q. Gentner
Professor Douglas A. Nitz
Professor Edward Vul

2019

Copyright

Vy Ai Vo, 2019

All Rights Reserved

The Dissertation of Vy Ai Vo is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2019

DEDICATION

I dedicate this work to the many people who supported and inspired me, whether they knew it or not. This includes my parents and siblings; my dearest friends; and of course, my partner Brandon, who is a constant source of both support and intellectual inspiration. But it also includes my favorite authors and philosophers; professors new and old; and those who have lived their lives dedicated to exploration and understanding.

EPIGRAPH

The tiny shock of seeing your assumptions undermined, and your gut-level conviction overturned, provides a high that I call an UnAha! experience.

There is an odd exhilaration to be found in showing your “gut” instinct that it is wrong. The gut is driven towards conviction. It provides to decisions a sense of moral certitude; “rightness” lives in the gut. *Counter-examples keep the capacity for doubt alive*. . . .re-examine where your certitude is coming from, and get deeper insights as a result.

-- Venkatesh Rao, 2008. ribbonfarm: experiments in refactored perception.

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Epigraph v

Table of Contents vi

List of Abbreviations vii

List of Figures viii

List of Tables ix

Acknowledgments x

Vita xii

Abstract of the Dissertation xiii

Introduction 1

Chapter 1 Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex 6

Chapter 2 Altering the spatial scope of covert attention induces differential patterns of voxel receptive field modulations 23

Chapter 3 Long-term and short-term memory representations in retinotopic visual cortex: cortical reinstatement and source information 77

Conclusion 110

LIST OF ABBREVIATIONS

fMRI: functional magnetic resonance imaging

BOLD: blood-oxygenation-level dependent

RF: receptive field

vRF: voxel receptive field

IEM: inverted encoding model

FDR: false discovery rate

V1: primary visual cortex, or striate cortex

IPS: intraparietal sulcus

WM: working memory

LTM: long-term memory

LIST OF FIGURES

Figure 1.1: Covert spatial attention task and hypothesized representation changes with shifts of spatial attention. 9

Figure 1.2: Changes in voxel receptive fields (vRFs) across attention conditions. 13

Figure 1.3: Spatial discriminability increases with attention and is mediated by position changes in vRFs 15

Figure 1.4: Multivariate inverted encoding model (IEM) used to reconstruct the mapping probe stimuli. 16

Figure 1.5: Reconstruction parameters as a function of mapping stimulus distance from the covertly attended locations and attention hemifield (attended vs. ignored). 17

Figure 1.6: A layered spatial encoding model reveals how different sets of vRF changes lead to enhancements in multivariate stimulus reconstructions. 18

Figure 2.1: Covert spatial attention task 28

Figure 2.2: vRF changes between the focused and divided attention conditions, averaged across all participants 34

Figure 2.3: Estimated mean parameter values for each covert attention condition, taken from the mixed effects model fits 37

Figure 2.4: A comparison of vRF properties when attending at central fixation point and covertly attending to 1 or 3 peripheral target 42

Figure 2.5: Spatial discriminability metric 45

Figure 2.6: The relative importance of each vRF parameter in each covert attention condition in determining spatial discriminability near the attended target 47

Figure 2.S1: Psychophysical curves fit to all the scanner data 67

Figure 2.S2: Estimated means for the effect of condition and ROI on two metrics of vRF quality 69

Figure 3.1: Memory tasks and mapping task presented to participants in the scanner 83

Figure 3.2: Memory recall across both scanner tasks is similar 94

Figure 3.3: Reconstructions of remembered spatial positions over time 97

Figure 3.4: Decoding between memory tasks is successful only in parietal regions (bottom row) in the late delay period 100

LIST OF TABLES

Table 1.1: vRF selection statistics, pooled across participants 10

Table 1.2: Mean coefficients for polynomial fits of how vRF parameter change is modulated by distance from the attended location..... 14

Table 1.3: Two-way ANOVA results for reconstruction parameter changes 17

Table 1.4: RMSE (and 95% CIs) between reconstructions from the reduced dataset (only using voxels with RFs) or from different versions of the layered IEM using the same voxels 19

Table 2.1: Statistical tests of vRF size, amplitude, and baseline for the two covert attention conditions: divided (attend 3) and focused (attend 1) 38

Table 2.S1: Number of voxels which survive each threshold before inclusion in the vRF analyses..... 68

Table 2.S2: Statistical tests of vRF size, amplitude, and baseline against the fixation task 70

ACKNOWLEDGEMENTS

I am grateful to have worked with many brilliant, kind, and hard-working people who deserve mention here. Foremost being my mentor John Serences, whose clear insight—which always seems so simple and straightforward in retrospect—and late-night email humor has carried me through many challenges. His combination of rigorous scientific perspective and empathetic support was an integral part of my learning, and of my experience in graduate school. Of course, I must thank all of the members of the Serences Lab who joined me for this journey in discovery: their discursive musings on cognition, perception, and neural computation both inspired and improved this work. In particular, the enthusiasm of two students before me, Tommy Sprague and Sirawaj Itthipuripat, often kept me going, and I am glad to have worked with them on several research projects. But the community and positive environment in the lab and at UCSD enlivened me, too, and often provided some much-needed stress-relief.

Funding from the National Science Foundation Graduate Research Fellowship program gave me the freedom to pursue some of these questions.

The chapter in this dissertation could not have been written without my co-authors, who collected data, generated ideas, and provided critical feedback, among other contributions.

Chapter 1 is a full reprint of the material as it appears in: Vo, Vy A., Sprague, Thomas C., Serences, John T. (2017). Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex. *Journal of Neuroscience 37*(12), 3386-3401. The dissertation author was the primary investigator and author of this paper.

Chapter 2 is being prepared for submission. The author list and working title is: Vo, Vy A., Serences, John T. Altering the spatial scope of covert attention induces differential patterns of voxel receptive field modulations. The dissertation author was the primary investigator and author of this manuscript.

Chapter 3 is being prepared for submission with colleagues either currently at or formerly at the University of Chicago and the University of California, San Diego. The author list

and working title is: Vo, Vy A., Sutterer, David W., Foster, Joshua J., Sprague, Thomas C., Awh, Edward, Serences, John T. Long-term and short-term memory representations in retinotopic visual cortex: cortical reinstatement and source information. The dissertation author was the primary investigator and author of this paper.

VITA

2011	Bachelor of Arts, Swarthmore College
2011 - 2013	Research Assistant, University of Rochester
2017	Master of Science, University of California, San Diego
2019	Doctor of Philosophy, University of California, San Diego
2019	Cognitive Neuroscientist / Deep Learning R&D Engineer, Intel Corporation

PUBLICATIONS

Itthipuripat, S.*, Vo, V.A.*, Sprague, T.C., Serences, J.T. (2019) Value-driven attentional capture enhances distractor representations in early visual cortex. *bioRxiv*.

Sprague, T.C.*, Adam, K.C.S.*, Foster, J.J.*, Rahmati, M.*, Sutterer, D.W.*, Vo, V.A**.*** (2018). Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. *eNeuro 5*(3).

Sprague, T.C., Itthipuripat, S., Vo, V.A., and Serences, J.T. (2018). Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex. *Journal of Neurophysiology 119*(6), 2153-2165.

Vo, V.A., Sprague, T.C., and Serences, J.T. (2017). Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex. *Journal of Neuroscience 37*(12), 3386-3401.

Vo, V.A**.**, Li, R., Kornell, N., Pouget, A., Cantlon, J.F. (2014). Young children bet on their numerical skills: Metacognition in the numerical domain. *Psychological Science 25*(9), 1712-1721.

ABSTRACT OF THE DISSERTATION

The effects of cognitive state on visual encoding properties in human cortex

by

Vy Ai Vo

Doctor of Philosophy in Neurosciences with a Specialization in Computational Neurosciences

University of California San Diego, 2019

Professor John T. Serences, Chair

Understanding the format of neural representations is a key problem in neuroscience. I present a series of functional magnetic resonance imaging (fMRI) studies of the human visual system to investigate the format of how visual information is represented, and how the format changes due to different cognitive demands. First, I investigate how covert spatial attention modifies visual encoding properties at two different levels of analysis—at the level of a single spatial encoding unit, the receptive field, and at the level of the encoding population. Second, I examine how the contents of visual memory are represented across both working memory and long-term memory, systems that have traditionally been considered distinct entities. Together these data reveal how different retinotopic regions of visual cortex flexibly encode information about the visual world and how they are altered in response to varying cognitive demands.

INTRODUCTION

A modern computational theory of mind posits that all mental processes (e.g., memory recall, spatial navigation, mental imagery) are simply computations performed on neural representations. For example, recalling a vivid visual memory likely depends on several representations of the contents of memory, instantiated in patterns of neural activity across several areas of the brain. There is likely activity in the hippocampus that distinguishes the recalled memory from other similar memories in long-term storage (Chavlis & Poirazi, 2017; Norman & O'Reilly, 2003). There is also likely activity in visual cortex corresponding to the specific visual features and details of the memory (Favila, Samide, Sweigart, & Kuhl, 2018; Kuhl & Chun, 2014). Understanding the format of neural representations is a key problem in the study of the human mind and brain. This work contains a set of studies of the human visual system to investigate how two cognitive processes—attention and memory—rely on specific formats of visual information, and how these processes may change the format of a representation to serve cognitive demands.

The visual system is an appealing topic of study here for several reasons: decades of prior study have described how visual input through the retina is transformed into representations of separate visual features in neocortex, and more recently, how these representations are further transformed into representations of complex objects such as faces (Chang & Tsao, 2017). This body of knowledge allows us to formulate specific *algorithmic* descriptions of how visual inputs are encoded and transformed (Marr, 1982). Humans also rely heavily on the visual system to perform many tasks, and have dedicated a large portion of cortex to it which can be easily measured with noninvasive techniques such as functional magnetic resonance imaging (fMRI). Finally, a long tradition of studying visual attention and memory in cognitive psychology has provided a taxonomy of behavior to help constrain the specific tasks used here. In particular, many have theorized that attention is a key component to both episodic memory retrieval and working memory maintenance (Awh, Vogel, & Oh, 2006;

Cabeza, Ciaramelli, Olson, & Moscovitch, 2008). While the specific relationship between attention and memory is not directly under investigation here, these theories do suggest that understanding how each one alters the format of neural representations may provide some insight into this topic.

In Chapters 1 and 2, I use fMRI to investigate how covert spatial attention alters the representation of relevant visual information. Using fMRI, rather than electrophysiology, has the advantage of being able to measure all the encoding units which form a representation of the entire visual field. This enables me to examine attentional modulations at different scales, or levels of analysis. The attention studies require 3 sets of analyses to link these levels of analysis. First, I examine how covert spatial attention modulates the response properties of the spatial receptive fields (RFs) of single voxels. Spatial receptive fields are considered a basic unit of the visual system, encoding information about retinotopic space from early visual areas (V1) to higher-level parietal areas (IPS2), which are thought to handle complex cognitive tasks. A single voxel contains thousands of spatially-selective, retinotopically organized neurons. Despite the huge difference in scale, both voxel RFs in humans and single-unit RFs in macaques have been shown to have similar response properties (Dumoulin & Wandell, 2008; Harvey & Dumoulin, 2011) and even similar modulations with covert spatial attention (Klein, Harvey, & Dumoulin, 2014). Second, I use the entire population of voxel RFs to assess how covert attention alters population-level representations of visual space. Here I combine information across all the measured voxel RFs to form a representation of the attended location. I find that, as expected, attention increases spatial information at the attended location. Third and finally, I can determine which modulations at the level of a single voxel contribute to these increases in spatial information. Together, these data bridge a gap between levels of analysis in the study of visual attention.

In Chapter 3, I investigate a few questions about the representational format of vivid visual memories, either maintained from a recent percept (short-term or working memory, WM)

or retrieved from a long-term store (long-term memory, LTM). Separate studies of each memory system have shown that both WM and LTM recall rely on a reactivation, or reinstatement, of neural activity in sensory regions involved in processing the original percept. However, it is unknown whether reinstated visual memories are represented in a similar format across retinotopic cortex, or whether some brain regions maintain distinct information about the type of memory. I used an inverted encoding model (Brouwer & Heeger, 2009; Ester, Sprague, & Serences, 2015; Sprague, Ester, & Serences, 2014, 2016) to examine the contents of visual memories and assess the similarity of their representations across a WM task and a well-equated LTM task. I then assessed whether these representations contained unique information about the type of memory with a linear decoder. The results generate several interesting hypotheses about the relationship between WM, LTM, and attention.

These three studies provide a crucial test of the flexibility of visual encoding across different regions of retinotopic cortex, and provide insight into how common encoding properties support distinct cognitive tasks.

References

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and memory. *Neuroscience*, *139*, 201–208. https://doi.org/https://doi.org/10.1016/j.neuroscience.2005.08.023

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of Neuroscience*, *29*(44), 13992–4003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*(8), 613–625. https://doi.org/10.1038/nrn2459

Chang, L., & Tsao, D. Y. (2017). The Code for Facial Identity in the Primate Brain. *Cell*, *169*(6), 1013–1020.e14. https://doi.org/10.1016/j.cell.2017.05.011

Chavlis, S., & Poirazi, P. (2017). Pattern separation in the hippocampus through the eyes of computational modeling. *Synapse*, *71*. https://doi.org/10.1002/syn.21972

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal representations of stimulus features are amplified during memory retrieval and flexibly aligned with top-down goals. *Journal of Neuroscience*. https://doi.org/10.1101/280974

Harvey, B. M., & Dumoulin, S. O. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *The Journal of Neuroscience*, *31*(38), 13604–12. https://doi.org/10.1523/JNEUROSCI.2572-11.2011

Klein, B. P., Harvey, B. M., & Dumoulin, S. O. (2014). Attraction of Position Preference by Spatial Attention throughout Human Visual Cortex. *Neuron*, 1–11. https://doi.org/10.1016/j.neuron.2014.08.047

Kuhl, B. A., & Chun, M. M. (2014). Successful Remembering Elicits Event-Specific Activity Patterns in Lateral Parietal Cortex. *Journal of Neuroscience*, *34*(23), 8051–8060. https://doi.org/10.1523/JNEUROSCI.4328-13.2014

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. New York, NY: W.H. Freeman and Company.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Sprague, T. C., Ester, E. F., & Serences, J. T. (2014). Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology*, *24*(18), 2174–2180. https://doi.org/10.1016/j.cub.2014.07.066

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron*, *91*(3), 694–707. https://doi.org/10.1016/j.neuron.2016.07.006

CHAPTER 1: Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex

Behavioral/Cognitive

Spatial Tuning Shifts Increase the Discriminability and Fidelity of Population Codes in Visual Cortex

Vy A. Vo,[1] Thomas C. Sprague,[1,2] and John T. Serences[1,3,4]

[1]Neurosciences Graduate Program, University of California, San Diego, La Jolla, California 92093, [2]Department of Psychology and Center for Neural Science, New York University, New York, New York 10003, [3]Department of Psychology, University of California, San Diego, La Jolla, California 92093, and [4]Kavli Institute for Brain and Mind, University of California, San Diego, La Jolla, California 92093

Selective visual attention enables organisms to enhance the representation of behaviorally relevant stimuli by altering the encoding properties of single receptive fields (RFs). Yet we know little about how the attentional modulations of single RFs contribute to the encoding of an entire visual scene. Addressing this issue requires (1) measuring a group of RFs that tile a continuous portion of visual space, (2) constructing a population-level measurement of spatial representations based on these RFs, and (3) linking how different types of RF attentional modulations change the population-level representation. To accomplish these aims, we used fMRI to characterize the responses of thousands of voxels in retinotopically organized human cortex. First, we found that the response modulations of voxel RFs (vRFs) depend on the spatial relationship between the RF center and the visual location of the attended target. Second, we used two analyses to assess the spatial encoding quality of a population of voxels. We found that attention increased fine spatial discriminability and representational fidelity near the attended target. Third, we linked these findings by manipulating the observed vRF attentional modulations and recomputing our measures of the fidelity of population codes. Surprisingly, we discovered that attentional enhancements of population-level representations largely depend on position shifts of vRFs, rather than changes in size or gain. Our data suggest that position shifts of single RFs are a principal mechanism by which attention enhances population-level representations in visual cortex.

Key words: attention; encoding models; labeled lines; population codes; receptive fields; vision

Significance Statement

Although changes in the gain and size of RFs have dominated our view of how attention modulates visual information codes, such hypotheses have largely relied on the extrapolation of single-cell responses to population responses. Here we use fMRI to relate changes in single voxel receptive fields (vRFs) to changes in population-level representations. We find that vRF position shifts contribute more to population-level enhancements of visual information than changes in vRF size or gain. This finding suggests that position shifts are a principal mechanism by which spatial attention enhances population codes for relevant visual information. This poses challenges for labeled line theories of information processing, suggesting that downstream regions likely rely on distributed inputs rather than single neuron-to-neuron mappings.

Introduction

Spatial receptive fields (RFs) are a core component of visual information processing throughout the visual system. They are modified by selective visual attention to improve the fidelity of sensory representations, likely enabling more precise, accurate perception (Desimone and Duncan, 1995; Anton-Erxleben and Carrasco, 2013). Prior studies in nonhuman primates have found that covert spatial attention changes the position, size, and amplitude of responses in single-cell RFs in early cortical areas, such as V1, V4, and MT (Moran and Desimone, 1985; Connor et al., 1996, 1997; Womelsdorf et al., 2006, 2008; Roberts et al., 2007; David et al., 2008). Recent neuroimaging studies have also shown that single-voxel RFs (vRFs) undergo similar response changes with attention, shifting toward the attended target or changing in size (Klein et al., 2014; de Haas et al., 2014; Kay et al., 2015; Sheremata and Silver, 2015). Most accounts suggest that these RF modulations improve the spatial representations of the attended target, either by boosting the signal-to-noise ratio by increasing response amplitude, or by increasing the spatial resolution by decreas-

Received Nov. 9, 2016; revised Feb. 13, 2017; accepted Feb. 16, 2017.

Author contributions: V.A.V., T.C.S., and J.T.S. designed research; V.A.V. and T.C.S. performed research; V.A.V., T.C.S., and J.T.S. contributed unpublished reagents/analytic tools; V.A.V. and T.C.S. analyzed data; V.A.V. wrote the paper.

This work was supported by National Science Foundation Graduate Research Fellowships to V.A.V. and T.C.S., National Eye Institute Grant R01-EY025872, and James S. McDonnell Foundation Scholar Award to J.T.S. We thank the lab, Rosanne Rademaker, and Edward Vul for comments on analyses and on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to either Vy A. Vo or John T. Serences, Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA 92093-0634. E-mail: vyaivo@ucsd.edu or jserences@ucsd.edu.

DOI:10.1523/JNEUROSCI.3484-16.2017

Copyright © 2017 the authors 0270-6474/17/373386-16$15.00/0

ing RF size (Desimone and Duncan, 1995; Anton-Erxleben and Carrasco, 2013; Cohen and Maunsell, 2014). These mechanisms are akin to turning up the volume (gain increase) or to using smaller pixels to encode a digital image (size decrease).

Despite these documented modulations, it is not yet clear how different types of RF modulations are combined to facilitate robust population codes. Recent studies have only begun to explore how interactions between neurons may affect the coding properties of the population (Anton-Erxleben and Carrasco, 2013; Cohen and Maunsell, 2014). Yet analyzing these data at a population level is crucial for understanding how spatial attention changes the overall representation of an attended area. Prior fMRI studies that measured many voxel RFs (vRFs) across space were often unable to report the full pattern of response modulations with respect to the attended target because subjects attended to the mapping stimulus, rather than to a fixed point in space (Sprague and Serences, 2013; Kay et al., 2015; Sheremata and Silver, 2015). Studies that fixed the locus of attention have reported mixed results on vRF modulations (Klein et al., 2014; de Haas et al., 2014). The first aim of this study was thus to evaluate how properties of vRFs in retinotopic areas change with attention, especially near the peripheral attention target.

The second aim of the study was to evaluate how different types of RF modulations contribute to population-level enhancements of an attended region of space. Single RFs in early visual areas are fundamentally local encoding models that are relatively uninformative about regions outside their immediate borders. To study their relationship to a population-level representation of space, other metrics are needed to integrate information across all local encoding units (e.g., vRFs) to evaluate how attentional modulations impact the quality of population codes. Here, we used two different population-level metrics of spatial encoding fidelity to investigate these questions, and to determine how changes in vRF amplitude, size, or position affect the population-level representations. First, we used a measure related to Fisher information to evaluate the spatial discriminability of population codes. Second, we used a spatial encoding model that incorporates information across voxels to form representations of stimuli in the mapped visual field (Brouwer and Heeger, 2009; Sprague and Serences, 2013; Sprague et al., 2015).

We found that vRF position shifts increase both the spatial discriminability around the attended region as well as the fidelity of stimulus reconstructions near the attended target. Surprisingly, shifts in vRF position accounted for more of the population-level enhancements with attention than changes in vRF size or gain. This finding is unexpected in the context of "labeled-line" models of information processing, which posit that visual representations rely on RFs that transmit consistent "labels" for visual features, such as spatial position. Our findings suggest that apparent shifts in the labels of RFs play an important role in the attentional enhancement of visual information.

Materials and Methods

Task design and participants

We collected data from 9 human participants (4 female), 6 of whom had previously completed a set of retinotopic mapping scans in the laboratory (Participants AA, AB, AC, AI, and AL in Sprague and Serences, 2013; Participants AA, AC, and AI in Sprague et al., 2014; all participants in Ester et al., 2015). All participants provided written informed consent and were compensated for their time ($20/h) as approved by the local University of California, San Diego Institutional Review Board. Participants practiced both the attention task and the localizer task before entering the scanner. A minimum of 4 h of scanning was required to complete the entire analysis, so one participant was excluded due to insufficient data (they only completed 2 h). Another participant was excluded for inconsistent behavioral performance, with average task accuracy at chance (48.6%). This yielded a total of 7 participants who completed the entire experiment (3 2 h scan sessions per participant).

Participants centrally fixated a gray rectangular screen (120 × 90 cm) viewed via a head-coil mounted mirror (~3.85 m viewing distance). They attended one of three fixed locations on the screen: the fixation point or a target to the lower left or lower right of fixation. During each 2000 ms trial, subjects reported a change in the attention target. When subjects attended fixation, they reported whether a brief contrast change (100–400 ms, starting 300–1000 ms into the trial) was dimmer or brighter than the baseline contrast. The peripheral attention targets were two pentagons (0.17° radius; 50% contrast) centered 2.1° to the left and right of fixation (Fig. 1*a*). When subjects attended a peripheral target, they reported whether it rotated clockwise or counterclockwise (rotation duration 100–300 ms, starting 300–1600 ms into the trial). Intertrial intervals randomly varied between 1000 and 3000 ms in 500 ms increments (mean intertrial interval: 2000 ms). The magnitude of the contrast change or the rotation was adjusted on each run to keep task performance for each participant near 75% (mean = 75.90%, bootstrapped 95% CI [72.46%, 79.20%]), with no significant difference between conditions as evaluated with a one-way repeated-measures ANOVA randomization test ($F_{(1,11)} = 0.220$, randomized $p = 0.800$). For 4 participants, we collected 6 runs on the attend periphery tasks without a change in the luminance of the fixation stimulus. Performance on the attend periphery tasks was stable across runs with and without the luminance change (repeated-measures ANOVA with run type × random participants factor; $p = 0.439$, null F distribution using randomized labels for 10,000 iterations). Therefore, these data were collapsed across scan sessions with and without changes in fixation luminance.

On 51 of the 61 trials in each run, a full-contrast 6 Hz flickering checkerboard (0.68° radius; 1.67 cycles/deg) appeared for 2000 ms at 1 of 51 different locations across the screen to map the spatial sensitivity of visually responsive voxels. These mapping stimuli covered a region of the screen approximately subtending 9° horizontal and 6° vertical when their position was jittered. When one of these checkerboards overlapped with any of the static attention targets, they were partially masked with a small circular aperture the same color as the screen background (0.16°/0.25° radius aperture for fixation/pentagon, respectively) that allowed the stimulus to remain visible. Participants were instructed to ignore the task-irrelevant flickering checkerboards throughout the experiment. During the 10 null trials on each scan, the participant continued to perform the attention task but no checkerboard was presented. Null trials and mapping stimulus trials were presented in a pseudorandom interleaved order.

The location of the checkerboard mapping stimulus on each trial was determined by generating an evenly spaced triangular grid (0.84° between grid points) and centering the checkerboard on one of these grid points. The location of the checkerboard mapping stimulus was then jittered a random amount from these grid points (±0.42°/0.37° horizontal/vertical). When subjects attended the peripheral target, half of the runs were presented at the discrete grid positions so that we could achieve more stable stimulus reconstructions (see Stimulus reconstructions using an inverted spatial encoding model).

MRI

We obtained all structural and functional MR images using a GE 3T MR750 scanner at the University of California, San Diego. We collected all functional images (19.2 cm × 19.2 cm FOV, 64 × 64 acquisition matrix, 35 interleaved slices, 3 mm^3 voxels with 0 mm slice gap, 128 volumes per scan run) using a gradient echo planar pulse sequence (2000 ms TR, 30 ms TE, 90° flip angle) and a 32-channel head coil (Nova Medical). Five dummy scans preceded each functional run. A high-resolution structural image was acquired at the end of each session using a FSPGR T1-weighted pulse sequence (25.6 cm × 25.6 cm FOV, 256 × 192 acquisition matrix, 8.136/3.172 ms TR/TE, 192 slices, 9° flip angle, 1 mm^3 voxels). All functional scans were coregistered to the anatomical images acquired during the same session, and this anatomical was in turn coregistered to the anatomical acquired during the retinotopy scan.

EPI images were unwarped with a custom script from UCSD's Center for Functional Magnetic Resonance Imaging using FSL and AFNI. All

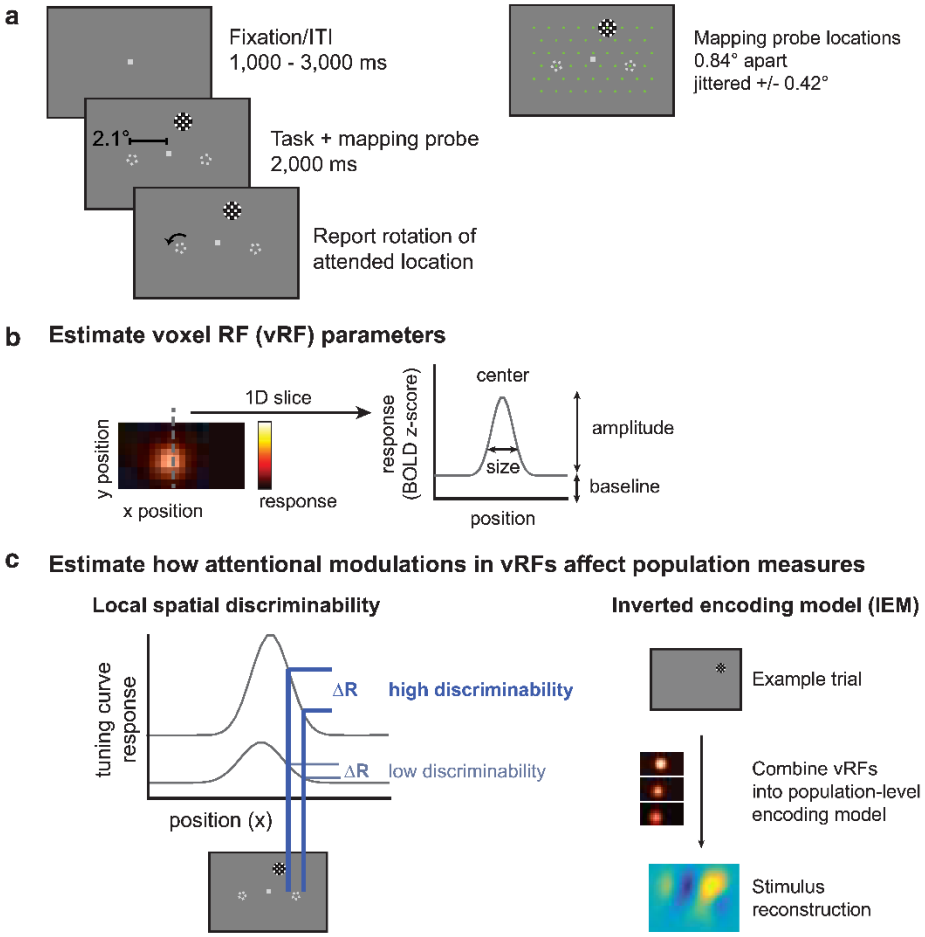


Figure 1. Covert spatial attention task and fMRI-based analyses used to link single voxels to population-level measurements. ***a***, Subjects fixated centrally and attended to brief rotations in the pentagon stimulus on the left or right while a flickering checkerboard probe stimulus appeared at 1 of 51 grid locations across the visual field. On control runs, subjects attended to a contrast change at fixation. fMRI data measured during this attention task are used to create visualizable estimates of vRFs and stimulus reconstructions. ***b***, A receptive field model is fit to the responses of each voxel and can be described by its *x* and *y* position (center), response baseline, response amplitude, and size (FWHM). ***c***, Given a population of voxels in a retinotopic region, such as V1, we examine two different measures of spatial information in the population. The first, a spatial discriminability metric, scales with the slope of the tuning curve at a given location in space (see Materials and Methods). The second relies on a multivariate IEM for space. By reconstructing images of the mapping stimulus on each test trial, we can measure how population-level spatial information changes with attention. We then can model how changes in individual vRFs affect both of these population measures.

subsequent preprocessing was performed in BrainVoyager 2.6.1, including slice-time correction, six-parameter affine motion correction, and temporal high-pass filtering to remove slow signal drifts over the course of each run. Data were then transformed into Talairach space and resampled to have a $3 \times 3 \times 3$ mm voxel size. Finally, the BOLD signal in each voxel was transformed into *z* scores on a scan-by-scan basis. All subsequent analyses were performed in MATLAB using custom scripts (available online on Open Science Framework: osf.io/s9vqv).

Independent localizer task

We constrained our analyses to visually responsive voxels in occipital and parietal cortex using a separate localizer task (3–5 runs per participant). On 14 trials, participants fixated centrally and viewed a full-field flickering checkerboard (10 Hz, 11.0/8.3° width/height) for 8000 ms. Participants detected whether a small area (2D Gaussian, $\sigma = 0.2°$) within the checkerboard dimmed in contrast. Contrast dimming occurred between 500 and 4000 ms after the start of the trial and lasted between 2000 and 3000 ms (all uniformly sampled in 500 ms steps). This contrast change occurred infrequently (randomly on 5 of 14 trials) at a random location within the checkerboard. The average contrast change was varied between runs to maintain consistent performance at ~75% accuracy (mean performance 78.0%). On 8 trials, participants simply fixated throughout the trial without a checkerboard being presented. We then used a standard GLM with a canonical two-gamma HRF (peak at 5 s, undershoot peak at 15 s, response undershoot ratio 6, response dispersion 1, undershoot dispersion 1) to estimate the response to the checkerboard stimulus in each voxel. For all subsequent analyses, only voxels in the retinotopically defined areas V1, V2, V3, V4, V3A/B, and IPS0 with a significantly positive BOLD response to the localizer task (at FDR $q = 0.05$) were included (Benjamini and Yekutieli, 2001).

Estimating single-trial BOLD responses

For all subsequent analyses, we used trial-wise BOLD *z* scores. We estimated these by creating a boxcar model marking the duration of each

Table 1. vRF selection statistics, pooled across participants ($N = 7$)

ROI	Total no. of localized voxels	No. of voxels after GLM thresholding	No. of voxels after regularizability threshold	No. of voxels after cross-validation threshold	No. of voxels after removing difference score outliers	Percentage that survive all thresholds	RMSE fit error for surviving voxels
V1	3723	3540	2438	989	931	25.01	0.1105
V2	4154	3970	3115	1405	1339	32.23	0.1087
V3	3698	3519	2839	1520	1435	38.81	0.0994
V4	1702	1492	1118	361	336	19.74	0.0783
V3A/B	1988	1922	1440	443	416	20.93	0.0893
IPS0	1567	1492	800	114	110	7.02	0.0882
V1-V4	13,277	12,521	9510	4275	4041	30.44	0.1032
V3A/B and IPS0	3555	3414	2240	557	526	14.80	0.0894
Total	16,832	15,935	11,750	4832	4567	27.13	0.1016

checkerboard mapping stimulus and convolving it with a canonical two-gamma HRF (peak at 5 s, undershoot peak at 15 s, response undershoot ratio 6, response dispersion 1, undershoot dispersion 1). To standardize our data across runs, we z scored the BOLD responses within each run and concatenated the z scores across runs. We then solved a GLM to find the response to each predictor.

Statistical procedures

All reported CIs are computed by resampling the data with replacement (i.e., bootstrapping). The number of iterations for each bootstrapping procedure varied (depending on available computing power and time for that procedure) and are therefore reported with each result. For tests comparing a bootstrapped distribution against zero, p values were computed by conducting two one-tailed tests against 0 (e.g., mean(param_change < 0) and mean(param_change > 0)) and doubling the smaller p value. All repeated tests were FDR corrected ($q = 0.05$).

vRF estimation, fitting, and parameter analysis

We first estimated vRFs for each attention condition to investigate (1) how vRF parameters changed when participants attended to different locations and (2) the spatial pattern vRF changes across visual space. We note here that prior reports have referred to similar vRF models as population receptive fields to emphasize the fact that each voxel contains a population of spatially tuned neurons (Dumoulin and Wandell, 2008; Wandell and Winawer, 2015). However, because we are comparing modulations at different scales in the present study (i.e., modulations in single voxels and in patterns of responses across many voxels), we will refer to these single voxel measurements as vRFs, and will reserve the term "population" exclusively for multivariate measures involving several voxels, allowing our terminology to align with theories of population coding (Ma et al., 2006).

We estimated vRFs using a modified version of a previously described technique (Sprague and Serences, 2013). This method estimates a single voxel's spatial sensitivity by modeling its BOLD responses as a linear combination of discrete, smooth spatial filters tiled evenly across the mapped portion of the visual field. These spatial filters (or spatial channels) form our modeled basis set. We then regressed the BOLD z scores (v voxels $\times$ n trials) onto a design matrix with predicted channel responses for each trial (C, k channels $\times$ n trials) by solving Equation 1 as follows:

$$B = WC \tag{1}$$

for the matrix W (v voxels $\times$ k channels).

Each of the k channels in the basis set was defined as a 2D cosine that was fixed to reach 0 at a set distance from the filter center as follows:

$$f(r) = \left(0.5\left(\cos\left(\frac{r\pi}{s}\right) + 0.5\right)\right)^7 \quad \text{for } r < s \tag{2}$$

where r is the distance from the filter center and s is the size constant. Setting a zero baseline in this function ensured that we could estimate a stable baseline for each voxel by restricting the response of the channel to a known subregion of the visual display. Because the estimated vRF size depends on the size of the filters, we made the filters fairly small (1.08° FWHM) and dense (91 filters arranged in a 13 horizontal/7 vertical grid, each spaced 0.83° apart). We then discretized the filters by sampling them in a high-resolution 2D grid of 135 $\times$ 101 pixels spanning 10° $\times$ 5°. The discretized filters (k filters $\times$ p pixels) were multiplied with a mask of the checkerboard stimulus on every trial (p pixels $\times$ n trials) so that the design matrix C contained predictions of the spatial channel responses on every trial of the mapping task.

To fit our estimated vRFs with a unimodal function, we used ridge regression to solve Equation 1. This is a common regularization method that sparsifies the regression solution by penalizing the regressors with many small weights (Hoerl and Kennard, 1970; Lee et al., 2013). This meant solving for an estimate of W by the following:

$$\hat{W}^T = (CC^T + \lambda I)^{-1}CB^T \tag{3}$$

where λ is the ridge parameter penalty term and I is a $k \times k$ identity matrix. We estimated an optimal λ for each voxel by evaluating Equation 3 over a range of λ values (0–750) for a balanced number of runs of the attention task (e.g., an equal number of runs from each attention condition). We then computed the Bayesian Information Criterion for each of these λ values, estimating the degrees of freedom in the ridge regression as follows:

$$df = \text{trace}\,(C(C^TC + \lambda I)^{-1}C^T) \tag{4}$$

The λ with the smallest Bayesian Information Criterion was selected for each voxel. Because the attention comparisons are done within voxels, the varying λ penalty across voxels could not explain the attention effects we observed.

To select reliable voxels for analysis, we next implemented a set of conservative thresholding steps (Table 1). We first needed to select voxels with reliable visual responses, so we only kept voxels with trial β weights that predicted at least 50% of the BOLD time courses in each scan session. Second, we only used voxels that could be successfully regularized with ridge regression. Any voxels with the maximum λ (750) were discarded, as this indicated that the ridge regression solution had not converged. Finally, we verified that the resulting regression model could predict an independent dataset, so we performed exhaustive leave-one-run-out cross validation for each attention condition. This ensured that the λ estimated across attention conditions produced reliable data for each condition separately. We estimated W using data from all but one run (Eq. 3) and used this to predict the BOLD GLM trial estimate of the left-out run (Eq. 2), separately for each condition. We then computed the mean correlation between the predicted and real BOLD GLM trial estimates across cross-validation folds for each voxel. It is not possible to calculate a coefficient of determination on regularized data because the process of ridge regression changes the scale of the predicted data (Huth et al., 2012). We only kept voxels where this cross-validation $r > 0.25$ for all 3 conditions.

To quantify each vRF, we fit the spatial RF profile of each voxel with a smooth 2D function with four parameters: center, size, baseline, and amplitude (Fig. 1*b*; Eq. 2). Here, we define the vRF baseline as the voxel's response that does not reliably depend on the position of the mapping stimulus (i.e., its constant offset). The vRF amplitude is defined as the spatially selective increase in a voxel's response above this baseline. To-

gether, these two parameters index how much of the voxel's response is due to a change in mapping stimulus position. Finally, the size and location parameters estimate the spatial selectivity and the spatial position preference of the vRFs, respectively. We first downsampled the vRFs by multiplying the estimated weights $\hat{W}$ for each voxel (a $1 \times k$ channel vector) by a smaller version of the spatial grid that contained the basis set (68×51 pixel grid; $10° \times 5°$). This speeded up the process of fitting the pixelwise surface with Equation 2. This fitting process began with a coarse grid search that first found the best fit in a discrete grid of possible vRF parameters (center sampled in 1° steps over the mapped portion of the visual field; size constant logarithmically sampled at 20 points between FWHM of $10^{0.01}$ degrees and 10^{1} degrees). At each grid point, we estimated the best fit amplitude and baseline using linear regression. The grid point fit with the smallest root mean square error (RMSE) provided the initialization seed to a continuous error function optimization algorithm (fmincon in MATLAB). This fit had several constraints: the final solution must place the center within 2 grid points of the seeded fit (parameterized by position and size) and within the mapped visual field; the amplitude must be between 0 and 5; the baseline must be between (5 and 5 BOLD z score units. Occasionally, this nonlinear fitting algorithm did not converge and resulted in a larger error than the original seed. In this case, we took the best fit grid point as the final fit.

To test whether vRF fit parameters changed when participants focused spatial attention at different positions, we compared fits during each attend periphery condition with fits during the attend fixation condition. We computed a difference score (attend peripheral(attend fixation) to describe the magnitude of the attentional modulation. For example, a difference score of(2° in the FWHM of the vRF would indicate that the response profile width decreased when the participant attended to the peripheral target. This analysis revealed a subset of voxels with very large difference scores, which we determined to be due to noisy data or poor fits via manual inspection. Accordingly, we performed a final thresholding step for all vRF-based analyses: an elimination of outlier voxels with difference scores >3 times the SD of the population mean, where the population consists of the parameter difference scores for a given region of interest (ROI) (Table 1). After removing these outliers, we tested whether the vRF parameter difference scores differed significantly from 0 within a visual ROI by bootstrapping the distribution of difference scores across participants 10,000 times.

To determine whether these vRF changes were modulated by their position in the visual field, we first calculated each vRF's distance from the attended location (*v_dist_attn*) using its position during the fixation task. We then fit an nth order polynomial to the vRF difference scores as a function of *v_dist_attn*, where $n = 0$, 1, or 2. This corresponds to a constant offset (0th order), a linear fit (first order), or a quadratic or parabolic fit (second order). These fits were cross-validated by fitting on 50% of the vRF difference scores and calculating goodness-of-fit (residual sum of squares and R^2) on each of the 10,000 cross-validation iterations. These cross-validation iterations also provided CIs on the coefficients for each polynomial. The most parsimonious fit was chosen by performing a nested F test on the average residual sum of squares for each polynomial model.

We also tested whether vRF attentional modulations depended on hemisphere or visual hemifield, akin to the results reported for IPS0-IPS5 in Sheremata and Silver (2015). We sorted the voxels in each attention condition as contralateral or ipsilateral to the attended target. We then performed a series of nonparametric bootstrapped tests similar to a two-way ANOVA with attended hemifield and voxel hemisphere as factors. The vRFs were resampled with replacement across subjects 10,000 times. We then evaluated the two main effects and the interaction by computing a difference in the means of the groups or a difference in the slope between the group means, respectively. None of the tests for the effect of hemisphere and the interaction survived FDR correction, so we do not report those results here. We speculate that this null result is likely due to a lack of reliable voxels in anterior parietal cortex areas IPS1–5 in our study, where previous reports have found larger laterality effects (Sheremata and Silver, 2015).

Population analysis

Fine spatial discriminability metric. To compute the spatial discriminability of a population of vRFs, we estimated the spatial derivative of each vRF at every point in the mapped visual field in 0.1° steps (Fig. 1*c*). This was done by taking the slope of the vRF along the x and y direction at each pixel in the image of the visual field and squaring this value (Scolari and Serences, 2009, 2010). This measurement is a descriptor of how well a population code can discriminate small changes in the spatial arrangement of the stimulus array, which depends on the rising and falling edges of a tuning curve rather than the difference between the peak response and a baseline response (Regan and Beverley, 1985; Pouget et al., 2003; Butts and Goldman, 2006; Navalpakkam and Itti, 2007; Scolari and Serences, 2009, 2010). To restrict our measurements to the relevant area near the peripheral target, we computed discriminability values within 1 degree of the center of each target across both spatial dimensions (x and y). These were summed and divided by the maximum discriminability value in that population to make the results comparable despite changes in vRF coverage or responsiveness.

Stimulus reconstructions using an inverted spatial encoding model. In addition to computing the discriminability metric described above, we also reconstructed an image of the entire visual field on each trial using a population-level encoding model. Compared with the local spatial discriminability index, this is a more sensitive method of assessing the amount of spatial information encoded in an entire population of voxels because it exploits the pattern of response differences across voxels, rather than treating each voxel as an independent encoding unit (Serences and Saproo, 2012; Sprague et al., 2015).

We trained the spatial encoding model using a procedure similar to the vRF estimation analysis described above (see Fig. 4*a*). This yields an estimated matrix of weights, $\hat{W}_2$, which specifies how much each voxel in a region of interest responds to each of the spatial channels (Brouwer and Heeger, 2009; Serences and Saproo, 2012; Sprague and Serences, 2013; Sprague et al., 2015). We then solved Equation 1 using the Moore-Penrose pseudoinverse with no regularization as follows:

$$\hat{W}_2 = BC^T(CC^T)^{-1} \tag{5}$$

C was constructed using a set of 54 evenly tiled spatial filters (Eq. 2; 9 horizontal/6 vertical; spaced 1.25° apart; 1.56° FWHM). $\hat{W}_2$ was estimated using the data from the jittered position runs. This was done separately for each participant, using a training set balanced across the conditions of interest (e.g., an equal number of attend left and attend right runs and all attend fixation runs because fixation is the neutral condition).

To reconstruct a representation of the mapped visual space, we inverted the model and multiplied the pseudoinverse of the estimated weight matrix $\hat{W}_2$ with a test dataset from the discrete position runs (B_2), yielding estimated channel activations for each trial (C_2; k_2 channels by t test trials) (Eq. 6). Thus, we refer to this analysis as the inverted encoding model (IEM) as follows:

$$\hat{C}_2 = (\hat{W}_2^T \hat{W}_2)^{-1}\hat{W}_2^T B_2 \tag{6}$$

Because of mathematical constraints on inverting W_2 (number of voxels must be greater than number of channels), we included all voxels in each ROI instead of just the subset of well-fit voxels used in the vRF analyses described above. We computed Equation 6 twice using different test datasets, once for the discrete position attend left runs and once for the discrete position attend right runs.

When we multiply the resulting channel activations by a grid of pixels that define the spatial channels, we obtain a spatial representation of the entire visual field on each trial. This image contains a stimulus reconstruction showing where the checkerboard should have been given the trained model and the activation pattern across all voxels in the independent test set. The stimulus reconstructions were then fit in the same manner as the vRFs, using Equation 1 to estimate the center, size, amplitude, and baseline (mean fit RMSE across all ROI reconstructions 0.114; 95% CI [0.102, 0.312]). Here, the baseline is an estimate of the multivariate reconstruction that is spatially nonselective (i.e., not significantly modulated by the position of the mapping stimulus). The amplitude

describes the maximal increase in that reconstruction relative to baseline when the mapping stimulus is on the screen.

To assess how attention changed reconstructions of the mapping stimulus across the visual field, we first computed a difference score that described the effect of attention by folding the visual field in half (i.e., collapsing across hemifield) and comparing parameters in the attended vs ignored hemifield. We excluded the reconstructions that fell along the vertical meridian (3 of 51 stimulus positions). This allowed us to control for the overall effect of eccentricity while remaining sensitive to other spatial patterns in stimulus reconstruction modulations.

We then set up a single factor repeated-measures omnibus ANOVA to determine which pairs of ROI and parameter (e.g., V1 size, V1 amplitude) were affected by either attention or Euclidean distance from the target stimuli. The attention factor had two levels (attend/ignore) and the distance factor had 6 levels (6 evenly spaced distance bins from 0° to 2.54°). Based on the results of this omnibus test, we tested any significant ROI-parameter combination in a two-way repeated-measures ANOVA of attention × distance. To estimate the p values for these tests, we generated empirical null distributions of the F scores by randomizing the labels within each factor 10,000 times within each participant. Reported p values are the percentage of the randomized F scores that are greater than or equal to the real F scores.

Layered spatial encoding model to link vRFs to multivariate stimulus reconstructions. To test how changes in the response properties of the underlying vRFs contributed to changes in the fidelity of region-level stimulus reconstructions, we generated simulated patterns of voxel activity on every trial by predicting the response to each stimulus based on the vRF fit parameters. We then used this simulated data to estimate and invert a population-level spatial encoding model, as described above (see Fig. 6*a*).

For these simulations, we could only use well-fit voxels to generate simulated BOLD time series. This constrained the analysis to ROIs with at least as many vRFs as spatial filters used to estimate the spatial encoding model. To ensure that we could include most participants in the layered encoding model analysis, we created two large ROIs by merging the smaller retinotopically defined regions described above. The occipital ROI consisted of V1, V2, V3, and V4 defined for each subject. The posterior parietal ROI consisted of V3A/B and IPS0. The vRFs in the parietal ROI show distinct patterns of attentional modulations (see Fig. 2*e*), suggesting that V3A/B and IPS0 are both anatomically and functionally distinct from the occipital regions (see also de Haas et al., 2014). Although merging ROIs increased the number of voxels available for the encoding model analysis, we still did not have enough voxels in the parietal ROI to estimate the layered encoding model for 3 of the 7 participants (A1, AR, AU). However, the remaining data from 4 participants were sufficient to produce stable subject-averaged results.

To simulate each voxel's BOLD response on every trial that the participant completed in the real experiment, we first created a high-resolution set of spatial channels (21 × 11 channels spaced 0.5° apart, FWHM = 0.65°) and generated weights for each channel based on the vRF fit obtained from prior analysis. That is, we evaluated Equation 2 for each channel at the vRF's fit center and adjusted the response gain by multiplying this result by the fit amplitude and adding the fit baseline. We then added independent Gaussian noise to each of these channel weights, simulating a small amount of variance in the voxel's response ($\sigma = 0.5$). Each voxel's channel weights were then multiplied by the idealized channel response on each trial (i.e., the channel filter convolved with the stimulus mask), effectively simulating the BOLD response on each trial for the entire population of voxels based on their measured vRFs. We added Gaussian noise to this simulated response as well ($\sigma = 0.5$). We then computed stimulus reconstructions using the same method as described above (the IEM in Stimulus reconstructions using an inverted spatial encoding model), averaging resulting reconstructions across participants and like positions before fitting.

To ensure the stability of the reconstructions that were based on simulated data, we repeated the simulations 100 times and averaged across the fits of all iterations to generate the plots in Figure 6*b*. Then, to compare how well the layered model reproduced the attentional modulations observed in stimulus reconstructions generated with real data, we calculated an error metric between the layered IEM and the real data. We first calculated reconstruction difference scores across attention condition (attended − ignored; see Stimulus reconstructions using an inverted spatial encoding model). This yielded 24 difference scores each for both attention conditions in both the layered IEM data and the empirical data. Because the empirical data did not have any repeated iterations, we averaged across all 100 iterations of the layered model to match the dimensionality of the real reconstructions (2 conditions × 24 difference scores × 4 parameters). We could then calculate the RMSE between the difference scores from the full empirical dataset (i.e., the data shown in Fig. 5) and the modeled data. This was used as a metric to describe the goodness-of-fit of each layered IEM.

We then tested how different vRF attentional modulations contributed to changes in the population-level stimulus reconstructions. To test how shifts in vRF centers contributed to population-level information, we modeled voxels that had the same fit center across both attention conditions, simulated their BOLD responses on each trial, and generated stimulus reconstructions from these data. The voxel's vRF center was defined as the vRF center fit from the neutral attend fixation data. A similar procedure was repeated for all reported combinations of parameter changes across conditions. Again, whichever parameter was held constant took its value from the neutral attend fixation condition.

To calculate the confidence intervals on the RMSE changes in Figure 6*c*, we resampled with replacement across the 100 model iterations and refit the average across these 100 instances. This resampling procedure was repeated 500 times to generate a distribution of fits to the model data. We then took the difference between the RMSE of the null model, in which no parameters varied between attention conditions, and the RMSE of the model which held some number of vRF parameters constant across attention conditions.

Results

Modulations of vRF properties with spatial attention

We estimated single vRFs for each voxel in 6 retinotopically identified visual areas from V1 to IPS0. The estimation of vRFs was done independently for each attention condition so that we could compare a single voxel's spatial tuning across conditions.

To confirm that the fit sizes were consistent with previous results, we fit a line to the estimated sizes as a function of the vRF center eccentricity. First, we combined all vRFs across participants and conditions in each ROI. We then binned the vRF centers every 0.25° from fixation and calculated the mean size (Fig. 2*b*). We first replicated an increase in vRF size with increasing eccentricity, and an increase in the slope of this relationship across visual regions (Gattass et al., 2005; Dumoulin and Wandell, 2008; Amano et al., 2009; Harvey and Dumoulin, 2011) (Fig. 2*b*). These observations confirm that our method produced reasonable vRF estimates that were consistent with previous reports.

Next, we examined how covert attention to the peripheral attention targets modulated vRF properties, relative to the attend fixation condition. Overall, the center position of vRFs shifted significantly closer to the attended location ($p < 0.005$ in all ROIs, Fig. 2*d*). This finding is consistent with previous reports in humans and in monkeys for both covert attention tasks and saccade tasks (Womelsdorf et al., 2006, 2008; Klein et al., 2014; Zirnsak et al., 2014).

Although we did observe changes in the size of individual vRFs, the mean change was not significantly different from zero ($p > 0.05$ in all ROIs). Size increases have been previously reported in tasks that required subjects to attend to the mapping stimulus, which moved on each trial (Sprague and Serences, 2013; Kay et al., 2015; Sheremata and Silver, 2015). Accordingly, if attention causes the center of RFs to shift toward the attended target, these combined shifts in position would average out to form a larger RF estimate. In contrast, mapping vRFs while maintaining a fixed locus of attention would nullify the size increase, consistent with the results we ob-

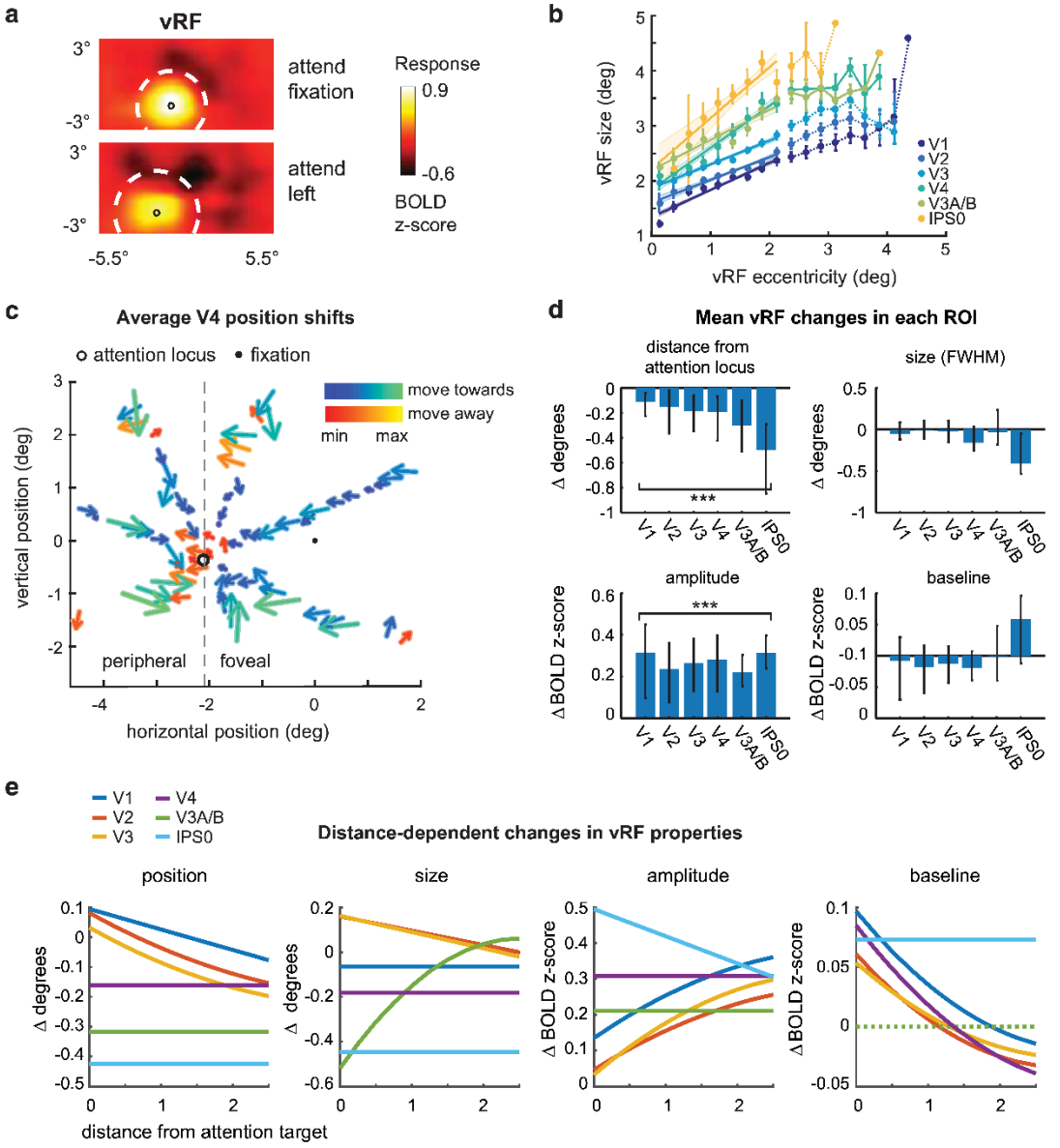


Figure 2. Changes in vRFs across attention conditions. We separately estimated vRFs for every voxel in visual and posterior parietal areas, discarding poorly estimated or noisy voxels (Table 1). Unless otherwise specified, figure data are averaged across subjects. Error bars indicate 95% CIs computed by resampling the data distribution. ***a***, An example vRF shows that attending covertly to the left location shifts the center of the receptive field profile to the left compared with the neutral attend fixation condition. Voxel is from Subject AR in area V3A/B. ***b***, Our vRF estimates reproduced the canonical size-eccentricity relationship (positive slope in all ROIs, $p <$ minimum possible p value, 1/10,000 iterations) and the increase in slope between visual regions. ***c***, Preferred position changes of V4 vRFs with covert spatial attention. We binned each vRF by its position during the attend fixation condition. The origin of each arrow is the center of each position bin. The end of the arrow shows the average position shift of the vRFs within that position bin during the attend peripheral conditions (left/right are collapsed and shown as attend left). The majority of vRFs shift toward the attended location (blue-green color map vs red-yellow). ***d***, Mean changes in vRF parameters (attend peripheral target − attend fixation) in each visual area. ***e***, Attentional modulations of each vRF parameter plotted by the vRF's distance from the attention target computed from its position during the attend fixation task (Table 2).

served (Fig. 2*d*). Another study which also found increases in vRF size with attention required subjects to attend the fixation point while they manipulated the perceptual load, or difficulty, of the attention task (de Haas et al., 2014). In our study, we intentionally kept task performance constant and could not evaluate effects of difficulty on the parameters of vRFs.

We also found an overall increase in vRF amplitude with attention ($p < 0.001$ for all tests). Because these measures were

Table 2. Mean coefficients for polynomial fits of how vRF parameter change is modulated by distance from the attended location (*v_dist_attn*)[a]

	Position	Size	Amplitude	Baseline
V1	−0.069, 0.095	−0.064	−0.019, 0.137, 0.135	0.011, −0.073, 0.010
V2	0.015, −0.133, 0.082	−0.064, 0.160	−0.016, 0.125, 0.046	0.011, −0.066, 0.061
V3	0.017, −0.135, 0.033	−0.073, 0.163	−0.025, 0.170, 0.032	0.009, −0.053, 0.054
V4	−0.162	−0.181	0.308	0.011, −0.078, 0.085
V3A/B	−0.318	−0.091, 0.461, −0.520	0.210	<0.001
IPS0	−0.425	−0.445	−0.076, 0.495	0.073

[a]Number of reported coefficients in the table correspond to the polynomial order, which has yielded the most parsimonious fit to the data (e.g., 1 coefficient for $n = 0$, 2 coefficients for $n = 1$, etc.).

calculated relative to a fixation task, these data suggest that covert spatial attention to a peripheral location caused widespread position and gain modulations in all vRFs across the visual field.

It is unclear whether these attentional modulations are limited to areas near the attended target, or whether they are uniform across the visual field. For example, vRF position shifts could result in a radial convergence of RFs toward the attended target, or a uniform shift of RFs along a vector extending from fixation to the attention or saccade target (Tolias et al., 2001; Klein et al., 2014; Zirnsak et al., 2014). Furthermore, reports of other RF properties (such as size) modulating with attention have been mixed (Connor et al., 1996, 1997; Womelsdorf et al., 2008; Niebergall et al., 2011; Sprague and Serences, 2013; Klein et al., 2014; de Haas et al., 2014; Kay et al., 2015; Sheremata and Silver, 2015). We therefore examined whether each of the vRF parameter changes was dependent on the vRF's location in the visual field, relative to the attended location. First, we created radial distance bins centered on the left or right attended locations, and sorted voxels into these bins based on their preferred position during the fixation condition. After this sorting procedure, data from the right condition were flipped and collapsed with the left condition.

When we plotted vRF position changes in each bin, we found that spatial attention caused vRF position shifts that converged on the attended location (two-tailed sign test on vector direction, $p < 0.001$ in all ROIs). That is, vRFs shifted closer to the attended location (Fig. 2*c*), compared with when subjects attended fixation (mean shift across all vRFs and ROIs: −0.239°, 95% CI [−0.566, −0.048], Fig. 2*d*). Small eye movements toward the attended location cannot explain receptive field convergence: this would cause all vRFs to shift in the same horizontal direction, rather than radially converging on one point. These data are consistent with results from both humans (Klein et al., 2014) and macaques (Connor et al., 1996, 1997; Womelsdorf et al., 2006, 2008) that use a similar task. However, the prior study in humans focused only on vRFs with preferred locations that were foveal to the attended location, and the studies in macaques only report RF position changes in V4 and MT. By contrast, our data show that vRF centers converge on the attended location across all visual areas, including primary visual cortex, and that this pattern of modulations includes vRFs peripheral to the attended target.

These plots (Fig. 2*a*,*d*) also suggested that vRFs farther from the attended location underwent larger position changes than vRFs near the attended location. That is, the magnitude of the attentional modulation may be dependent on the distance between the vRF and the attended target. To test for this, we fit a polynomial to the vRF parameter changes as a function of distance from the attended location (see Materials and Methods). We selected the most parsimonious fit ranging from a mean change in vRF parameter (0th order polynomial) to a parabolic change (second-order polynomial) by conducting a nested *F* test (Table 2). The best polynomial fits are plotted in Figure 2*e*.

This analysis allowed us to characterize trends in vRF attentional modulations across space. It also implicitly tests whether voxels contralateral to the attended target respond differently than ipsilateral voxels. This is because vRFs near the attended target will mostly originate from the contralateral side of visual cortex. Therefore, any fit lines with a significant slope imply there is a difference between contralateral and ipsilateral voxels (Sheremata and Silver, 2015). A separate test described in Materials and Methods confirmed that contralateral voxels differed significantly from ipsilateral voxels in the areas where we saw the highest fit slopes in Figure 2*e* (FDR-corrected $p < 0.05$ for position: V1, V2; size: V3; amplitude: V3, IPS0). However, because the fit lines illustrate how these changes occur over space, we discuss those data here instead (Fig. 2*e*).

In early visual areas V1–V3, vRFs near the attention target were slightly repelled from the target, whereas vRFs farther from the target were attracted toward the target. In later visual areas, vRFs were uniformly attracted toward the attention target. We saw a similar pattern of results with size modulations: early visual areas showed an increase in vRF size near the attention target, and decreased size farther away. However, in areas V3A/B and IPS0, vRF size decreased near the attention target.

The pattern of vRF amplitude modulations was also segregated between the early and later visual areas. All vRFs increased in amplitude with attention, but the slope of this relationship inverted from early to later visual areas. In V1–V3, the slope if positive, such that voxels ~2° away from the attention target increase in amplitude more than voxels right at the target position. The amplitude increase is constant in V4 and V3A/B. Finally, in IPS0, the slope inverts to become negative, so that voxels near the attention target increase in amplitude more than voxels farther away. Last, we found an increase in vRF baseline near the attended target in V1-V4, but a uniform increase in baseline in IPS0. Overall, we found that the type and magnitude of the attentional modulation in different visual areas changes as a function of the spatial relationship between vRFs and the attended target. This is consistent with findings from macaque neurophysiology, which had suggested that amplitude and size changes depend on where the RF is located in relation to the attended target (Connor et al., 1996; Niebergall et al., 2011).

These fits only describe the modulations with respect to the voxel's position during the attend fixation task. However, these parameter changes likely interact with one another, such that a voxel that shifts toward the attended location will also increase in amplitude. Hence, to determine how the joint patterns of vRF modulations change the spatial information content of a representation, in the next section we discuss two different population-level measures that combine data across the population of vRFs in each ROI.

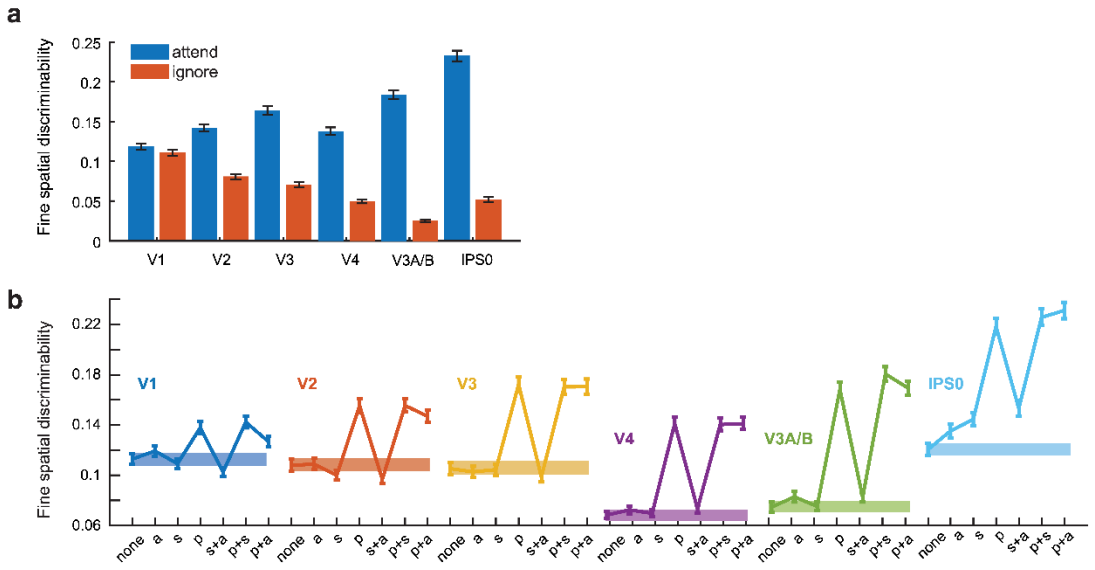


Figure 3. Spatial discriminability increases with attention and is mediated by position changes in vRFs. Error bars indicate bootstrapped 95% CIs. ***a***, We formulated a measurement to describe the ability of a population of voxels to make fine spatial discriminations around the attention target. We used the properties of each voxel's spatial tuning curve to make this measurement (see Materials and Methods). Spatial discriminability increased when subjects attended the target, compared with when they ignored the target in the opposite hemifield (resampled $p <$ minimum possible p value 1/1000 for all ROIs). ***b***, The discriminability metric was recomputed for vRFs with a variety of attentional modulations. Abscissa labels from left to right: None, vRF parameters during the neural attend fixation condition; a, amplitude; s, size; p, position. Spatial discriminability increased significantly when we applied position changes measured during the attend L/R task to the vRFs compared with when we applied no parameter changes (solid bar). By contrast, applying size changes did not change spatial discriminability in most ROIs, although it did cause a small increase in IPS0.

Increases in spatial discriminability depend primarily on vRF position shifts

Next, we assessed how different types of RF modulations influenced the precision of population-level codes for spatial position. We first computed a discriminability metric that described the ability of a population of tuning curves (here, vRFs) to support fine spatial judgments (see Materials and Methods). When we computed this metric based on the measured vRF properties from each condition, spatial discriminability near the attended target increased relative to the ignored target in the opposite visual hemifield (Fig. 3*a*).

We then tested how different types of vRF modulations (such as size changes or position shifts) affected this spatial discriminability metric. As a baseline comparison, we first computed discriminability based on vRFs estimated during the attend fixation runs for each subject. We then added different sets of observed attentional modulations to the population before recomputing spatial discriminability. For example, we shifted all the vRF centers to match the measurements when a subject was attending to the left target and computed discriminability near the attended target. Because the response baseline of a vRF does not affect the discriminability metric, we excluded this type of attentional modulation from these analyses.

Across all ROIs, we found that vRF position shifts played the biggest role in increasing fine spatial discriminability compared with changes in size or changes in amplitude (Fig. 3*b*). Position modulations alone led to a large increase in spatial discriminability, whereas other combinations of parameter modulations only had an impact if we added in position shifts (i.e., a change in size and position increased discriminability, but size alone did not). The only departure from these patterns was observed in IPS0, where all attentional modulation types increased spatial discriminability, but position changes increased spatial discriminability the most.

Spatial attention increases the fidelity of population-level stimulus reconstructions

By design, the spatial discriminability metric we computed is only informative about local spatial representations and cannot assess how different patterns of vRF modulations might result in representational changes across the visual field. To address this point, we built a multivariate spatial encoding model to measure how attention changes the representations of visual information in disparate parts of space. This also allowed us to further test the effects of vRF modulations on the encoding properties of the population, including response baseline changes that were not captured by our discriminability metric.

The spatial IEM reconstructed an image of the entire visual field on each test trial. We first trained the model using the responses of each voxel on a set of training trials with known mapping stimulus positions. We then created image reconstructions on independent test trials by inverting the model and multiplying it by the voxel responses during each test trial (Fig. 4*a*; see Materials and Methods). Each image contained a representation of where the mapping stimulus should have been given the pattern of voxel activations on that particular trial. The IEM successfully reconstructed the task-irrelevant mapping stimuli using activation patterns across voxels in each visual area from V1 through IPS0 (Fig. 4*b*; grand mean error between fit and actual position 2.40°, 95% CI 0.55°, 4.97°).

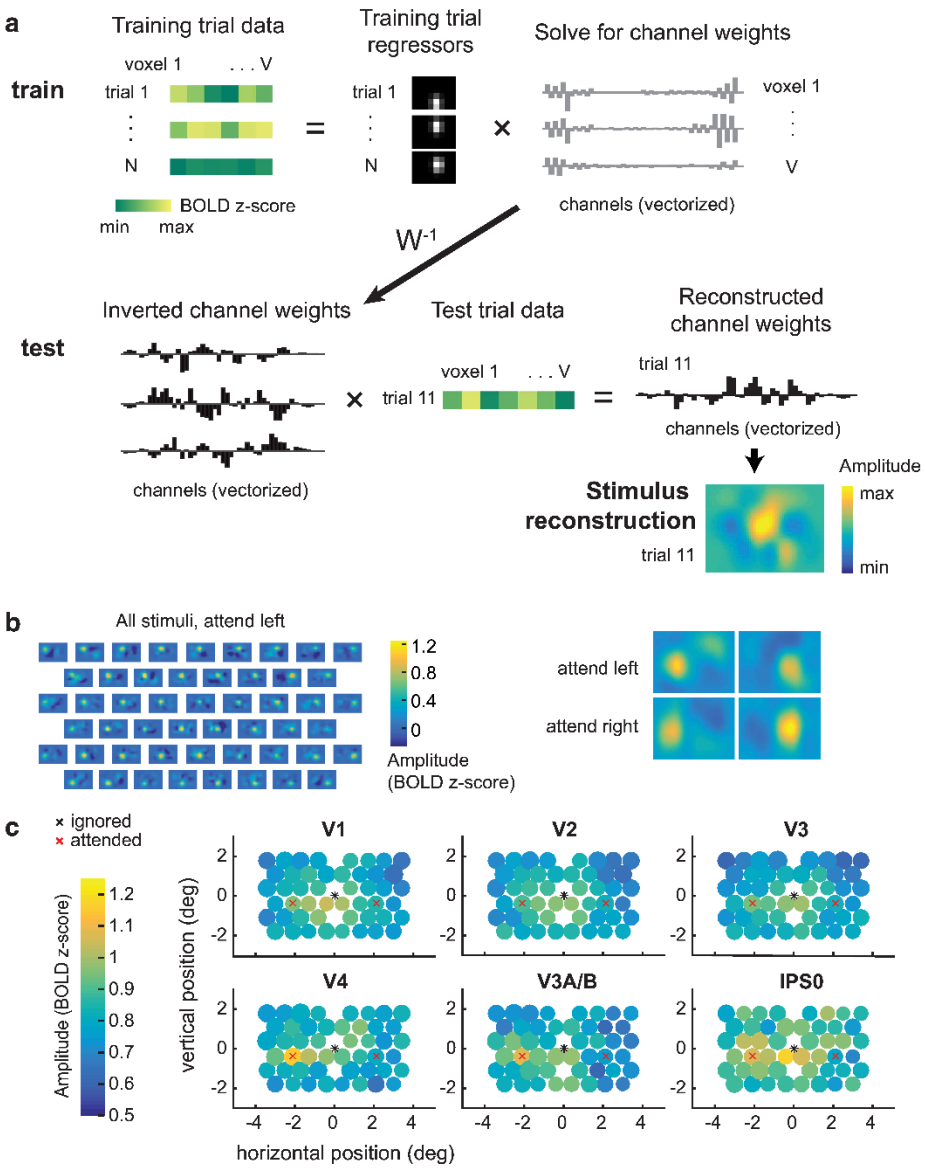


Figure 4. Multivariate IEM used to reconstruct the mapping probe stimuli. ***a***, To train the IEM, we first take the BOLD data from all voxels within a visual region from a subset of training trials. Then, we solve for a set of channel weights using least-squares regression. To reconstruct the stimulus, we invert this weight matrix and multiply it with BOLD data from the same voxels during a test trial. This yields a reconstructed channel response profile, which can be transformed into a reconstruction of the mapping stimulus on every trial in each attention condition. Data shown are examples from Participant AR for a subset of V1 voxels. ***b***, Example stimulus reconstructions for Participant AI in area V1. These reconstructions were averaged across trials with the same position, yielding 51 reconstructions: one for each unique position in the test dataset. Left, The same averaged position reconstructions are shown for each condition. The amplitude on the left is higher when attending left, and on the right when attending right. ***c***, Average reconstruction sizes and amplitudes for each stimulus position (collapsed across condition; left, attended). The diameter of the circle depicts the average fit FWHM of the reconstructions at that spatial position. Reconstruction amplitude was greater in the attended hemifield compared with the ignored hemifield in areas V3A/B and V4 ($p < 0.005$; Table 3; Fig. 5).

We used these stimulus reconstructions as a proxy for the quality of the spatial representations encoded in a population of voxels. This is line with previous studies showing that stimulus reconstructions change in amplitude or size as a function of cognitive demands. (Brouwer and Heeger, 2013; Ester et al., 2013; Sprague and Serences, 2013; Sprague et al., 2014, 2015, 2016).

First, we compared how reconstructed representations of each mapping stimulus changed as subjects shifted their spatial attention. We ran a repeated-measures ANOVA of attention × dis-

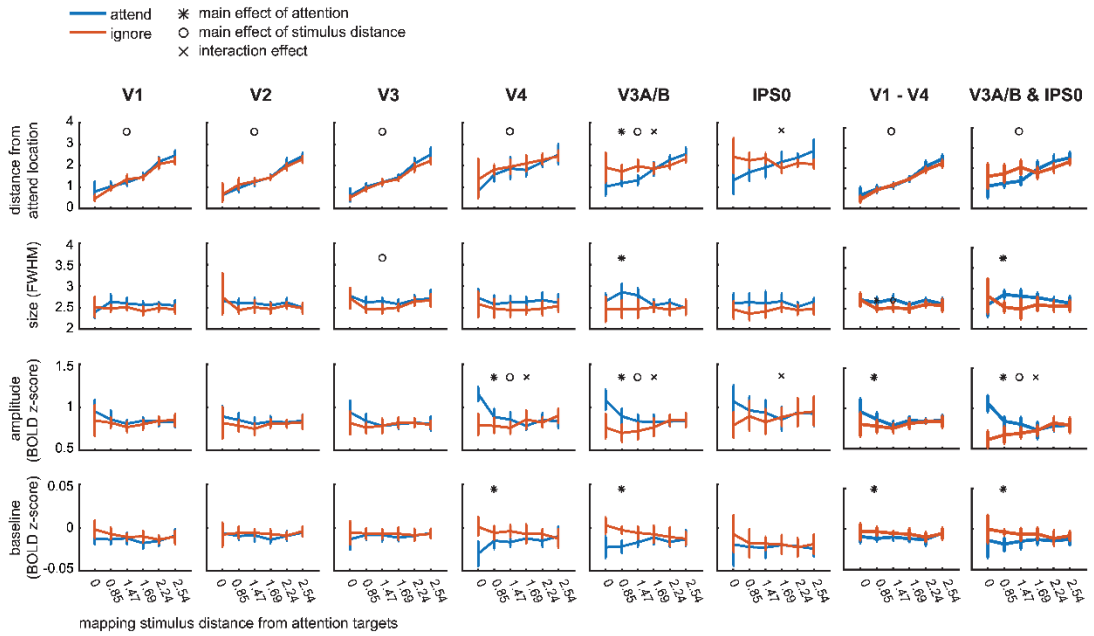


Figure 5. Reconstruction parameters as a function of mapping stimulus distance from the covertly attended locations and attention hemifield (attended vs ignored). For complete list of *p* values, see Table 3.

Table 3. Two-way ANOVA results for reconstruction parameter changes (*s_dist_attn* × attention hemifield)

	V1	V2	V3	V4	V3A/B	IPS0	V1-V4	V3A/B and IPS0
Omnibus test								
Position	<0.001[a]	<0.001[a]	<0.001[a]	<0.001[a]	<0.001[a]	0.001[a]	<0.001[a]	<0.001[a]
Size	0.216	0.565	0.019	0.428	0.006[a]	0.121	0.001[a]	0.110
Amplitude	0.174	0.579	0.024[a]	<0.001[a]	<0.001[a]	0.008[a]	0.016[a]	<0.001[a]
Baseline	0.088	0.734	0.934	<0.001[a]	0.001[a]	0.937	0.015[a]	0.241
Main effect of distance								
Position	<0.001[a]	<0.001[a]	<0.001[a]	<0.001[a]	<0.001[a]	0.192	<0.001[a]	<0.001[a]
Size	—	—	—	—	0.484	—	0.019[a]	—
Amplitude	—	—	0.140	0.002[a]	0.005[a]	0.478	0.100	0.002[a]
Baseline	—	—	—	0.829	0.916	—	0.210	—
Main effect of attention								
Position	0.371	0.916	0.346	0.067	0.005[a]	0.254	0.401	0.343
Size	—	—	—	—	0.003[a]	—	0.005[a]	—
Amplitude	—	—	0.069	<0.001[a]	0.005[a]	0.158	0.049[a]	0.004
Baseline	—	—	—	0.001[a]	<0.001[a]	—	0.004[a]	—
Interaction of distance and attention								
Position	0.052	0.588	0.541	0.657	<0.001[a]	<0.001[a]	0.121	0.026
Size	—	—	—	—	0.077	—	0.271	—
Amplitude	—	—	0.064	<0.001[a]	<0.001[a]	0.004[a]	0.224	<0.001[a]
Baseline	—	—	—	0.019	0.011	—	0.370	—

[a]*p* value passed FDR correction ($q = 0.05$, corrected across ROIs and comparisons within each parameter).

tance bin for each reconstruction fit parameter (see Materials and Methods). Here, a main effect of attention would suggest that stimulus reconstructions in the attended hemifield changed in a consistent way compared with the ignored hemifield. A main effect of distance would suggest that stimulus reconstruction changes had a consistent spatial pattern across both the attended and ignored hemifields. This would occur when a stimulus' representation was altered with distance from the attention target. For example, the stimulus reconstruction center should vary linearly with the stimulus' true distance from the attention target. And last, an interaction effect would suggest that the distance effect was dependent on whether the reconstruction belonged to the attended or ignored hemifield. In our task, the reconstructed stimuli are always irrelevant to the task of the observer. We therefore predicted an interaction effect where spatial attention would selectively modulate stimulus reconstructions within the hemifield of the attended location, but not the opposite hemifield (Connor et al., 1996, 1997).

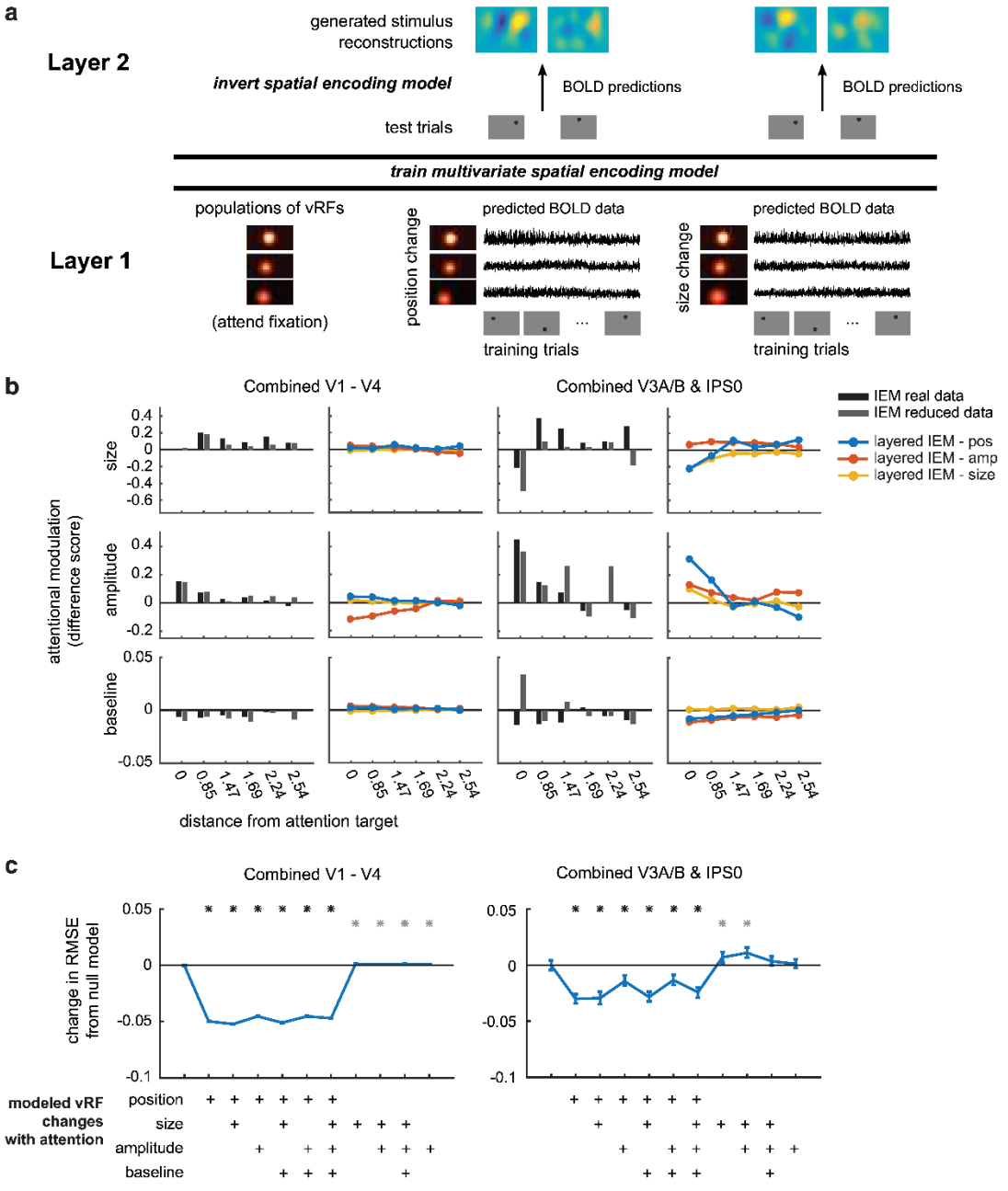


Figure 6. A layered spatial encoding model reveals how different sets of vRF changes lead to enhancements in multivariate stimulus reconstructions. ***a***, The first layer of the model uses the vRF fits to generate BOLD data from every subject's real trial sequence. Then the BOLD data from all voxels within one ROI are used to train a multivariate spatial encoding model and reconstruct the mapping stimuli. ***b***, Change in reconstruction amplitude in the attended versus the ignored hemifield. We only show reconstruction parameters with significant attentional modulations in the prior IEM analysis (Fig. 5; Table 3). Stimulus reconstructions computed with a reduced number of voxels (gray bar) largely reproduce the pattern of attentional modulations observed in IEMs computed with all voxels (black bar). Furthermore, a comparison of layered IEMs using simulated data revealed that vRF position changes (blue lines) in the first layer of the model are better at reproducing the amplitude modulations in the stimulus reconstructions of the parietal ROI than models that simulate changes in vRF size or amplitude (yellow and red lines). ***c***, RMSE between each set of IEM fits and the full empirical dataset fits shown in Figure 5. The null baseline model (far left) is a layered IEM where the vRF parameters are the same across all attention conditions. We then added vRF attentional modulations for each parameter as shown in the matrix below the plot, where all models with position changes are on the left side. $*p < 0.05$ (FDR-corrected), models differed significantly from the null baseline model. Gray asterisks indicate an increase in RMSE from the null model. Black asterisks indicate a decrease.

Table 4. RMSE (and 95% CIs) between reconstructions from the reduced dataset (only using voxels with RFs) or from different versions of the layered IEM using the same voxels[a]

	Reduced data	p/s/a/b	p/a/b	p/s/b	s/a/b	p/a
Combined occipital V1–V4	0.133 (0.109, 0.170)	0.146 (0.146, 0.146)	0.148 (0.147, 0.148)	0.142 (0.141, 0.143)	0.194 (0.194, 0.194)	0.148 (0.148, 0.148)
Combined parietal V3A/B and IPS0	0.834 (0.744, 0.959)	0.415 (0.410, 0.419)	0.426 (0.422, 0.431)	0.411 (0.407, 0.416)	0.443 (0.439, 0.447)	0.425 (0.421, 0.430)

[a]To generate CIs, the resampling of the real data is performed at the level of the fits to the reconstructions, whereas resampling layered IEM RMSEs is described in Materials and Methods. Abbreviations for vRF changes included in each model: None, vRF parameters identical to fixation condition; a, amplitude; s, size; p, position; b.

We found that reconstruction amplitude was selectively increased near the attended location in V4, V3A/B, and IPS0 (interaction effect, bootstrapped $p < 0.005$; Fig. 5; Table 3). This can be interpreted as a local boost in signal-to-noise ratio. Prior reports found that attending to the mapping stimulus, as opposed to attending to a peripheral target in the current experiment, caused an increase in the amplitude of all stimulus reconstructions (Sprague and Serences, 2013). That is, representations of task-relevant stimuli increased in signal-to-noise ratio. We find here that even representations of task-irrelevant stimuli near the attended region of space increase in amplitude, consistent with the idea of an attentional "spotlight," which boosts the fidelity of spatial representations near the attention target.

Although the amplitude interaction effect was present in most visual areas we tested (Fig. 5), we found other effects limited to V3A/B and IPS0 that involved modulations in stimulus representations in the ignored hemifield. In these regions, we found that stimulus reconstructions in the ignored hemifield shifted away from the ignored target location (interaction, bootstrapped $p < 0.005$). We also observed a relative size increase near the ignored attention stimulus in IPS0 (interaction, bootstrapped $p < 0.005$). These results suggest that stimulus reconstructions in the ignored hemifield are less spatially precise in posterior parietal cortex. Finally, there was also a main effect of attention on reconstruction size and baseline in areas V3, V4, and V3A/B (bootstrapped $p < 0.005$). However, unlike the interaction effect in IPS0, these size and baseline changes did not vary as a function of distance between the reconstruction and the attended target location.

Using a layered encoding model to explore how single vRFs change population-level codes

In our final analysis, we used a layered spatial encoding model to determine how changes in vRF properties affected the representations of mapping stimuli in the multivariate reconstructions discussed in the previous section (Figs. 1*c*, 4*a*). The goal of this analysis was to determine which vRF modulations contribute the most to the observed increase in the amplitude of stimulus representations around the attended location (Fig. 5). This analysis thus complements our analysis of the spatial discriminability metric, which demonstrated that vRF position changes significantly increased the ability of the population to make fine spatial discriminations near the attention target (Fig. 3*b*).

The layered spatial encoding model we built links the response properties of single voxels to the encoding properties of a whole population of voxels in a region of visual cortex (Fig. 6*a*). In the first layer of the model, we used the fit vRFs to generate simulated BOLD data from each voxel under different attention conditions. We then repeated the multivoxel stimulus reconstruction analysis on this simulated data to model population results for the second layer of the model. This approach allowed us to perform virtual experiments to test how changes in the first layer impacted the second layer. That is, we manipulated which vRF parameters changed with attention (first layer) and observed the resulting changes in the population-based stimulus reconstructions (second layer). For example, we could test whether an overall increase in vRF response gain with attention would be necessary or sufficient to reproduce the amplitude increases observed in the empirical stimulus reconstructions reported in Figure 5. These virtual experiments also allowed us to compare the relative impact of one type of response modulation (e.g., size changes) with other types of response modulations (e.g., position shifts).

Because the population-level stimulus reconstructions require many voxels from each subject to produce stable and reliable results, we combined the data across several regions in each individual subject before estimating the IEM (for a longer discussion, see Materials and Methods). This yielded one occipital region that combined data from areas V1, V2, V3, and V4, and one posterior parietal region that combined data from V3A/B and IPS0. We repeated the IEM analysis described in the previous section on these larger regions and found that the pattern of attentional modulations observed earlier was consistent in the large ROIs (Fig. 5). Next, to verify whether we could perform the layered IEM using a reduced number of voxels, we reran the IEM analysis but only used the data from voxels with well-fit vRFs. The reduced dataset with fewer voxels reproduced the main pattern of results we observed in the previous section. In particular, covert attention led to an increase in the amplitude of reconstructions near the locus of attention (Fig. 6*b*, black vs gray bars).

We then investigated the contribution of each vRF parameter to the population-level stimulus reconstructions, in a comparison akin to the spatial discriminability analysis in Figure 3. A model that only simulated the observed vRF amplitude or vRF size modulations did not predict the observed increase in reconstruction amplitude near the attention target (Fig. 6*b*, red lines). However, a layered model that only simulated vRF position changes did predict a large increase in reconstruction amplitude near the attention target in the parietal ROI (Fig. 6*b*, blue line on right). This is consistent with the effects observed in the full dataset (Fig. 5; Table 3), where we only observed an interaction of stimulus distance and attention in the parietal ROI.

To more formally quantify each manipulation of the layered IEM, we calculated an error metric to describe how well each model reproduced the attentional modulations in the empirical data (using the RMSE) (Table 4). We compared each model's RMSE with a baseline model, which did not simulate any vRF attentional modulations (Fig. 6*c*, far left). This null baseline should have the highest error, and any good models should decrease the RMSE between the simulated data and the empirical data. Conversely, a model with higher RMSE is worse at accounting for the empirical data compared with the null model. In both the occipital and parietal ROIs, adding vRF position shifts to the layered model decreased RMSE while abolishing position shifts generally increased the model error (Fig. 6*c*). These data are consistent with the results from the spatial discriminability analysis. Together, they suggest that shifts in vRF position have the largest impact on population-level representations, whereas changes in vRF size or gain play smaller roles in changing the fidelity of the population code.

Table 4. Continued

s/a	p/s	p	a	s	None
0.194 (0.194, 0.194)	0.141 (0.141, 0.141)	0.143 (0.143, 0.143)	0.194 (0.194, 0.194)	0.194 (0.194, 0.194)	0.193 (0.193, 0.193)
0.451 (0.447, 0.456)	0.410 (0.405, 0.416)	0.410 (0.406, 0.414)	0.441 (0.437, 0.445)	0.447 (0.442, 0.452)	0.834 (0.744, 0.959)

Discussion

By simultaneously measuring the response properties of both single voxels and populations of voxels within retinotopic areas of visual cortex, we could link attentional modulations of spatial encoding properties across scales. Our data provide an initial account of how different types of RF modulations improve the quality of population codes for visual space. First, we show how vRF attentional modulations depended on the distance between the vRF's preferred position and the static attention target (Fig. 2). We then found that shifts in the preferred position of vRFs near the attended target increased the spatial discrimination capacity of a population of voxels (Fig. 3), as well as the amplitude of stimulus reconstructions based on response patterns across all voxels in a ROI (Figs. 5, 6).

Attentional modulations of spatial RFs

We provide new data on how vRF responses are modulated around a covertly attended static target (Sprague and Serences, 2013; Klein et al., 2014; de Haas et al., 2014; Kay et al., 2015; Sheremata and Silver, 2015). Like prior macaque studies, we find that vRF position shifts depend on the vRF's distance from the attended target (Connor et al., 1996, 1997). However, we also found that the pattern of attentional modulations differs across the visual hierarchy. In V4, V3A/B, and IPS0, voxels shift toward the attended target, whereas in earlier areas, vRFs near the attended target are slightly repelled from it (Fig. 2*e*). We also found distinct patterns of size modulations: vRF size increased near the attention target in early visual areas but decreased in parietal areas V3A/B and IPS0. Comparison with the existing literature suggests that patterns of RF size modulations likely depend on the nature of the spatial attention task. In fMRI tasks where subjects attended to the mapping stimulus, rather than a static position, researchers report that average vRF size increases with attention (Sprague and Serences, 2013; Kay et al., 2015; Sheremata and Silver, 2015). RFs in macaque area MT shrink when measured with a mapping probe smaller than the stimulus but increase in size when macaques track the mapping probes as they move across the screen (Womelsdorf et al., 2006, 2008; Anton-Erxleben et al., 2009; Niebergall et al., 2011). This may be because the RFs shift position to track the probe, causing an apparent increase in overall size. Last, manipulating perceptual load at fixation also increases vRF size in human visual cortex (de Haas et al., 2014). Together, these observations demonstrate that the pattern of RF response modulations depends both on task demands and on the spatial relationship between the attended target and the encoding unit's RF.

We note that, although the similarity between attentional modulations of single cell RFs and single vRFs is compelling, their properties are derived from different input signals and are not interchangeable. fMRI voxels in retinotopically organized regions of visual cortex sample from a broad array of neurons with approximately the same spatial tuning preferences, so a position shift in a vRF could either be driven by a change in the preferred position of single neurons, or by a change in the gain profile across neurons tuned to slightly different locations in the visual field. Similarly, single neuron RFs receive input from smaller RFs in earlier visual areas, and a position shift could arise from either mechanism described above (McAdams and Maunsell, 1999; Baruch and Yeshurun, 2014; Dhruv and Carandini, 2014). Because of this inherent ambiguity when measuring the encoding properties of a locally tuned unit, it is useful to compare them with attentional modulations of population-level representations.

Attention boosts the spatial encoding fidelity of a population

We first measured the overall capacity of a population of voxels to make fine spatial discriminations in a region of space. We found that attention increased spatial discriminability near the attended target, relative to the ignored target. We then performed virtual experiments on the vRFs contributing to the population to determine how they affected the spatial discriminability metric. We report that vRF position shifts increased spatial discriminability significantly more than vRF size changes or gain changes (Fig. 3).

Because the spatial discriminability metric (Fig. 3) is only informative about a local portion of space, we performed a second population analysis to reconstruct an image of the entire visual field on each trial using a multivariate IEM. Attention increased the amplitude of stimulus reconstructions near the attention target, indicating an increase in representational fidelity that accompanied the increase in spatial discriminability. In addition, a layered spatial encoding model revealed that shifts in vRF position could account for these attentional enhancements in the population-level stimulus reconstructions, but changes in vRF size could not. Together, our data demonstrate that shifts in position of many RFs may be a dominant way that single encoding units alter the properties of a population spatial code.

Although population-level information increased the most with changes in vRF position, we reiterate that these position changes could arise from spatially specific patterns of gain modulations in input RFs. If this is true, it is possible that gain modulations with attention may exert their largest effects on the downstream population, where these patterns of gain changes become apparent shifts in vRF position. However, this remains an open question for future work to address.

Our findings also underscore the fact that changes in the spatial encoding properties of single units do not directly translate into analogous changes in the encoding properties of a population of those same units. For example, an overall change in vRF size does not necessarily change the size of the population-level representation (Sprague and Serences, 2013; Kay et al., 2015). Although we found that single units shifted their preferred position toward the attended target, population-level representations did not generally shift with attention. When the population code did shift its encoded position, we found that it was typically representations of the ignored stimulus that shifted farther from the true stimulus location (Fig. 5), consistent with more error-prone representations of irrelevant stimuli. These types of differences further emphasize the need to understand the effects of cognitive state on population codes for the entire visual scene, rather than focusing solely on single units.

Last, we note that our population-level data do not address the open question of whether RF attentional modulations have perceptual consequences because it is not clear how the spatial encoding models measured here are linked to visual perception and behavior (Koenderink, 1990; Rose, 1999; Anton-Erxleben and Carrasco, 2013; Klein et al., 2016). Further investigation into these topics should include task manipulations to investigate how attentional modulations of both vRFs and population-level metrics track psychophysical performance.

Tuning shifts and labeled lines

Historically, shifts in the tuning of a RF have not been considered one of the main mechanisms by which attention modulates population-level information, although recent reports suggest that this view is being reconsidered (David et al., 2008; Anton-Erxleben and Carrasco, 2013). This may be due to "labeled-line" theories of visual information processing, which posit that a single neuron has a consistent feature label, which downstream neurons rely on to perform computations and transmit stable information (Barlow, 1972; Doetsch, 2000; David et al., 2008). When a spatial RF shifts position as a function of cognitive state (e.g., attention), that single neuron's feature label is no longer consistent. Without an accompanying shift in the downstream neurons receiving the changing feature label, such a change could disrupt the stability of the population code. However, our results suggest that population-level spatial representations remain relatively stable, and are even enhanced, when the tuning of the underlying vRFs shift in position, size, and gain.

An alternate proposal to a labeled line code relies on the joint information encoded across a population of cells (Erickson, 1982; Doetsch, 2000). This may occur at several scales; for example, V2 could use the pattern of information from V1 inputs to form a visual representation. This idea is more akin to an encoder-decoder model in which the downstream decoder does not need information about the altered representations in each of the encoder units but instead relies on a population readout rule (Seriès et al., 2009). The population readout rule could incorporate knowledge about the "labels" of the encoder units but could perform equally well by relying on relative changes in the pattern across units to resolve the information encoded in the population. However, further exploration of population readout rules in visual cortex are needed to test this hypothesis.

In conclusion, the spatial encoding properties of the visual system can be measured and modeled at many different spatial scales. Here, we report how these properties change with attention for single voxels and for a group of voxels in each ROI. Notably, single vRF modulations do not propagate directly to analogous changes in large-scale codes. Instead, we observed that attentional modulations of vRF position play a dominant role in modulating the amplitude of population-level representations. Future research is needed to resolve how shifts in RF labels are generated, how information is read out from a population, and how these multiscale attentional modulations affect visual perception and behavior.

Notes

Supplemental material for this article is available at https://osf.io/s9vqv/. This repository includes Supplemental Methods and Figures, along with the full dataset and code needed to reproduce the analysis. This material has not been peer reviewed.

References

Amano K, Wandell BA, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT $^{+}$ complex. J Neurophysiol 102:2704–2718. CrossRef Medline

Anton-Erxleben K, Carrasco M (2013) Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. Nat Rev Neurosci 14:188–200. CrossRef Medline

Anton-Erxleben K, Stephan VM, Treue S (2009) Attention reshapes center-surround receptive field structure in macaque cortical area MT. Cereb Cortex 19:2466–2478. CrossRef Medline

Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1:371–394. CrossRef Medline

Baruch O, Yeshurun Y (2014) Attentional attraction of receptive fields can explain spatial and temporal effects of attention. Vis Cogn 22:704–736. CrossRef

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188. CrossRef

Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. J Neurosci 29:13992–14003. CrossRef Medline

Brouwer GJ, Heeger DJ (2013) Categorical clustering of the neural representation of color. J Neurosci 33:15454–15465. CrossRef Medline

Butts DA, Goldman MS (2006) Tuning curves, neuronal variability, and sensory coding. PLoS Biol 4:e92. CrossRef Medline

Cohen MR, Maunsell JHR (2014) Neuronal mechanisms of spatial attention in visual cerebral cortex. In: The Oxford handbook of attention, pp 318–345. Oxford: Oxford University.

Connor CE, Gallant JL, Preddie DC, Van Essen DC (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. J Neurophysiol 75:1306–1308. Medline

Connor CE, Preddie DC, Gallant JL, Van Essen DC (1997) Spatial attention effects in macaque area V4. J Neurosci 17:3201–3214. Medline

David SV, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron 59:509–521. CrossRef Medline

de Haas B, Schwarzkopf DS, Anderson EJ, Rees G (2014) Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. Curr Biol 24:R66–R67. CrossRef Medline

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222. CrossRef Medline

Dhruv NT, Carandini M (2014) Cascaded effects of spatial adaptation in the early visual system. Neuron 81:529–535. CrossRef Medline

Doetsch GS (2000) Patterns in the brain: neuronal population coding in the somatosensory system. Physiol Behav 69:187–201. CrossRef Medline

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647–660. CrossRef Medline

Erickson RP (1982) The across-fiber pattern theory: an organizing principle for molar neural function. In: Contributions to sensory physiology, Vol. 6, pp 79–110. New York: Academic Press.

Ester EF, Anderson DE, Serences JT, Awh E (2013) A neural measure of precision in visual working memory. J Cogn Neurosci 25:754–761. CrossRef Medline

Ester EF, Sprague TC, Serences JT (2015) Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. Neuron 87:893–905. CrossRef Medline

Gattass R, Nascimento-Silva S, Soares JG, Lima B, Jansen AK, Diogo AC, Farias MF, Botelho MM, Mariani OS, Azzi J, Fiorani M (2005) Cortical visual areas in monkeys: location, topography, connections, columns, plasticity and cortical dynamics. Philos Trans R Soc Lond B Biol Sci 360:709–731. CrossRef Medline

Harvey BM, Dumoulin SO (2011) The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. J Neurosci 31:13604–13612. CrossRef Medline

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12:55–67. CrossRef

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210–1224. CrossRef Medline

Kay KN, Weiner KS, Grill-Spector K (2015) Attention reduces spatial uncertainty in human ventral temporal cortex. Curr Biol 25:595–600. CrossRef Medline

Klein BP, Harvey BM, Dumoulin SO (2014) Attraction of position preference by spatial attention throughout human visual cortex. Neuron 84:227–237. CrossRef Medline

Klein BP, Paffen CL, Pas SF, Dumoulin SO (2016) Predicting bias in per-

ceived position using attention field models. J Vis 16:1–15. CrossRef Medline
Koenderink JJ (1990) The brain a geometry engine. Psychol Res 52:122–127. CrossRef Medline
Lee S, Papanikolaou A, Logothetis NK, Smirnakis SM, Keliris GA (2013) A new method for estimating population receptive field topography in visual cortex. Neuroimage 81:144–157. CrossRef Medline
Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nat Neurosci 9:1432–1438. CrossRef Medline
McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431–441. Medline
Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science 229:782–784. CrossRef Medline
Navalpakkam V, Itti L (2007) Search goal tunes visual features optimally. Neuron 53:605–617. CrossRef Medline
Niebergall R, Khayat PS, Treue S, Martinez-Trujillo JC (2011) Expansion of MT neurons excitatory receptive fields during covert attentive tracking. J Neurosci 31:15499–15510. CrossRef Medline
Pouget A, Dayan P, Zemel RS (2003) Inference and computation with population codes. Annu Rev Neurosci 26:381–410. CrossRef Medline
Regan D, Beverley KI (1985) Postadaptation orientation discrimination. J Opt Soc Am 2:147–155. CrossRef
Roberts M, Delicato LS, Herrero J, Gieselmann MA, Thiele A (2007) Attention alters spatial integration in macaque V1 in an eccentricity-dependent manner. Nat Neurosci 10:1483–1491. CrossRef Medline
Rose D (1999) The historical roots of the theories of local signs and labelled lines. Perception 28:675–685. CrossRef Medline
Scolari M, Serences JT (2009) Adaptive allocation of attentional gain. J Neurosci 29:11933–11942. CrossRef Medline
Scolari M, Serences JT (2010) Basing perceptual decisions on the most informative sensory neurons. J Neurophysiol 104:2266–2273. CrossRef Medline
Serences JT, Saproo S (2012) Computational advances towards linking BOLD and behavior. Neuropsychologia 50:435–446. CrossRef Medline
Seriès P, Stocker AA, Simoncelli EP (2009) Is the homunculus "aware" of sensory adaptation? Neural Comput 21:3271–3304. CrossRef Medline
Sheremata SL, Silver MA (2015) Hemisphere-dependent attentional modulation of human parietal visual field representations. J Neurosci 35:508–517. CrossRef Medline
Sprague TC, Serences JT (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat Neurosci 16:1879–1887. CrossRef Medline
Sprague TC, Ester EF, Serences JT (2014) Reconstructions of information in visual spatial working memory degrade with memory load. Curr Biol 24:2174–2180. CrossRef Medline
Sprague TC, Saproo S, Serences JT (2015) Visual attention mitigates information loss in small- and large-scale neural codes. Trends Cogn Sci 19:215–226. CrossRef Medline
Sprague TC, Ester EF, Serences JT (2016) Restoring latent visual working memory representations in human cortex. Neuron 91:694–707. CrossRef Medline
Tolias AS, Moore T, Smirnakis SM, Tehovnik EJ, Siapas AG, Schiller PH (2001) Eye movements modulate visual receptive fields of V4 neurons. Neuron 29:757–767. CrossRef Medline
Wandell BA, Winawer J (2015) Computational neuroimaging and population receptive fields. Trends Cogn Sci 19:349–357. CrossRef Medline
Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nat Neurosci 9:1156–1160. CrossRef Medline
Womelsdorf T, Anton-Erxleben K, Treue S (2008) Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. J Neurosci 28:8934–8944. CrossRef Medline
Zirnsak M, Steinmetz NA, Noudoost B, Xu KZ, Moore T (2014) Visual space is compressed in prefrontal cortex before eye movements. Nature 507:504–507. CrossRef Medline

Chapter 1 is a full reprint of the material as it appears in: Vo, Vy A., Sprague, Thomas C., Serences, John T. (2017). Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex. *Journal of Neuroscience 37*(12), 3386-3401. The dissertation author was the primary investigator and author of this paper.

CHAPTER 2: Altering the spatial scope of covert attention induces differential patterns of voxel receptive field modulations

Abstract

Many real-world tasks that require covert spatial attention involve attending to multiple locations in space. However, our knowledge of how this affects visual encoding and information processing is limited. In this fMRI study, human participants covertly attended to either 1 or 3 isoeccentric locations in space and performed a fine spatial discrimination task. For comparison, they also performed an attention task at fixation. We estimated the spatial receptive fields (RFs) of thousands of voxels across retinotopic visual cortex, from early visual area V1 to higher-order parietal areas that are be strongly modulated by attention. We found that dividing attention across 3 locations generally increases the size and response gain of voxel RFs (vRFs), compared to focusing attention at 1 location. Contrary to a prediction from a unitary attention field model, we find that vRFs near the central fixation point shift outward toward the periphery when participants divided their spatial attention across 3 locations. Finally, we find that changes in vRF response gain during the divided attention condition are important for increasing spatial discriminability near the attended location.

Introduction

Covert spatial attention enables humans and other primates to attend to visual inputs without moving the head or eyes. In some cases, maintaining a fixed gaze at a central location while processing peripheral visual inputs can enhance visuomotor precision (Dessing, Rey, & Beek, 2012). Indeed, many real-world examples of covert spatial attention involve covertly attending to multiple locations in space. Yet most work that investigates how covert attention impacts visual information processing uses a single peripheral target. Here we set out to investigate how the visual system enhances the representations of multiple covertly attended locations in space.

Prior work has shown that covert spatial attention to a single static location alters the basic response properties of spatial receptive fields (RFs). Electrophysiological studies in nonhuman primates have demonstrated that covert attention can alter the position, size, and

response gain of spatial RFs of single units through visual cortex (Anton-Erxleben, Stephan, & Treue, 2009; Connor, Gallant, Preddie, & Van Essen, 1996; Connor, Preddie, Gallant, & Van Essen, 1997; David, Hayden, Mazer, & Gallant, 2008; Moran & Desimone, 1985; Niebergall, Khayat, Treue, & Martinez-Trujillo, 2011b, 2011a; Roberts, Delicato, Herrero, Gieselmann, & Thiele, 2007; Womelsdorf, Anton-Erxleben, Pieper, & Treue, 2006; Womelsdorf, Anton-Erxleben, & Treue, 2008). Several groups have reported that the centers of single unit spatial RFs shift toward a covertly attended location in space (Anton-Erxleben et al., 2009; Connor et al., 1996, 1997, Womelsdorf et al., 2006, 2008). In addition, others have found that RF size can increase or decrease depending on the spatial relationship between the RF and the attended location (Anton-Erxleben et al., 2009; Niebergall et al., 2011a). More recent work using functional magnetic resonance imaging (fMRI) has also shown that covert attention modulates the responses of single voxel RFs (de Haas, Schwarzkopf, Anderson, & Rees, 2014; Kay, Weiner, & Grill-Spector, 2015; Klein, Harvey, & Dumoulin, 2014; Puckett & DeYoe, 2015; Sheremata & Silver, 2015; van Es, Theeuwes, & Knapen, 2018; Vo, Sprague, & Serences, 2017). Although this work necessarily relies on measuring the aggregate responses of thousands of single units – such that they are typically referred to as “population receptive fields” (pRFs) – their findings correspond very well to the single unit literature. For example, when covertly attending to a single static location, the spatial RF centers of single voxels tends to shift toward the attended location (Klein et al., 2014; Vo et al., 2017) and can both increase or decrease in size (Vo et al., 2017). We note here that to avoid confusion with analyses which aggregate over a group of spatial RFs, we reserve the term “population” to refer to a group of voxels, and instead refer to these single-voxel measurements as voxel receptive fields (vRFs).

Still, it is unknown whether covertly attending to multiple static targets would alter the pattern of vRF modulations. To our knowledge, the only studies which have mapped spatial RFs during a divided attention task used a multiple object tracking task in which the attended locations moved across the screen (Niebergall et al., 2011a, 2011b). They observed increases

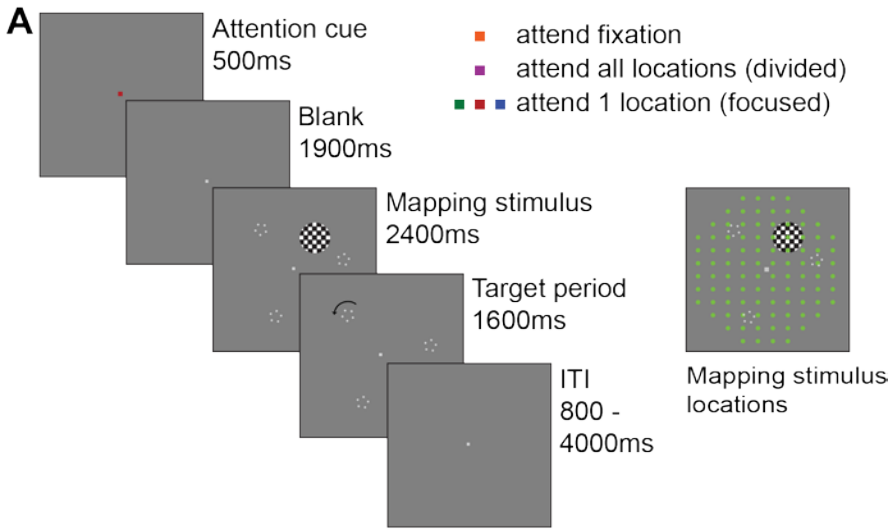
in the RF size of cells in area MT when the attended stimuli passed through the RF (Niebergall et al., 2011a). This increase in size may be driven by the fact that the attended locations are moving, rather than static. Indeed, in fMRI studies which require the participant to attend to a stimulus moving across the screen, researchers have always reported vRF size increases (Kay et al., 2015; Sheremata & Silver, 2015; Sprague & Serences, 2013). However, dividing attention across multiple static locations may elicit altogether different patterns of RF modulations. Some work has measured univariate (i.e. mean) fMRI activity with a small and large scope of attention. They find that this manipulation elicits more widespread activity across retinotopic cortex (Herrmann, Montaser-Kouhsari, Carrasco, & Heeger, 2010; Itthipuripat, Garcia, Rungratsameetaweemana, Sprague, & Serences, 2014; McMains & Somers, 2004, 2005; M. M. Müller, Malinowski, Gruber, & Hillyard, 2003). While they did not measure individual voxel RFs, their data do suggest that vRFs within the scope of attention may increase in response amplitude or baseline.

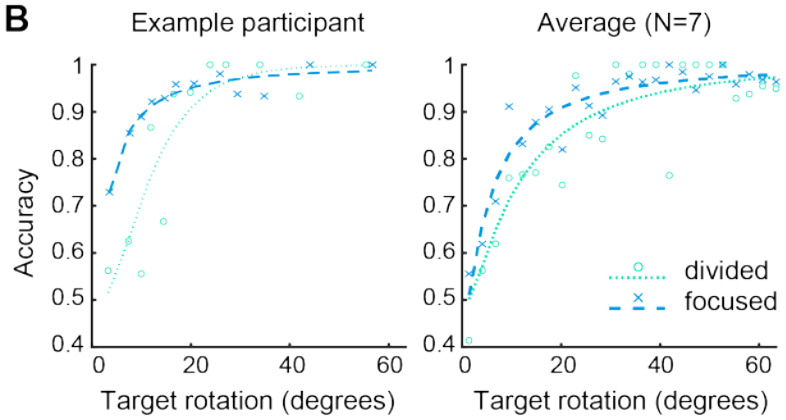
Altering the spatial scope of attention also provides a test of theories that posit an "attention field" to account for changes in RF properties. An attention field is defined as a spatially limited area (typically modeled as a Gaussian) that can modulate the responses of RFs which overlap with the attention field (Reynolds & Heeger, 2009). It has been proposed that a multiplicative interaction between a spatial attention field and individual RFs can account for apparent changes in RF position (Klein et al., 2014; van Es et al., 2018; Womelsdorf et al., 2008). Under this account, expanding the scope of spatial attention would produce a different pattern of vRF modulations. For example, if attention was allocated to several isoeccentric targets, one might expect a single large attention field centered on the fixation point. This would generate the prediction that all spatial RFs would move toward the center of the attention field, rather than toward the peripheral targets.

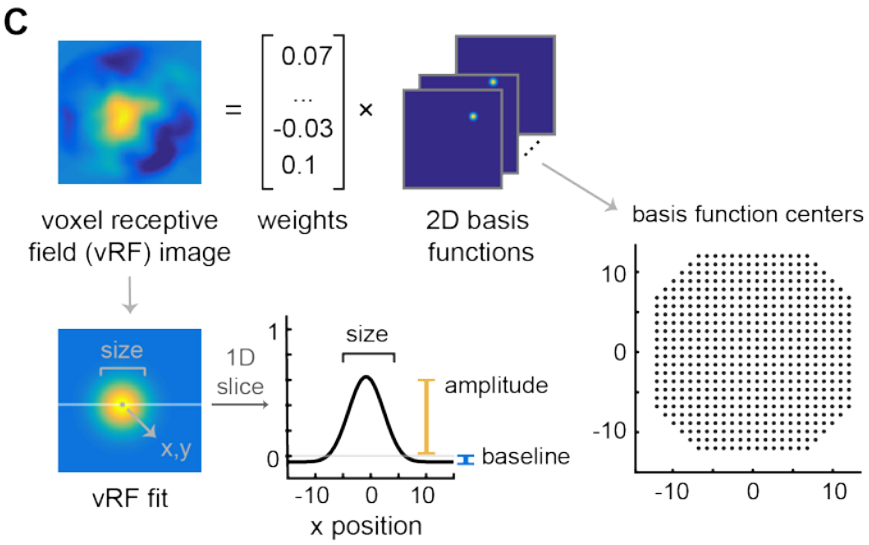
In the present study, we measured fMRI responses while human participants covertly attended to either 1 or 3 peripheral locations (Figure 2.1). They then made fine spatial

discrimination judgments at the peripheral location. For comparison, we also had participants perform an overt attention task at the central fixation point. First we estimated spatial RFs for thousands of voxels across retinotopic cortex during each attention condition. We characterized how expanding the spatial scope of covert attention altered the vRF response properties. Then we investigated how these response modulations affected the ability of a population of RFs to make fine spatial discriminations.

Figure 2.1. Covert spatial attention task. A) Participants were cued to a single location (focused condition) or to all three peripheral locations (divided condition). Their task was to discriminate the direction (clockwise or counter-clockwise) of a small rotation in a stimulus occupying the attended location. Only one stimulus rotated on each trial. In the fixation condition, participants were cued to attend the central fixation point and discriminated the direction of a luminance change. B) Psychophysical curves illustrating lower thresholds, or heightened sensitivity, in the focused compared to the divided attention condition. C) We assumed that each voxel receptive field (vRF) could be represented as a weighted sum of a set of 2D basis functions (Methods). These basis functions can also be thought of as spatial filters. After solving for each vRF based on its BOLD response to the mapping stimuli, we fit the vRF image with an isotropic, Gaussian-like function that was specified by 5 parameters (x, y, size, amplitude, and baseline).


A
Attention cue
500ms
Blank
1900ms
Mapping stimulus
2400ms
Target period
1600ms
ITI
800 -
4000ms
attend fixation
attend all locations (divided)
attend 1 location (focused)
Mapping stimulus
locations


B
Example participant
Average (N=7)
Accuracy
Target rotation (degrees)
Target rotation (degrees)
divided
focused


C
0.07
...
-0.03
0.1
voxel receptive
field (vRF) image
weights
2D basis
functions
basis function centers
size
x,y
vRF fit
1D
slice
size
amplitude
baseline
x position

Results

We trained 7 participants on a covert spatial attention task which required fine position discriminations. On each trial, participants either focused their attention on a single location in the periphery, divided their attention across three locations, or attended to the central fixation point and made a two-alternative forced choice discrimination judgment (Figure 2.1A). In the covert attention conditions, their task was to detect a movement in one of the three peripheral target stimuli and discriminate the direction of its rotation. In the attend fixation condition, they discriminated a change in luminance of the fixation point. The attention cues were always 100% valid, and visual stimulation was identical across conditions. In order to control for changes in task performance across runs and sessions, we used an adaptive and ongoing psychometric procedure (Watson and Pelli, 1983; Methods) to determine the stimulus change needed to achieve 60%, 75%, or 90% performance on each attention condition, in each run. This ensured that participants maintained a consistent level of effort throughout all fMRI scans.

While they directed their attention toward the spatial location(s), we presented a task-irrelevant probe stimulus at different locations across the screen to map the spatial sensitivity of visually responsive voxels in several retinotopic regions of interest (ROIs) from V1 to dorsal parietal region IPS2. Each participant completed 2-3 sessions of behavioral training and 3 two-hour sessions of fMRI scanning (Methods).

Behavioral performance

We first examined whether the covert attention manipulation affected behavioral performance on a fine spatial discrimination task. Dividing covert attention across 3 targets should increase their behavioral threshold (i.e. increase the amount of rotation required to achieve 75% accuracy). We fit psychometric functions for each participant and condition (Figure 2.1B, Figure 2.S1) based on the behavioral data collected during the fMRI scanning sessions. On average, participants required a 19.28° rotation in the divided condition (95% CIs [9.87, 32.33]) and a 14.21° rotation in the focused condition (95% CIs [6.82, 24.74] to achieve 75%

accuracy. We then computed each participant's change in threshold between the divided and focused conditions. Across our participant sample, the behavioral threshold for the divided condition was larger than the focused condition by approximately 4.04° (95% CIs [1.74°, 6.10°], $p < .001$). These data demonstrate that the attentional manipulation was effective, and that participants were worse at making fine position discriminations in the divided attention task.

vRF data quality

We then estimated spatial RFs for each voxel and each participant in retinotopic regions V1, V2, V3, V4, V3A/B, IPS0, IPS1, and IPS2 using a regularized linear spatial encoding model (Figure 2.1C, Methods). This generated an image of a vRF that could assume any shape. We then fit this image with a unimodal, Gaussian-like isotropic function which parameterized each vRF and described its x-y center, size, amplitude, and baseline. Here we define baseline as any spatially non-specific response. The amplitude is the height of the spatially specific response. We then applied a series of strict thresholds to ensure that we were only analyzing voxels that are well-described by the spatial encoding model, and subsequently fit well by the Gaussian-like function. Table 2.S1 lists the number of voxels that survived each threshold.

We then examined two metrics describing the quality of the estimated voxel RFs to ensure that we could make direct comparisons across conditions and note trends across retinotopic regions of interest (ROIs). The first metric described the ability of the vRF to predict held-out responses (i.e., the ability of the regularized spatial encoding model to describe voxel responses), and the second metric described the goodness-of-fit between the vRF and the Gaussian-like function (Supplemental Methods). If vRF quality was unexpectedly low in a specific condition or ROI, it would indicate a failure of our methods to estimate proper vRFs. We then examined how both metrics varied with condition or ROI. The biggest effect was that ROIs in early visual areas were best described by the spatial encoding model, which is to be expected. Otherwise, the metrics did not vary much across either factor (Figure 2.S2).

These procedures confirmed that we could compare vRF parameters across ROIs and attention conditions. We then moved on to the principal aims of the study: first, to investigate how manipulating attentional scope would change vRF response properties, and second, to determine which vRF modulations increased the spatial discriminability of the entire population.

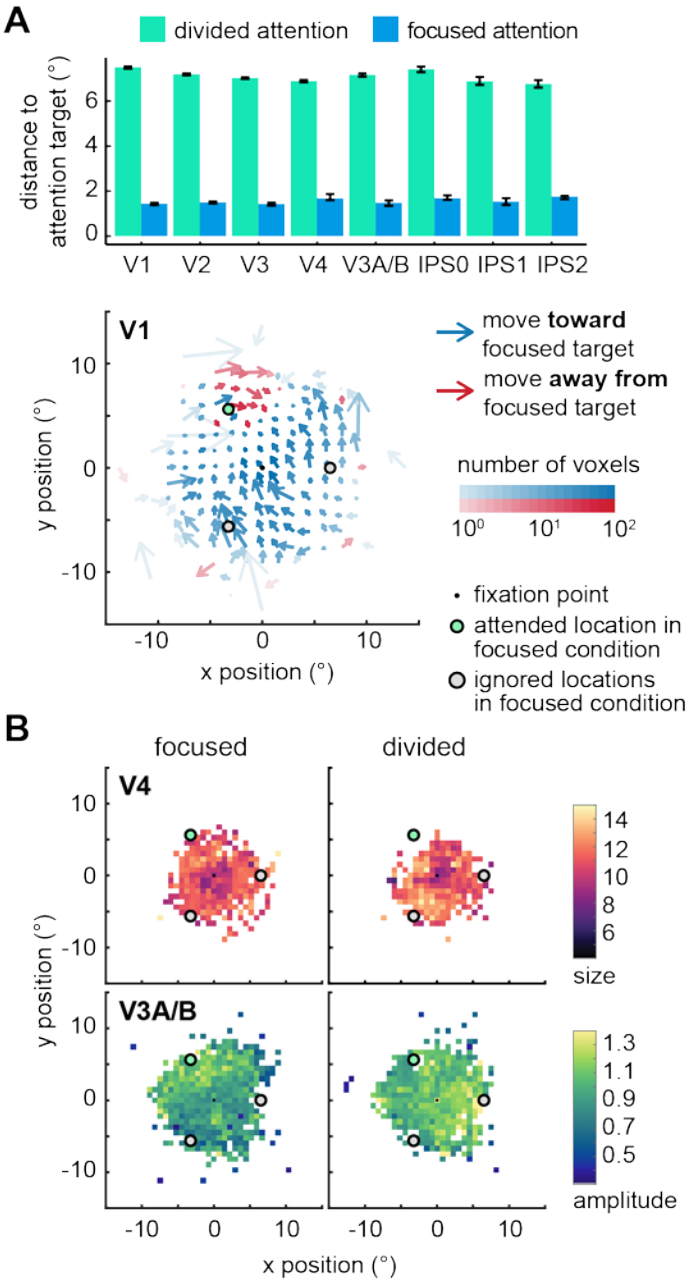
Modulations of vRF parameters with attention condition

To address the first aim, we compared the vRF parameters (i.e., position, size, amplitude, and baseline) across the two covert attention conditions. Previous work on attentional modulations of spatial RFs compared a covert attention condition to an easy task at fixation. However, an overt attention task, in which the focus of gaze and the focus of attention are the same, likely recruits fundamentally different mechanisms compared to tasks that require covertly dissociating gaze from attention (Kelley, Serences, Giesbrecht, & Yantis, 2008; Moore, Armstrong, & Fallah, 2003; Moore & Fallah, 2004; Umiltà, Riggio, Dascola, & Rizzolatti, 1991). Therefore, since the conditions of interest are the two covert attention manipulations, comparing them to one another is the most direct way to characterize how changes in the spatial scope of covert attention impact voxel-level tuning characteristics. The fixation task and its relation to the covert attention conditions will be discussed in more detail in the next section.

First, we examined whether changing the spatial scope of covert attention affected vRF positions. Previous reports have shown that changes in vRF position are widespread across the visual field (Klein et al., 2014; van Es et al., 2018; Vo et al., 2017), so we started by evaluating position shifts across all measured vRFs, irrespective of their initial spatial center. We computed the distance between the x-y center of each vRF and the attended target. In the focused condition, we estimated a separate vRF for each attention target (e.g., 1 vRF when attending target 1 and ignoring targets 2 and 3; another vRF when attending target 2 and ignoring targets 1 and 3, etc.). We only included the distance between the vRF and whichever target was attended at the time, yielding 3 values for the focused condition. Then we compared these values to the distance between a single vRF in the divided condition to each of the 3 attended

targets, yielding another 3 values. We found that spatial vRFs in the focused condition were closer to the attended targets overall (Figure 2.2A, 2.2B). This effect was observed in all ROIs (median position difference ranged from 4.98° in IPS2 to 6.06° in V1; all $p < .001$).

Figure 2.2. vRF changes between the focused and divided attention conditions, averaged across all participants. All attention locations are coregistered to these common points, with the green circle indicating the coregistered focused attention target. A) Top: vRFs in the focused condition are closer to the attended location. Bottom: A vector plot illustrating the changes in vRF position between the focused and divided conditions. vRFs move toward the attended location (green circle). The start of each arrow is the average vRF location in the divided condition, and the end of the arrow is the average location of those same voxels when attending to the upper left position (green circle) in the focused condition. B) Average size and amplitude in each condition for example ROIs V4 and V3A/B. Each plot is a 2D histogram of the parameter values in that part of space. The data shown here is only for 1 focused location (upper right target). Both size and amplitude are greater in the divided condition near the targets which are ignored in the focused attention condition (gray).


A
divided attention
focused attention
distance to attention target (°)
6
4
2
0
V1 V2 V3 V4 V3A/B IPS0 IPS1 IPS2
V1
move toward focused target
move away from focused target
number of voxels
10^0 10^1 10^2
y position (°)
10
0
-10
x position (°)
fixation point
attended location in focused condition
ignored locations in focused condition
B
focused
divided
V4
V3A/B
14
12
10
8
6
size
1.3
1.1
0.9
0.7
0.5
amplitude

We then examined the remaining parameters. While the effects of covert attention on RF size, amplitude, and baseline are less well-understood, they appear to be dependent on the spatial relationship between the x-y position of the RF and the attended location (Anton-Erxleben et al., 2009; Connor et al., 1996, 1997; de Haas et al., 2014; Kay et al., 2015; Klein et al., 2014; Niebergall et al., 2011a; Sheremata & Silver, 2015; Sprague & Serences, 2013; van Es et al., 2018; Vo et al., 2017; Womelsdorf et al., 2008). We first tested this hypothesis by modeling each vRF parameter (size, amplitude, or baseline) as a function of its distance from the attended location and condition. We did this with a linear mixed effects model, which is akin to a traditional ANOVA but can handle unbalanced data. Our data was unbalanced because each participant contributed a varying number of voxels which covered different parts of space. We fit a different model was fit for each ROI (Methods). First, we fit a model predicting the vRF parameter as a linear function the vRF's distance from the attended location. Then we added the factor of condition to see if it accounted for significantly more variance in the vRF parameters than the model with only the distance. This allowed us to look at the overall effect of condition on each parameter while accounting for variation over space. The main effect of condition was significant in all ROI-parameter combinations, even after multiple comparisons correction (Figure 2.3).

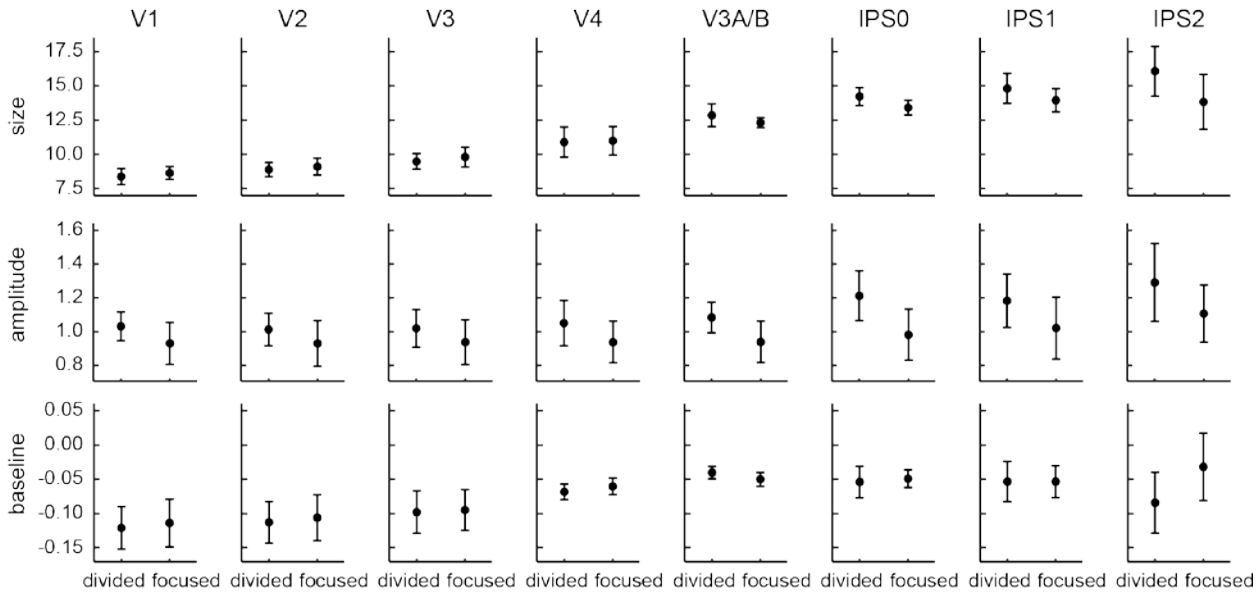


Figure 2.3. Estimated mean parameter values for each covert attention condition, taken from the mixed effects model fits. All ROIs and parameters showed a significant effect of condition (FDR q = 0.05).

We limited our subsequent investigations to areas near the peripheral targets, which are always attended in the divided condition or either attended or ignored in the focused condition. Studies that have reported effects on size or response gain (Anton-Erxleben et al., 2009; Connor et al., 1996, 1997; de Haas et al., 2014; Kay et al., 2015; Klein et al., 2014; Niebergall et al., 2011a; Sheremata & Silver, 2015; Sprague & Serences, 2013; van Es et al., 2018; Vo et al., 2017; Womelsdorf et al., 2008) and baseline (Anton-Erxleben et al., 2009) find the largest effects near the attended stimulus.

Changes in size, amplitude, and baseline with attentional scope

We tested three hypotheses about the two covert attention conditions (Table 2.1). After selecting only voxels near the peripheral targets, we compared the distributions of parameter values across condition by bootstrapping the medians (see Methods*, vRF parameter statistics*). We then corrected for multiple comparisons across all tests performed about a single parameter (e.g. about vRF size), including all the repeated tests for each ROI (e.g. 3 hypotheses * 8 ROIs = 24 tests).

Table 2.1. Statistical tests of vRF size, amplitude, and baseline for the two covert attention conditions: divided (attend 3) and focused (attend 1). Results highlighted in gray go against the trend observed in most ROIs.

Comparison	*Size*	*Amplitude*	*Baseline*
1. Attended vs. ignored targets (focused condition only)	Att > Ig: V3, V4, V3AB Ig > Att: IPS1, IPS2	Att > Ig: V3, V4, V3AB, IPS0, IPS1	Att > Ig: V1, V2, V3, V4, V3AB, IPS0, IPS1, IPS2
2. Focused condition vRFs near attended target vs. divided condition vRFs near the same target	Div > Foc: V4, IPS0, IPS2	Div > Foc: V1	Foc > Div: V4, IPS0
3. Focused condition vRFs near ignored targets vs. divided condition vRFs near the same targets	Div > Foc: V2, V3, V4, V3AB, IPS0 Foc > Div: IPS1	Div > Foc: V1, V3, V4, V3AB, IPS0, IPS1	Div > Foc: V1, V2, V3, V4, V3AB, IPS0

We first looked within the focused attention condition and compared vRFs near the attended targets to those near the ignored targets (hypothesis 1, Table 2.1). Both vRF amplitude and baseline were higher near the attended areas across most ROIs (median difference in amplitude 0.041 in V3 to 0.169 in IPS0; median difference in baseline 0.037 in IPS2 to 0.073 in V3). However, the effect of attention on vRF differed between early ROIs with smaller spatial RFs and later ROIs with larger RFs. vRF sizes near the attended areas were larger in V3, V4, and V3A/B (median difference 0.370°, 0.221°, 0.015°, respectively; all $p < .001$). Yet in IPS1 and IPS2, vRFs near the attended stimuli shrank compared to those near the ignored stimuli (median difference -1.240°, -2.090° respectively; $p < .001$).

We then turned to comparisons between the divided and focused attention conditions. We investigated an area around each of the single targets in the focused condition, and compared it to the same area in the divided condition (hypothesis 2, Table 2.1). The effects were consistent in ROIs that survived our statistical thresholds. First, we observed that vRF sizes were larger in the divided condition, but only in higher-order visual regions (median difference and p-value: V4 0.898°, $p < .001$; IPS0 0.012°, $p = 0.02$; IPS1 2.807, $p = 0.004$). A size increase would also be consistent with a decrease in resolution in the divided condition,

which corresponds to a higher psychophysical threshold (Figure 2.1). vRF amplitude only differed in area V1 (median difference 0.067, $p = 0.018$). And finally, response baseline was higher in the focused condition in V4 (median difference -0.027, $p = 0.004$) & IPS0 (median difference -0.023, $p < .001$). These results suggest that response gain near the attended targets was largely similar between the two conditions.

Lastly, we investigated the areas around the two ignored targets in the focused condition, and compared it to the same areas in the divided condition (hypothesis 3, Table 2.1). Note that these areas were attended in the divided condition, but ignored in the focused condition. The differences in size and amplitude were similar to the trends we observed near the attended targets: vRF sizes were larger and amplitude was higher in the divided condition. However, the regions where we observed a significant effect differed. The size increase was observed in earlier ROIs (e.g. median differences: V2 0.357°, $p < .001$; V3 0.637°, $p < .001$) and in more ROIs overall (Figure 2.2B). We also found that in IPS1, the effect reverses: vRFs near the ignored targets in the focused condition were larger than the same (attended) areas in the divided condition (median difference -1.243°, $p < .001$). Taken together with the size effect reversal observed in hypothesis 1, the results suggest that there might be an interaction between ROI (or RF size) and the scope of attention. That is, ROIs with large RFs will decrease in size if the scope of attention is sufficiently small.

Both the amplitude and baseline effects we observe near the attended areas in the divided condition are robust. They are both higher than the same ignored areas in the focused condition (median amplitude difference ranges from 0.042 in V1 to 0.210 in IPS0; median baseline difference ranges from 0.021 in IPS0 to 0.070 in V3; all $p < .001$). The baseline increases may be partially driven by increased general arousal that globally increases the BOLD response.

In sum, we found that increasing the spatial scope of attention generally increased vRF sizes, except in the late parietal ROIs. We also found that changing the scope of attention generally increased both vRF response gain and baseline near the attended targets.

Comparing the fixation condition to the covert attention conditions

Next we compared modulations in each covert attention condition to modulations that occurred when observers discriminated changes in luminance at the central fixation point. This is an overt, endogenous attention task that was controlled in difficulty just like the two covert attention tasks. We first examine how vRF position changes between the overt and covert attention tasks. vRFs in the focused attention condition were closer to the attended peripheral target than during the fixation task in all ROIs (Figure 2.4). This replicates several prior reports that both single-unit RFs and voxel RFs move toward a single covertly attended target when compared to attending fixation (Anton-Erxleben et al., 2009; Connor et al., 1996, 1997; Klein et al., 2014; Vo et al., 2017; Womelsdorf et al., 2006, 2008).

We then compared vRF position in the divided attention condition to the overt attention condition. This allowed us to test a prediction from an attention field model that attempts to account for shifts in spatial RF position. This model proposes that a Gaussian-shaped attention field combines multiplicatively with each spatial RF to predict the RF's position under covert attention (Klein et al., 2014; Womelsdorf et al., 2006). Under this account, we would predict that our divided attention task would generate a single, large attention field that peaked at the central fixation point and encompassed all 3 target location. When multiplied with the spatial RFs, one would observe that all RFs move towards the center of the attention field. Even RFs that are close to the central fixation point (i.e. foveal voxels) would move towards the center. Alternatively, we might base our predictions off of previously observed data that spatial RFs shift toward the attended target. This would generate the prediction that spatial RFs shift toward all 3 attended locations, such that foveal voxels would move away from the center towards the peripheral targets.

To test these predictions binned the vRFs based on their eccentricity when participants attended to the central fixation point. We then calculated the mean eccentricity of those same voxels when participants directed their attention towards 1 or all of the covert attention locations. We found that vRFs close to fixation moved outward, toward the peripheral targets, in both covert attention conditions (Figure 2.4A). We note that these effects are unlikely to be driven by active modulations in vRF position during the fixation task. A prior study reported that when attentional load is manipulated at the central fixation point, voxel RFs in early visual areas V1 – V3 moved out toward the periphery, away from the fixation point (de Haas et al., 2014). Any movement away from the center that we observe in the divided attention condition would be above and beyond a modulation in the fixation task. However, we cannot assess the extent to which such position shifts might occur in the attend-fixation task, as we do not have a purely 'passive' baseline task. Indeed, it is hard to imagine a truly neutral baseline task in this context, as other factors such as task engagement and difficulty would necessarily be conflated.

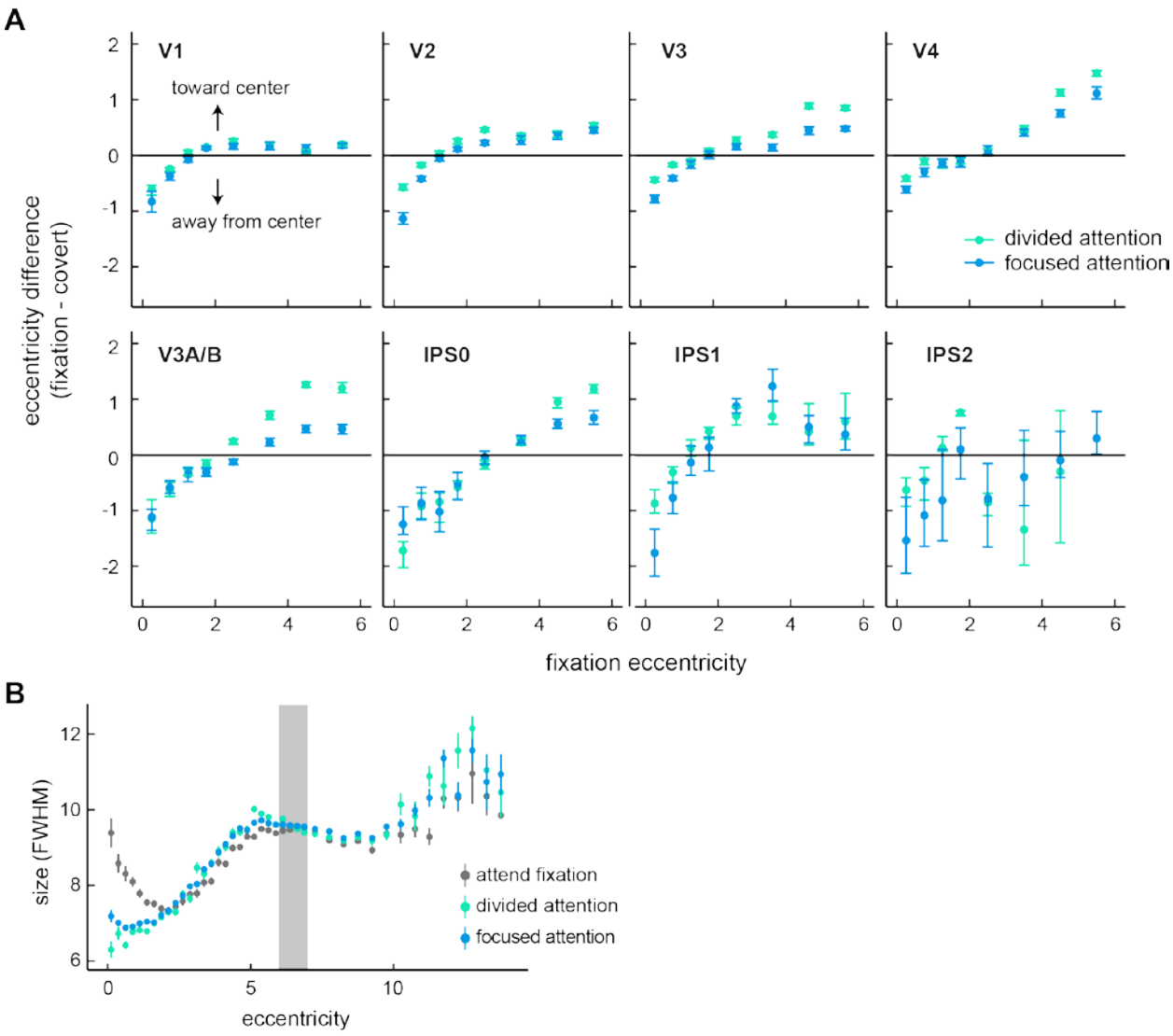


Figure 2.4. A comparison of vRF properties when attending at central fixation point and covertly attending to 1 or 3 peripheral target. Error bars are 95% CIs calculated by bootstrapping. Any error bars which do not overlap indicate a significant difference. A) In both covert attention conditions, vRFs near the central fixation point move outward (centrifugally) toward the peripheral attention targets. B) vRF size as a function of eccentricity in early visual area V1. The fixation task induces a significant increase in vRF size near the central fixation point. The light gray bar represents the diameter of the peripheral attention targets. All attention conditions show size modulations near the peripheral attention stimuli which disrupt the overall linear size-eccentricity trend.

We discuss the remainder of the comparisons between the covert attention tasks and the attend-fixation task in the Supplemental Results. These data require a lengthy discussion because we believe that the fixation task caused active modulations in at least some of the vRF

properties. We find, for example, that early visual vRFs (V1-V4) near the fovea are larger in the fixation task than the covert attention conditions (Figure 2.4C). This is consistent with an earlier report that increasing the cognitive load of a fixation task increases voxel RF size in V1 – V3 (de Haas et al., 2014).

Our primary interest was in investigating how the spatial scope of covert attention affects visual encoding. The first step in examining this question was to document the types of vRF properties which changed by expanding the scope of covert spatial attention. Next, we analyzed which of these vRF modulations might support better spatial discrimination at the attended locations.

Different sets of vRF modulations support higher spatial discriminability

We computed a discriminability metric that describes the ability of a population of RFs to support fine discrimination judgments (Pouget, Dayan, & Zemel, 2003; Seung & Sompolinsky, 1993). The discriminability metric is based on prior theoretical and empirical findings that fine discriminability is highest where the slope of a tuning curve is the highest (Butts & Goldman, 2006; Navalpakkam & Itti, 2007; Pouget et al., 2003; Pouget, Deneve, & Latham, 2001; Regan & Beverley, 1985; Scolari & Serences, 2010). The idea is that larger response differences from a single RF contribute to an increased ability to discriminate two adjacent points (Figure 2.5A; Methods). Note that this is a theoretical value based on the spatial RFs, and is not tied to the behavioral performance of our participants.

To compute the metric, we first take the responses of a given vRF at two spatial positions and compute the difference between the two. Since we were interested in fine spatial discriminability, we computed the response differences along adjacent points in small (0.1°) steps across the mapped visual field. This was repeated for every vRF in an ROI for a given participant. To combine across participants, we concatenated all the vRFs across participants and resampled those vRFs with replacement. Then we averaged the spatial discriminability values for those vRFs in an area around each peripheral target stimulus (Methods).

We then examined the vRFs in the focused attention condition and compared spatial discriminability near the attended target to discriminability near the ignored targets. We predicted that spatial discriminability should be higher near the attended targets. This was tested in two different ways. First, we computed spatial discriminability as described above using the 2D images of each vRF (e.g., Figure 2.1C, top). These vRF images can assume any shape. We found higher spatial discriminability near the attended location than the ignored locations across all ROIs (Figure 2.5B, solid bars; ~25.9% change across ROIs, all $p < .001$). Second, we computed spatial discriminability metric using the isotropic functions fit to each vRF (Figure 2.5B, patterned bars). Using the vRF fits caused an overall drop in discriminability values, compared to the vRF images. This suggests that fitting the vRF discards some informative modulations that are not well-described by a Gaussian-like, isotropic function. However, discriminability computed on the vRF fits still showed the same effect: in all ROIs, discriminability was higher near the attended location (~10% change across ROIs; all $p < .001$). Both these analyses replicate a previous report that spatial discriminability was higher near a single attended location compared to a single ignored location in the opposite hemifield (Vo et al., 2017).

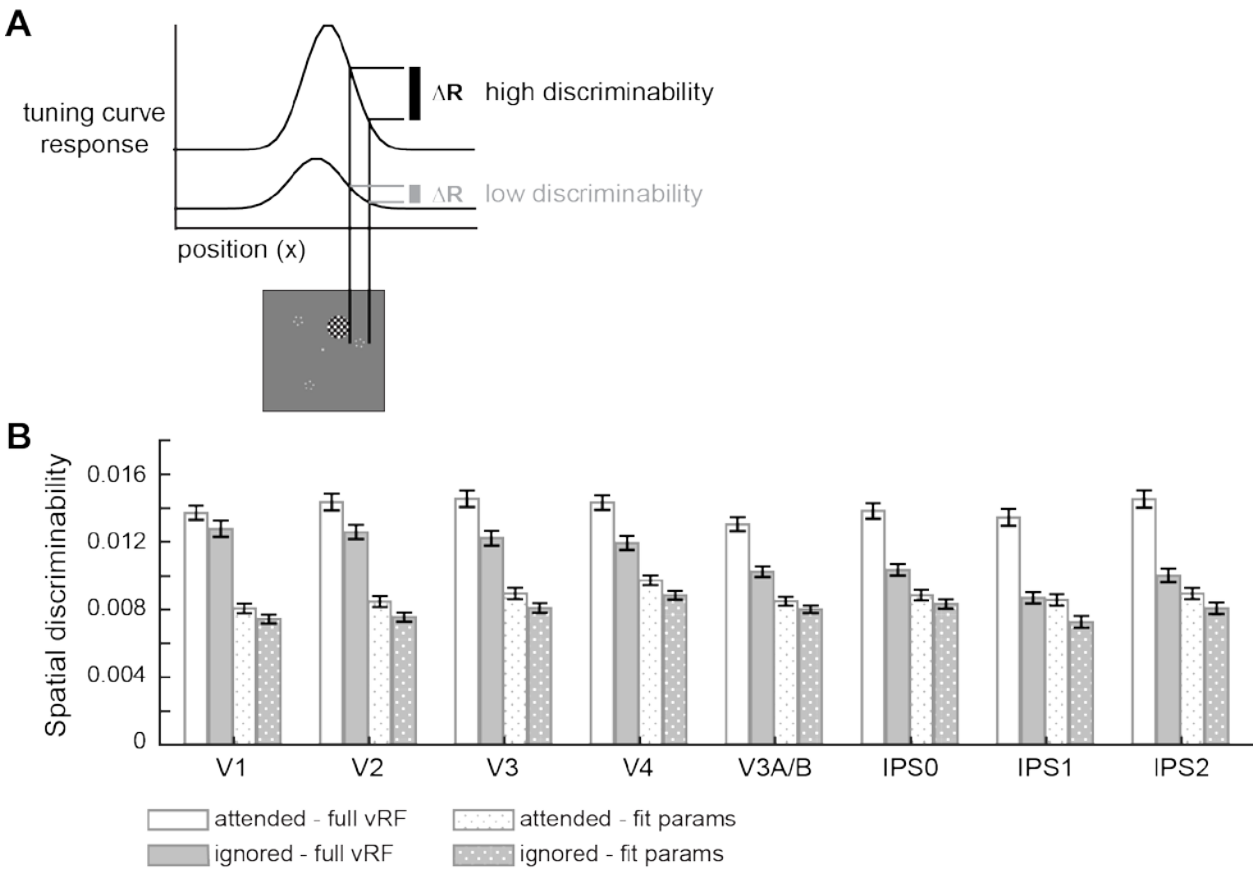


Figure 2.5. Spatial discriminability metric. A) The discriminability metric for a single voxel is computed by taking the difference in a RF's response to two spatial positions. These values are then averaged across voxels (Methods). Error bars are 95% CIs computed by bootstrapping. B) The spatial discriminability of vRFs in the focused condition, either computed using the vRF images or using the best fit 5-parameter functions. Attention increased spatial discriminability at the attended location compared to the ignored locations (all $p < .001$).

The next set of analyses addressed two questions: (1) which vRF modulations in each covert attention condition support high spatial discriminability near the attended locations? and (2) does focused attention rely on different vRF modulations than divided attention to increase discriminability near the attended locations? In order to address how a specific vRF modulation – e.g. a change in position, or an increase in amplitude – affected spatial discriminability, we needed reference vRFs that represent their response properties without covert spatial attention. This reference would allow us to compute the spatial discriminability of the vRFs near the

attended target without the specific modulations induced by covert attention. Then we could modify the reference vRFs by substituting in parameter values observed under each covert attention condition (e.g., the positions observed when spreading attention across all 3 peripheral targets), recompute the spatial discriminability value, and observe whether this significantly changed spatial discriminability near the target. We used the vRFs estimated during the overt attention task at fixation as the reference. Even if the attend-fixation vRFs were modulated by overt attention, they can still serve as a reference to show which response parameters are important for increasing spatial discriminability *near the covertly attended target*. Moreover, it seems likely that any modulations in the attend-fixation task would not affect spatial discriminability near the peripheral target, given that it was a non-spatial task performed at the central fixation point.

There are two possible predictions for these analyses. One prediction is that the same vRF parameters are needed to increase spatial discriminability in both the focused and divided conditions. Previous work has shown that position shifts increase spatial discriminability more than size or amplitude changes in a focused covert attention task (Vo et al., 2017). In the present dataset, we observed widespread shifts in vRF position toward the focused attention target, compared to the other two attention tasks. We also observed widespread position shifts when we compared vRFs between the divided attention task and the fixation task. Therefore, one might predict that position shifts may be the most important modulation in both covert attention tasks. On the other hand, other vRF parameters, such as size or amplitude, may be more important in determining spatial discriminability when the scope of attention is expanded. In this case, the parameter(s) that govern increases in spatial discriminability may differ across covert attention conditions.

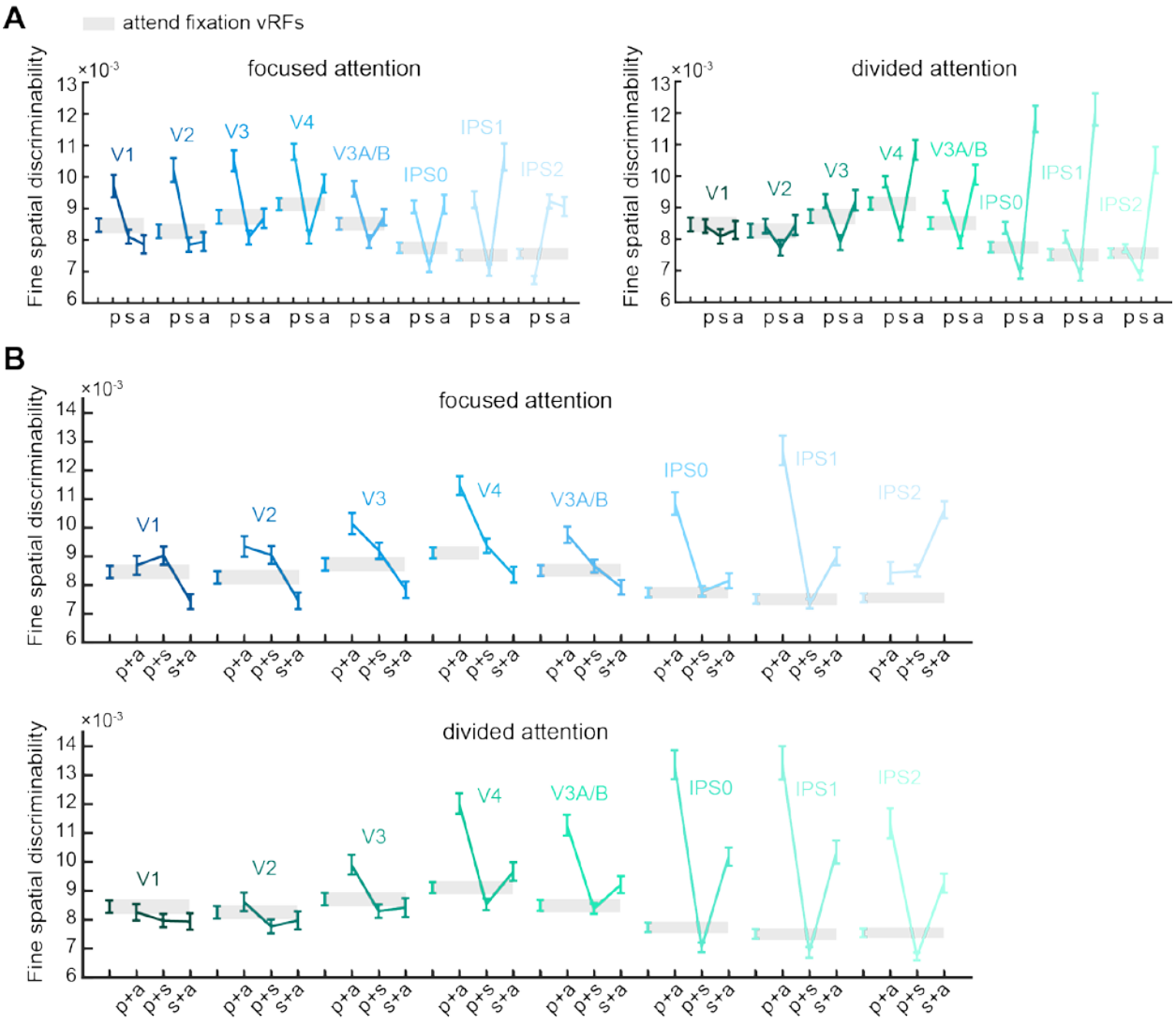


Figure 2.6. The relative importance of each vRF parameter in each covert attention condition in determining spatial discriminability near the attended target. [Key: p = position, s = size, a = amplitude]. Gray horizontal bars represent the reference spatial discriminability values near the peripheral targets during the overt attention task at the central fixation point. Error bars are 95% CIs computed by bootstrapping. Error bars that do not overlap demonstrate a significant difference between parameter manipulations for that ROI and condition. (A) Analyses which consider each the effect of each vRF parameter in isolation. We find that position is most important in the focused condition, but amplitude is important in the divided attention condition. B) Analyses which consider combinations of vRF parameters. Together, changes in both vRF position and amplitude (p + a) result in the highest spatial discriminability across covert attention conditions in nearly every ROI (except focused, IPS2).

First, we altered only one vRF parameter at a time and examined its effect on spatial discriminability (Figure 2.6A). For example, if we altered vRF position, we used that voxel's position from a covert attention task, but kept the voxel's size and amplitude from the fixation

task. Note that a baseline change does not affect the value of the discriminability metric, so it is not included in these modulations. In the focused condition, we found that position shifts increase spatial discriminability the most in areas V1 – IPS0, replicating Vo et al. 2017. However, in IPS1 and IPS2, we found that amplitude led to large increases in spatial discriminability over the fixation task (Figure 2.6A, left). Note that IPS1 & IPS2 were not analyzed in the previous study. In the divided condition, we found that vRF position alone and amplitude alone led to comparable spatial discriminability values in early visual areas V1 – V3. However, the difference in spatial discriminability across parameter manipulations was small. In later visual areas, however, vRF amplitude contributed to large increases in spatial discriminability (Figure 2.6A, right).

Second, we examined the effect of combinations of parameters (e.g., position and size together) on spatial discriminability near the covertly attended location. If vRF amplitude alone is important in the divided attention condition, all parameter combinations with amplitude should lead to higher discriminability. We found that across both covert attention conditions, the combination of position and amplitude changes led to the highest spatial discriminability values in most ROIs (Figure 2.6B). However, with focused covert attention, a combination of size and amplitude in IPS2 vRFs increased spatial discriminability the most. Given the low number of voxels in that ROI we are hesitant to draw strong conclusions about its deviation from the general pattern.

In sum, we found that a combination of position and amplitude changes appears to be a generalized mechanism for increasing spatial discriminability during two covert attention tasks. Yet when the spatial scope of attention is large, shifts in vRF position are not as informative as changes in vRF amplitude.

Discussion

We investigated how expanding the scope of covert spatial attention affects the response properties of voxel receptive fields in human retinotopic cortex. Dividing spatial

attention across 3 locations generally increased vRF size and response gain, compared to focusing spatial attention on a single location (Table 2.1, Figure 2.2B). We also found evidence against models that posit a single Gaussian attention field to explain changes in vRF position: vRFs near the fixation point shifted outward toward the peripheral attention targets, rather than inward toward the center of the attention field (Figure 2.4). We then assessed how different vRF modulations in each covert attention condition contributed to the ability of the population to make fine spatial discriminations. When participants divided their attention across 3 locations, shifts in vRF position were relatively unimportant: instead, changes in vRF amplitude improved spatial discriminability more (Figure 2.6A). Across both covert attention conditions, however, we find that a combination of position shifts and amplitude changes generally supports the highest spatial discriminability near the covertly attended locations.

Our data do not support a model which posits that a large, unitary Gaussian-shaped attention field can account for changes in vRF position. This model generates the prediction that spreading attention across 3 peripheral targets should cause all vRFs to move toward the center, contrary to what we observe (Figure 2.4). This hypothesis necessarily adheres to the view that spatial attention is unitary – i.e. that attention acts as a 'zoom lens' (Eriksen & St James, 1986; N. G. Müller, Bartelt, Donner, Villringer, & Brandt, 2003). However, several behavioral studies have reported attentional enhancements that are more consistent with a split focus of attention (Awh & Pashler, 2000), particularly when there are distractor stimuli inside the large scope of attention. This alternative is known as the 'multiple spotlight' hypothesis. Indeed, there is both neuroimaging and physiological evidence that spatial attention can be split into multiple foci. fMRI and EEG studies have found that portions of retinotopic cortex which correspond to the center of the attentional field, where distractors are placed, do not show any attentional enhancement, in contrast to portions of retinotopic cortex which correspond to the attended target stimuli (McMains & Somers, 2004, 2005; M. M. Müller et al., 2003). In a single-unit electrophysiology study, one group reported that single-unit responses actually decreased

when two attended locations surrounded the spatial RF (Niebergall et al., 2011a). Overall, our data are more consistent with a split focus of attention. However, we cannot rule out other proposals, such as a contiguous annular attention field, rapid switching between locations, or a combination of a Gaussian with active suppression of a distractor at the central fixation point (Cave, Bush, & Taylor, 2010; Cutzu & Tsotsos, 2003; Jans, Peters, & De Weerd, 2010). In our task, participants may have suppressed spatial representations of the central fixation point during the divided attention task, since the luminance changes on each trial may capture exogenous attention. This could cause vRFs to shift away from the central fixation point. However, we are not aware of any evidence that demonstrates changes in spatial vRF position due to distractor suppression.

Another prediction of a multiplicative Gaussian account of spatial RF changes is that with a narrow focus of attention relative to the size of the RF, spatial RFs will shrink (Womelsdorf et al., 2006). However, several neurophysiology studies have shown that spatial RFs can either increase or decrease in size with covert attention. Some have suggested that these effects may be due to the spatial relationship between the RF and the attended location (Anton-Erxleben et al., 2009), or due to an interaction between stimulus eccentricity and non-classical RF properties (Roberts et al., 2007). It is challenging to compare these physiological results to spatial RFs measured using fMRI, which depend on the aggregate responses of retinotopically organized single cells in visual cortex. For example, our analyses which examine vRF size near the attended location likely included a mixture of single cell RFs which entirely contain the attended stimulus, partially overlap with the stimulus, and are simply adjacent to the stimulus. We then observe that in many retinotopic ROIs, vRF size increases near the attended location, compared to the ignored location (Table 2.1). Given a previous observation that spatial attention inside a RF drives decreases in RF size (Anton-Erxleben et al., 2009), it seems likely that we are mostly measuring vRFs which partially overlap with the stimulus or are adjacent to it. However, we do find that in IPS1 & IPS2, the ROIs with the largest spatial RFs, median vRF

size decreases near the attended location compared to the ignored locations. In these parietal ROIs, we may largely be measuring RFs that entirely contain the attended stimulus. The interaction between vRF size and the scope of attention is an interesting effect that deserves further investigation. With a more effectively neutral fixation condition which does not modulate vRF size, one could sort vRFs based on their properties measured during the neutral task and observe how these factors interact with vRF modulations under covert attention.

Task difficulty may have affected the types of vRF modulations we observed. In our task, we interleaved both attention condition and trial difficulty, making it likely that participants were maintaining the same high level of effort throughout the scan. A previous study documented that increasing the cognitive load of an overt attention task at fixation modulates both the size and position of vRFs (de Haas et al., 2014), and we observe similar effects in our data. However, it is unknown how task difficulty affects vRF amplitude during a covert attention task. A previous study which examined how vRF modulations affect spatial discriminability held the task at a fixed level of difficulty (Vo et al., 2017). Compared to that dataset, we find that vRF amplitude is relatively important for spatial discriminability, even in the focused attention condition (IPS0 and above in Figure 2.6A, right). We speculate that this difference may be due to changes in task difficulty. In line with this hypothesis, electrophysiological work has also shown that task difficulty increases the spiking of single-units at the focus of spatial attention, and suppresses responses outside the focus of attention (Chen et al., 2008; Elizabeth Boudreau, Williford, & R Maunsell, 2006). A more detailed comparison of vRF properties under several difficulty levels would provide better evidence to test this hypothesis at the level of single fMRI voxels.

Manipulating the spatial scope of attention affected the pattern of results we saw in the spatial discriminability analysis. In particular, we observed that while changes in vRF position are important in the focused attention task, changes in vRF amplitude are more important in the divided attention task, particularly in higher visual areas. These data suggest that shifting the preferred position of a spatial RF and altering the response gain make distinct, non-overlapping

contributions to the ability of the population to make spatial discriminations at the covertly attended locations. This is particularly interesting in the context of the computational modeling literature which attempts to account for the variety of documented modulations in spatial RFs with covert attention. This work has shown that shifts in the tuning of spatial RFs can be generated by physiologically plausible mechanisms. One possibility is that shifts in tuning are produced by spatially specific patterns of gain modulations in an input layer. This hypothesis has been both informally noted in several places, as well as formally implemented in computational models (Anton-Erxleben et al., 2009; Baruch & Yeshurun, 2014; Compte & Wang, 2006; Miconi & VanRullen, 2016; Treue & Martinez-Trujillo, 2012; Vo et al., 2017). Another possibility is that attentional modulations occur by a combination of additive modulatory inputs and recurrent feedback between layers of processing, either locally or across ROIs (Compte & Wang, 2006; Miconi & VanRullen, 2016). It remains to be seen whether modulations in spatial RF gain with an increase in the spatial scope of covert attention is mediated by a similar mechanism.

We note that the spatial discriminability metric we use here is a theoretical description of a population's ability to discriminate between two spatial positions. Future work could explicitly test whether this metric tracks behavioral performance by collecting enough data at different thresholds to estimate a separate RF. Other methods of assessing the spatial information content of a population may also be useful comparisons to the discriminability metric. For example, one could compare these results to that from an inverted spatial encoding model that generates stimulus representations, or a decoding technique that generates confidence values. The prediction is that the same pattern of vRF modulations should increase the quality of stimulus reconstructions, or the confidence of a spatial decoder.

In sum, we found that increasing the spatial scope of covert attention induces differential patterns of vRF modulations. vRF size, amplitude, and baseline generally increased with a large scope of spatial attention. We also found that changes in vRF amplitude were more important

for spatial discriminability near the attended locations when participants divided their attention across 3 locations. These results provide interesting insights into how spatial attention affects visual encoding properties across retinotopic cortex.

Methods

Data and all analysis code will be posted on osf.io upon publication.

Participants

We recruited 8 human participants from the UC San Diego community (4 female). Each performed a number of behavioral sessions to train each participant and to gather data on their psychophysical thresholds (mean 2.125 sessions). We also collected retinotopic mapping data on 2 participants (anonymized IDs BX & BZ), and used previously collected data in the lab for the remaining participants (IDs BW, AV, BI, BR, BJ, & AR). We collected at least 3 sessions of experimental fMRI data for each participant. All participants provided written informed consent and were compensated for their time ($20/hour for fMRI, $10/hour for behavior) according to our protocol approved by the UC San Diego Institutional Review Board.

Participants completed two main tasks in the scanner: the covert spatial attention task described above, and an additional spatial mapping task to aid in the estimation of voxel receptive fields (vRFs). We used a slightly different checkerboard stimulus setup and trial counterbalancing for the first 2 participants, BW and AV, before adjusting the parameters to be identical for the remainder of the 6 participants (Supplemental Methods). This did not appear to impact the vRF data significantly. However, one of these participants (BW) reported difficulty distinguishing the cue colors in the scanner, and had a significantly higher psychophysical threshold than the other participants (Supplemental Results). We excluded this single participant for a total of N = 7.

Spatial attention task

This two-alternative forced choice task was adapted from prior work (Vo et al., 2017) to present 3 spatial attention targets, rather than only 2. The task itself is designed to require fine

spatial discrimination in addition to allocating attention to particular locations. While participants allocated their spatial attention to one or more targets, a task-irrelevant full-contrast checkerboard stimulus appeared at different locations on the screen (Figure 2.1). In the fMRI scanner, this checkerboard mapping stimulus allowed us to map the spatial selectivity profiles of thousands of voxels across the visual hierarchy.

On each trial of the covert attention task, participants viewed 3 targets (0.5° radius) placed along an isoeccentric ring (6.5° eccentricity) from a central fixation point (Figure 2.1). On every trial, 1 of the 3 targets would rotate clockwise (CW) or counter-clockwise (CCW), and a brief luminance change occurred in the gray central fixation point. In two of the three conditions, participants were asked to report the direction of rotation with a keypress (index finger for CW, middle finger for CCW). In the *divided* attention condition, participants were asked to attend to all 3 targets. In the *focused* attention condition, participants were given a 100% valid cue that indicated which of the 3 targets would rotate on that trial. In the remaining attend *fixation* condition, participants reported the direction of the luminance change (index finger for dimmer, middle finger for brighter). In total, participants needed to remember 5 different attention cues (attend all, attend target 1, attend target 2, attend target 3, and attend fixation).

The cue was indicated by a color change to the central fixation point. The color-condition cue pairings were randomized for each participant. All participants were trained on their cue pairings in psychophysical threshholding sessions outside of the scanner. An instruction screen reminded participants of the color-condition cue pairings before the start of each scan run.

The timing of each trial was as follows. The colored cue was presented for 500 ms, followed by a 1900 ms period where only a neutral gray fixation point was shown. The 3 attention targets then appeared with the mapping stimulus for 2400 ms. On most covert attention trials, the mapping stimulus disappeared and the 3 attention targets remained on the screen for 1600 ms, with the target occurring during a 300 - 1400 ms window post-stimulus offset. To ensure that participants would continue to direct their spatial attention at the cued

target throughout the trial while the mapping stimulus was on the screen, we inserted 10 catch trials in which the target change occurred while the checkerboard stimuli remained on the screen (300 - 1400 ms post-stimulus onset). Data from these trials were not analyzed or used to estimate psychophysical functions. Participants were instructed to respond as soon as they detected the target rotation, which lasted between 100 - 300 ms. Responses were accepted at any point after the rotation until the end of the intertrial interval; average 2400 ms; 5 possible durations uniformly distributed between 800 – 4000 ms).

The checkerboard mapping stimuli (5.12° radius) appeared in randomized order along points in a predetermined triangular grid that contained 24 locations. Across all 4 generated grids, the stimuli spanned a 27.56° diameter circular aperture on the screen. We shifted this grid 4 times for a total of 96 unique mapping locations (Figure 2.1; although see *Participant differences* in Supplemental Methods).

Pre-scan behavioral training

Before the participants entered the MRI, we measured the psychophysical thresholds for each condition of the covert spatial attention task. This allowed us to equate performance across all 3 attention conditions, ensuring that any differences we found between conditions were not due to changes in task difficulty. We also acquired eye-tracking data during these training sessions using an EyeLink 1000 Plus. Participants were required to complete a set number of trials in each condition, with their gaze within 1.5° of fixation. If their gaze shifted beyond this bounding region for 1250 ms (long threshold chosen to allow for blinks), the trial was aborted and participants were given a 5000 ms timeout indicating that they had broken fixation. All aborted trials were repeated at the end of the run. To familiarize participants with the size of the fixation window, they performed two brief training tasks described in the Supplemental Methods.

We introduced elements of the experiment in progressive stages. During the first 90-minute session, participants were informed about all the conditions of the spatial attention task.

However, they did not practice any trials of the fixation task until the last training session. To learn the color-cue pairings, participants were shown a text label underneath the colored cue ('1', '2', '3', or 'all') for the first 1-2 runs of the first session. This trial cue label was removed after the participant confirmed that they had learned the color-cue pairings. However, participants were always reminded of the color-condition cue pairings at the start of each run.

We used the QUEST toolbox to determine the task difficulty from trial-to-trial (Watson & Pelli, 1983) separately for each condition. The task difficulty for the covert attention task was controlled by the magnitude of the target stimulus rotation, where larger rotations led to better discrimination. The difficulty of the fixation task was controlled by the magnitude of the contrast change. For the first 2 runs, the QUEST parameters were set to their defaults so that participants could adjust to the task without influencing the estimation of their psychophysical curve. After this learning period, every trial was used in the QUEST estimation as long as participants maintained fixation within the bounding box.

Every run consisted of 50 total trials, 10 of which were catch trials that did not contribute to the psychophysical function. Participants completed a mean of 10.26 runs of the task with only the focused and divided attention conditions (range 7 – 19 runs). We fit a psychophysical curve (Weibull function) to the data after each session to determine whether their performance was stable and to examine whether they exhibited the expected difference in behavioral threshold between the focused and divided conditions. Participants then completed a mean of 9.29 runs with all 3 attention conditions (range 6 – 14 runs).

Attention task parameters in the scanner

The attention task in the scanner was similar to the previously practiced behavioral task with a few modifications. First, to aid with the estimation of the hemodynamic response function (HRF), we inserted 7 null trials in each run. On null trials, no cue color change occurred, no mapping stimulus was shown, and participants were simply instructed to fixate the central

fixation point. We also reduced the number of catch trials to 5. This yielded a total of 52 trials per run.

Second, we changed the frequency of each trial type to counterbalance across condition and target position. In the focused condition, we predicted based on prior studies that the voxel RF modulations would be dependent on the position of the attention target. Therefore we acquired an equal number of trials for the focused condition at each of the 3 targets as the total number of trials for the divided condition. This was also equivalent to the number of trials for the fixation condition. This yielded a total of 5 pairings of attention condition and target position. For example, if we acquired 192 trials for each focused target position, we would have 960 total trials across all attention conditions and target positions. This allowed us to estimate 5 different spatial RFs for each voxel (focused target 1, focused target 2, focused target 3, divided, and fixation).

Third, we used all the behavioral data we had acquired to maintain a consistent level of task performance throughout the fMRI sessions. At the beginning of each run, we used the last 50 trials from each attention condition to estimate the participant's psychophysical curves for the focused, divided, and fixation tasks. We then obtained the estimated stimulus change needed to achieve 3 different performance levels (60%, 75%, or 90% accuracy) in each condition. These difficulty levels were randomly interleaved into the trials for that run. Rather than exactly equating the difficulty between the two covert attention tasks, we exactly equated visual stimulation by forcing the stimulus change to be the same across tasks within a run. To do this we took the minimum value needed to achieve 60% accuracy on either task, the mean value needed to achieve 75% across both tasks, and maximum value needed to achieve 90% accuracy on either task. All these choices ensured that (1) we always sampled at least 3 different points along the psychophysical curve, allowing us to maintain an accurate ongoing estimate of behavioral threshold throughout the duration of the fMRI scans and (2) their effort was consistent across the fMRI sessions, and approximately equal across tasks.

We acquired 168 – 216 total trials per condition for each participant across 21 – 27 total runs in 3 fMRI sessions. A set of 3 runs contained a counterbalanced set of 120 trials, with one trial for each of the 5 combinations of condition and target position, at each difficulty level, and at each position in a spatial grid that spanned the entire mapped region of the screen. Note that all these numerical values differed slightly for participant AV, who was presented a slightly different stimulus grid with different counterbalancing decisions (Supplemental Methods).

Mapping task

In the scanner, each participant also completed at least 3 runs of an additional task that aided in voxel RF estimation and in an analysis not shown in the main text (see *Spatial reconstructions using an inverted encoding model* in the Supplemental Methods).

Magnetic resonance imaging

Participants viewed the screen through a mirror mounted on the head-coil, which allowed them to see a screen mounted in the scanner bore (21x15 cm, 29 cm viewing distance). We acquired data for both tasks in each 2-hour fMRI session. Participants made their responses using the 2 leftmost keys of a 4-key button box.

We obtained all structural and functional MR images from participants using a GE 3T MR750 scanner at the Center for Functional Magnetic Resonance Imaging (CFMRI) at University of California, San Diego. Unless otherwise specified, all data were collected using a 32-channel head coil (Nova Medical). We acquired an accelerated anatomical using parallel imaging (GE ASSET on a FSPGR T1-weighted sequence; 1x1x1 mm^3 voxel size; 8136 ms TR; 3172 ms TE; 8° flip angle; 172 slices; 1 mm slice gap; 256x192 cm matrix size) in each session for coregistration to the functional data and to a high-resolution anatomical from a separate retinotopic mapping session. We acquired whole-brain functional data using a multiband echo-planar imaging (EPI) protocol (Stanford Simultaneous Multi-Slice sequence) at a multiband factor of 8 (9 axial slices per band, 72 total slices; 2x2x2 mm^3 voxel size; 800 ms TR; 35 ms TE; 35° flip angle; 104x104 cm matrix size). Prior to each functional scan, 16 TRs were acquired as

reference data for image reconstruction. To reconstruct raw k-space data into NIFTI files, we ran custom scripts provided by CFMRI on UCSD servers or our internal lab servers. We also acquired forward and reverse phase encoding blips in each session to estimate the susceptibility off-resonance field (Andersson, Skare, & Ashburner, 2003). This was used to correct EPI signal distortion using FSL *topup* (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012; Smith et al., 2004). The corrected functional images were submitted to further preprocessing stages using BrainVoyager 20.2 and MATLAB R2016b with the NeuroElf toolbox.

In BrainVoyager, we performed slice-time correction, affine motion correction, and temporal high-pass filtering. Then the functional data were coregistered to the same-session anatomical and transformed to Talairach space. This data was then read into MATLAB. Each voxel's timecourse was first *z*-scored within run. We then built a design matrix with individual trial predictors convolved with a double-gamma HRF (peak = 5 s, undershoot peak = 15 s; response undershoot ratio = 6; response dispersion = 1; undershoot dispersion = 1) and a baseline predictor. This allowed us to calculate single-trial beta weights using a general linear model (GLM), which describe the magnitude of the voxel's response to a given stimulus. These beta weights served as input to the voxel RF estimation described below.

Voxel receptive field (vRF) estimation

We used a linear forward encoding model to estimate each vRF. This assumes that the BOLD responses B of each voxel can be described by the weights W on a bank of 2-dimensional spatial filters C, arrayed evenly across horizontal and vertical space (Equation 1). These spatial filters can also be thought of as spatial information channels.

(1) $B = CW$

Here B is a *n* x 1 trial vector, C is a *n* trial x *k* filter matrix, and W is a *k* filter x 1 vector.

To define each of the *k* spatial filters, we created a 2D raised cosine with varying centers and a constant size. We forced the cosine to reach 0 at a set distance from the center, defined

by a size constant s. This ensured that we could estimate a baseline response for each voxel. The cosine was also rescaled to reach a maximum of 1. The equation to define each filter was:

(2) $$f(r) = \left(0.5\left(\cos\left(\frac{r\pi}{s}\right) + 0.5\right)\right)^7 for\ r < s$$

where r is the distance from the filter center and s is the size constant.

In order to maximize our ability to detect small changes in position or shape of the vRFs, we made the bank of filters quite dense and small (1.30° FWHM, spaced 1.04° apart, size constant 3.28). We first created a 24 x 24 array of filters with their centers spanning the mapped screen (out to 14.5° eccentricity). Since we presented our mapping stimuli within a circular aperture, we removed any filters that did not overlap with any mapping stimuli, leaving a total of 516 filters. Each was defined as a vector containing *p* pixels. We chose a resolution of 101 x 101 pixels, which allowed us to resolve details up to 0.3°.

We then created the matrix C by multiplying a pixel-wise mask of the stimulus on each trial (*n* x *p*) with all of the spatial filters (*p* x *k*). This is our design matrix.

We solved for each vRF by regularized regression (Equation 3). The assumption is that each voxel has a sparse set of spatial channels which define its RF. Ridge regression penalizes regressors by forcing most of the weights to be very small, producing an effectively sparse solution (Hoerl & Kennard, 1970).

(3) $$\widehat{W} = (C^T C + \lambda I)^{-1} C^T B$$

The ridge penalty term λ was selected separately for each voxel, following an established procedure (Vo et al., 2017). We iterated through a range of λ values (1000 – 5000) and solved for $\widehat{W}$ at each value using all the available data (after balancing across conditions). We then calculated the Bayesian Information Criterion (BIC) for each solution and selected the λ with the lowest BIC.

Each vRF can be visualized by transforming the weights back into pixel space (i.e. multiplying the pixel-wise spatial filters (*p* x *k*) by the estimated $\widehat{W}$.

vRF fitting

To reduce each estimated vRF to a small set of parameters, we fit the pixel-wise image of each vRF with a smooth, unimodal, isotropic function that has a center, size, amplitude, and baseline. The fit function is identical to Equation 2, except we define an [x,y] center and use this to compute r as the Euclidean distance between the center and each pixel coordinate. Here the fit baseline describes the magnitude of the voxel's response that does not reliably depend on the position of the mapping stimulus – that is, it is spatially unspecific. Since we z-scored the response of each voxel on each run, a value of zero is exactly equal to the mean response of the voxel across an entire run.

The ridge regression produced results with variable scales depending on the strength of the lambda penalty parameter. To standardize the scale across voxels, we found the maximum absolute value of each vRF in the fixation condition and used that to normalize the same voxel across conditions. This meant that the vRF in the fixation condition had a maximum absolute value of 1. The same voxel in the divided condition might have a higher value (e.g., 1.2), which would indicate that its amplitude increased relative to the fixation condition. This enabled us to compare amplitude and baseline directly across condition, as well as across participants.

The best fit parameters for each vRF were found by first performing a grid search over the center and size (x & y coordinates in 1° steps, size in 30 logarithmically spaced points between $10^{0.01}$° and $10^{1.25}$° FWHM). For each grid point, we found the best fit amplitude and baseline by linear regression. The grid point with the best fit (i.e. the lowest error) was then refined by a constrained nonlinear optimization search (MATLAB *fmincon*) within 2 grid points of the best fit. In some cases, the result from the nonlinear optimization routine had a higher root mean squared error (RMSE) than the best grid search fit. We always took the fit with the lower error.

vRF selection

While we estimated vRFs for all voxels in retinotopically defined ROIs, not all of these voxels made it into the subsequent analyses. To ensure that we only analyzed well-estimated vRFs, we required each voxel to pass a series of quality checks.

First, all voxels needed to have a ridge penalty term that was lower than the maximum. This indicated that the regularized solution had not converged. Second, they also needed to successfully predict responses from a held-out run. We performed exhaustive leave-one-run-out cross-validation for each attention condition. Because ridge regression changes the scale of the predicted data, we quantified the quality of the prediction by computing a Pearson correlation coefficient (Huth, Nishimoto, Vu, & Gallant, 2012; Vo et al., 2017). We only kept voxels where the mean correlation across runs and conditions was greater than 0.1 (Table 2.S1; Figure 2.S2). Third, they needed to be described reasonably well by a unimodal, isotropic function (see vRF Fitting). We discarded voxels that had an estimated center beyond the mapped screen, a FWHM larger than 15°, or a fit RMSE of greater than 0.25. Lastly, we removed any voxels with outlier fit parameters. We defined outliers as values beyond 3 standard deviations of the mean, based on the entire population of vRFs across all participants. The number of voxels which survived each threshold in each ROI (pooled across participants) is shown in Table 2.S1.

vRF parameter statistics

We performed all statistical tests on vRF parameters in R 3.5.0 (code available online upon acceptance).

We treated the vRF position somewhat differently than the other parameters (size, amplitude, and baseline). The [x,y] center was transformed into two values: first, the voxel's eccentricity (distance to the central fixation point), and second, the voxel's distance to each of the three attention targets.

We then modeled each voxel's non-positional parameters (size, amplitude, and baseline) as a function of its distance from the attention target and condition. This analysis would help us evaluate which parameters were changing significantly as a function of condition,

and in which ROIs. Because each participant contained a variable number of voxels, we fit a mixed effects model instead of a two-way ANOVA to handle the unbalanced data. This was done using the “lme4” package in R. We used hierarchical model comparison to evaluate the significance of each factor. First, a null model was fit in which we only included a random intercept for each participant. Then we included the continuous factor of distance from the target, since this is known to modulate each parameter (Vo et al. 2017). Finally, we included a fixed effect of condition as well as a random slope for each condition. This entire procedure was repeated for each ROI-parameter combination (e.g., V1 size, V2 size, V3 size, etc.). All p-values generated from this procedure were corrected for multiple comparisons using the false-discovery rate (FDR) for dependent samples (Benjamini & Yekutieli, 2001).

We then tested specific hypotheses about each vRF parameter by comparing the medians between conditions with a non-parametric resampling (or bootstrapping) test. This was done using the “boot” package in R. We resampled with replacement across participants for 10,000 iterations to generate bootstrapped distributions of the voxel parameters. On each iteration, we took all the voxels that met the criteria of the statistical test for each participant in the iteration sample. We then computed the median of that voxel parameter distribution. To compare across conditions, we simply took the difference of the distributions across conditions. We generated 95% confidence intervals for the difference in medians by the percentile. We then assessed statistical significance by computing a two-tailed p-value on this difference distribution with the standard method: find the proportion of values below and above zero, take the smaller of the two, and multiply by two (Davison & Hinkley, 1997). All p-values were FDR-corrected for multiple comparisons across all ROIs and the number of tests for each parameter (e.g., 8 ROIs * 3 size tests = correcting across 24 p-values).

To compare whether vRF position changed significantly between the attention conditions, we evaluated whether a voxel’s distance to the attended location grew smaller or

larger between conditions. The median difference was then tested against 0 as described above.

For the remaining non-positional parameters, we listed hypotheses about how they changed near the attention targets (Table 2.1; Table 2.S2). To determine which voxels were "near" a given target, we used a distance cutoff that scaled with vRF size in each ROI. The cutoffs were based on independent data from previous reports (Klein et al., 2014), and confirmed with manual inspection of vRF distributions across eccentricity for each ROI (V1 & V2: 2°, V3: 2.5°, V4: 3°, V3AB: 4°, IPS0 & IPS1: 4.5°, IPS2: 5°). As stated above, we evaluated statistical significance for each hypothesis based on a resampling test. However, since we are only selecting a portion of voxels for analysis, there was occasionally not enough data to draw a valid conclusion. We required that each parameter distribution required at least 6 unique participants (which could generate at least 462 unique resampled iterations, which could generate a p-value of ~0.004). To limit the effect of extreme sampling bias across participants, we required that these 6 participants contain a minimum of 5 parameter values.

Spatial discriminability analysis

We estimated how well a population of vRFs could make fine position discriminations. This spatial discriminability metric is based on prior theoretical and empirical findings that fine discriminability is highest where the slope of a tuning curve is the highest (Butts & Goldman, 2006; Navalpakkam & Itti, 2007; Pouget et al., 2003, 2001; Regan & Beverley, 1985; Scolari & Serences, 2010). It has been computed similarly in previous work (Vo et al., 2017).

First we computed the squared spatial derivative of each voxel RF. We, and took the discrete derivative in both the x and y directions. This was then squared and summed across the two spatial dimensions. That is, we computed an image of the squared spatial derivative $D = (x_{i+1} - x_i)^2 + \left(y_{j+1} - y_j\right)^2$, where i and j are indices into either dimension.

In Figure 2.6, we report the average spatial discriminability values using the response of all voxels in a 0.5° wide annulus of pixels centered on each attended target location (155 pixels total). The annulus excluded the center of the target location, which contained no information about the direction of the target rotation. To generate 95% confidence intervals in a given ROI, we pooled all voxels across participants and resampled voxels with replacement for 2000 iterations. By pooling the voxels, we could achieve a more even sampling of vRFs across space to avoid any idiosyncratic biases within a single participant.

We repeated this procedure with different populations of vRFs to assess how the attentional modulations we observed affected spatial discriminability. To do so, we computed the metric based on the 2D pixelwise image of the vRF fit surface. We took the vRF fit parameters and generated the corresponding 2D image at a resolution of 0.1° per pixel. Because our primary interest was in spatial discriminability at the covertly attended targets, so we compared the attentional modulations in the divided or focused conditions to the fixation condition. First, we calculated the spatial discriminability of the voxel RFs as observed during the fixation condition. Then we manually 'modulated' these vRFs ourselves by keeping all of the RF parameters the same, except for one. For example, in one manipulation we kept vRF size, amplitude, and baseline the same as the fixation condition, but used the vRF positions observed during the focused attention condition. We then recomputed spatial discriminability for the new population of vRFs, and repeated this for each parameter and each covert attention condition.

Chapter 2 is being prepared for submission. The author list and working title is: Vo, Vy A., Serences, John T. Altering the spatial scope of covert attention induces differential patterns of voxel receptive field modulations. The dissertation author was the primary investigator and author of this manuscript.

Appendix

Supplemental Methods

Participant differences

The first two participants (AV & BW) were presented with a slightly different grid of checkerboard stimuli. Instead of a grid of 24 stimuli, we used 37 stimuli with a 3.73° radius. Rather than shifting the grid horizontally and vertically, we presented the grid at 4 different rotations about the central fixation point (0°, 15°, 30°, 45°). This generated 148 unique grid positions. Each run contained 37 mapping trials (spanning a complete grid across space), 5 catch trials, and 8 null trials for a total of 50 trials.

For these participants, we also collected the same number of trials for the focused condition as the divided or fixation condition. This meant that we had 1/3 of the data for each target in the focused condition. However, after applying the strict voxel thresholds (see main text Methods) to remove noisy or poorly estimated voxels, the data was of similar quality. Therefore, we included participant AV in all our main analyses. As noted in the main text, participant BW was excluded due to variable behavioral performance and very high thresholds (Figure 2.S1).

Fixation training

Before each participant practiced the spatial attention task, they performed two tasks to familiarize them with the eye-tracker. After the initial calibration, we presented them with a central fixation point and plotted their gaze position in real time. Participants were then allowed to explore the screen as they wished to see the real-time feedback. This self-paced demonstration allowed participants to observe the effects of blinks, head movements, and microsaccades on their gaze position.

After this first task, participants completed 20 trials of a simple covert spatial attention task. On each trial, a point on the screen appeared for 2000 ms. Participants were asked to press a button to report a change in the color of the peripheral stimulus, which occurred on every single trial at an unpredictable time between 250 ms and 1500ms (randomly drawn from

a uniform distribution). If their gaze was outside of the fixation region (within 1.5° eccentricity) for 1250 ms, a message was displayed ("You broke fixation! Timeout.") for a period of 3000 ms.

Supplemental Results

Psychophysical function fits

We concatenated each participant's data across all scan sessions. We then binned the trials by the size of the target rotation, for a total of 12 bins. The psychophysical curves shown in Figure 2.S1 are the resulting fits.

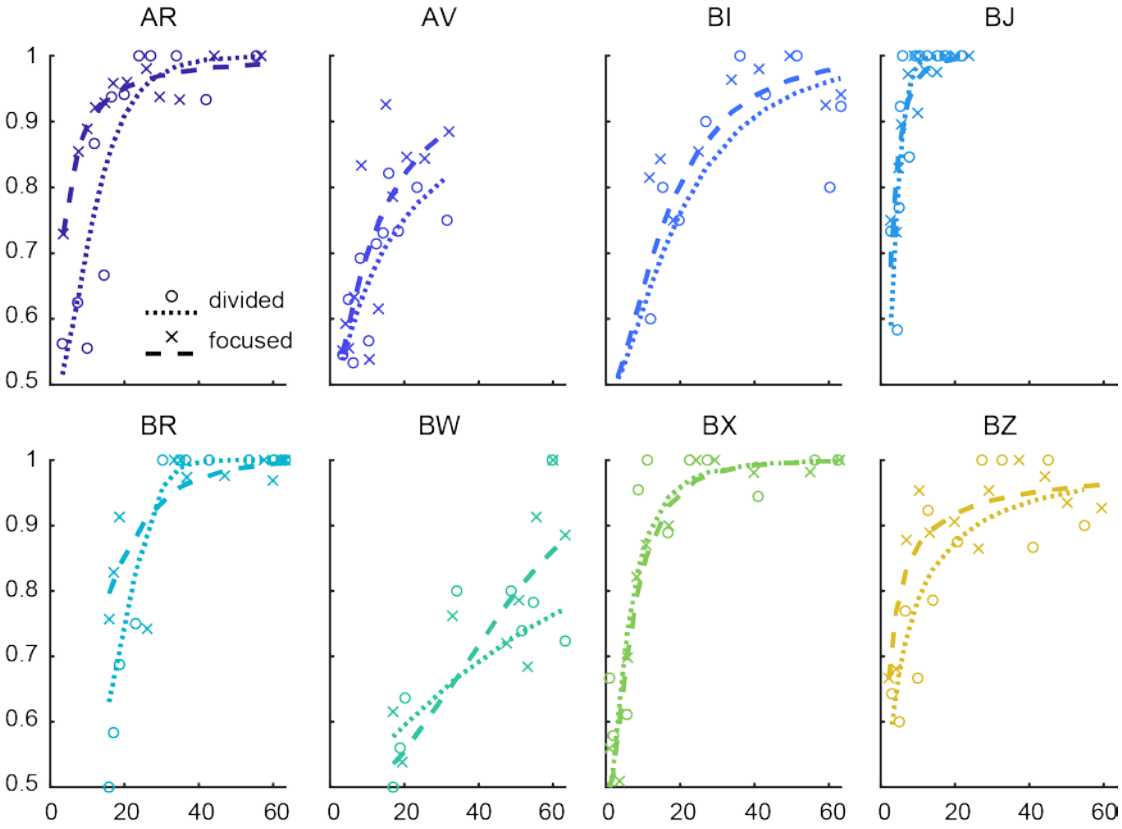


Figure 2.S1. Psychophysical curves fit to all the scanner data for each participant and condition. The excluded participant, BW, had very high estimated thresholds (55.65° & 43.80° for divided and focused, respectively) and inconsistent run-to-run performance.

vRF quality

We first eliminated poorly estimated vRFs by implementing a series of strict thresholds. Table 2.S2 shows the percentage of voxels that survive each threshold.

Table 2.S1. Number of voxels which survive each threshold before inclusion in the vRF analyses. Each column displays the number of voxels that survived that threshold, with the last column representing the total number included in the reported analyses.

ROI	Total measured	Regularizable with ridge	Cross-validation r > 0.1	Fit quality (RMSE & parameter)	Parameter outlier
V1	19654	9650	8077	6280	5597
V2	20023	10363	8635	6931	6256
V3	17874	9713	8159	6862	6136
V4	7013	3559	2827	2756	2534
V3AB	11582	5762	3890	3320	2996
IPS0	9315	3047	1462	1319	1213
IPS1	7573	1695	431	387	358
IPS2	5749	1334	111	107	98
Total	98783	45123	33592	27962	25188

After threshholding the voxels, we ran an analysis to ensure that we could directly compare vRF parameters across condition. Since the data are correlated across ROIs for reasons beyond our control (e.g., common vasculature, correlated responses in adjacent voxels), we did not perform direct statistical comparisons between ROIs. However, we were interested in comparing the relative effects of ROI and condition on vRF quality. We ran a two-way mixed effects model predicting two metrics of vRF quality by ROI and condition. Figure 2.S2 illustrates the estimated means of each factor on both metrics. The first metric is the ability of the estimated vRF to predict a set of held-out runs (held out in sets of 3 runs to ensure the training set is completely counterbalanced). Since the regularization procedure changes the scale of the predicted data, we use the correlation between the predicted and actual data, rather than R^2 (see methods in Huth et al., 2012). The second metric is simply the fit quality of the 5-parameter raised cosine.

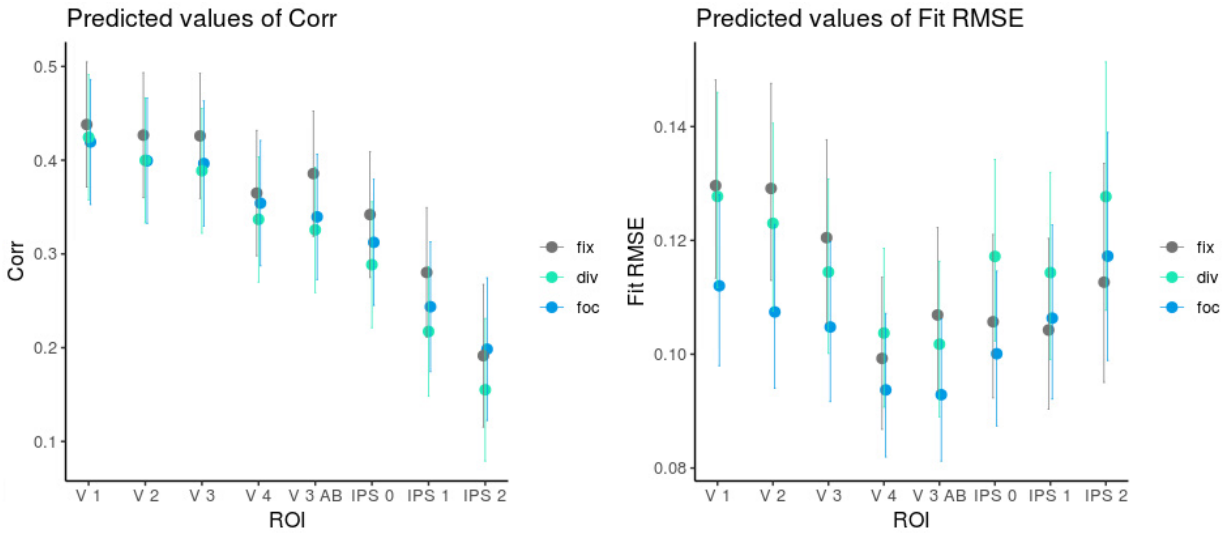


Figure 2.S2. Estimated means for the effect of condition and ROI on two metrics of vRF quality. A) The correlation measure (vRF ability to predict BOLD data from held out runs) shows that the spatial encoding model accounts for less and less of the signal in higher-order visual regions. However, the effect of condition is minor. B) Root mean square error (RMSE) of the isotropic raised cosine (see Methods) compared to the full vRF. While the effect of condition and ROI are now more comparable, their magnitudes is small (an average of ~0.1 units deviation across the whole vRF, where the vRF is normalized to a maximum absolute value of 1).

Results from fixation task

As with the comparison between the covert attention tasks, we listed a set of hypotheses to test about the vRF parameters. In addition to examining the area near the attended location, we also examined a region around the fovea. The results of these tests are in Table 2.S2.

Table 2.S2. Statistical tests of vRF size, amplitude, and baseline against the fixation task. Results highlighted in gray go against the trend observed in most ROIs.

Comparison	*Size*	*Amplitude*	*Baseline*
1. Focused vs. fixation near the attended targets	Foc > Fix: V1, V3, V3AB, IPS0, IPS1	Fix > Foc: V1, V2 Foc > Fix: V4, V3AB, IPS0	Foc > Fix: V1, V2, V3, V4, V3AB, IPS0, IPS1
2. Divided vs. fixation near all targets	Div > Fix: V1, V2, V3, V4, V3AB, IPS0	Fix > Div: V1, V2 Div > Fix: V3, V4, V3AB, IPS0	Div > Fix: V1, V2, V3, V4, V3AB, IPS0, IPS1
3. Focused vs. fixation near ignored targets	Foc > Fix: V1, V3AB, IPS0, IPS1 Fix > Foc: V4	Fix > Foc: V1, V2, V3, V4, V3AB, IPS0, IPS1	Fix > Foc: V1, V2, V4 Foc > Fix: IPS0, IPS2
4. Focused vs. fixation near central fixation point	Fix > Foc: V1, V2, V3, V4 Foc > Fix: V3AB, IPS0, IPS1	Fix > Foc: V2, V3, V3AB Foc > Fix: V4, IPS0, IPS1	Fix > Foc: V1, V2, V3, V4, IPS1
5. Divided vs. fixation near central fixation point	Fix > Div: V1, V2, V3, V4 Div > Fix: V3AB, IPS0, IPS1	Div > Fix: V3, V4, V3AB, IPS0, IPS1	Fix > Div: V1, V2, V3, V4, IPS0, IPS1

The general pattern of results is that covert attention increases vRF size, amplitude, and baseline near the attended locations. Near the ignored targets, vRFs in the fixation condition have higher amplitude than vRFs in the focused condition. These results are in line with previous findings, including those in the main text of this paper. However, we do observe that in areas V1 & V2, vRF amplitude near the attended targets is higher in the overt attention condition compared to either covert attention condition. This is an unexpected effect, as one might expect that the effect of covert attention is to increase the response gain of vRFs near the attended target. Further investigation is needed to understand the cause of this effect.

Near the fixation point, we see an interaction between vRF size and ROI. In early visual ROIs (V1 – V4), vRFs are larger near the fixation point. This is consistent with the pattern of effects observed in an earlier paper which manipulated the cognitive load of a task at fixation (de Haas et al., 2014, Supplement). The pattern of results for vRF amplitude near the fixation point is more complex, and inconsistent in comparisons between the focused condition and

fixation, and the divided condition and fixation. Again, further investigation is needed to pinpoint the cause of these effects.

Finally, we observe fairly consistent results in vRF baseline both near the peripheral locations and near the central fixation point. However, we find in IPS0 & IPS2 that vRF baseline is higher near the ignored targets in the focused condition than in the fixation condition. To our knowledge, there are no available data from prior studies to help interpret these results. Work that investigates spatial RF baselines with covert attention focus on RF properties near the attended location, or computes a difference between the attended and ignored locations.

References

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Anton-Erxleben, K., Stephan, V. M., & Treue, S. (2009). Attention reshapes center-surround receptive field structure in macaque cortical area MT. *Cerebral Cortex*, *19*(10), 2466–78. https://doi.org/10.1093/cercor/bhp002

Awh, E., & Pashler, H. (2000). Evidence for Split Attentional Foci. *Journal of Experimental Psychology*, *26*, 834–846. https://doi.org/10.1037/0096-1523.26.2.834

Baruch, O., & Yeshurun, Y. (2014). Attentional attraction of receptive fields can explain spatial and temporal effects of attention. *Visual Cognition*, *22*(5), 704–736. https://doi.org/10.1080/13506285.2014.911235

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188.

Butts, D. A., & Goldman, M. S. (2006). Tuning Curves, Neuronal Variability, and Sensory Coding. *PLoS Biology*, *4*(4), e92. https://doi.org/10.1371/journal.pbio.0040092

Cave, K. R., Bush, W. S., & Taylor, T. G. G. (2010). Split Attention as Part of a Flexible Attentional System for Complex Scenes: Comment on Jans, Peters, and De Weerd (2010). *Psychological Review*, *117*(2), 685–696. https://doi.org/10.1037/a0019083

Chen, Y., Martinez-Conde, S., Macknik, S. L., Bereshpolova, Y., Swadlow, H. A., & Alonso, J.-M. (2008). Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature Neuroscience*, *11*(8), 974–982. https://doi.org/10.1038/nn.2147

Compte, A., & Wang, X.-J. (2006). Tuning Curve Shift by Attention Modulation in Cortical Neurons: a Computational Study of its Mechanisms. *Cerebral Cortex*, *16*(6), 761–778. https://doi.org/10.1093/cercor/bhj021

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in Area V4 Depend on the Spatial Relationship Between Stimulus and Attention. *Journal of Neurophysiology*, *75*(3), 1306–1308.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *The Journal of Neuroscience*, *17*(9), 3201–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9096154

Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: Psychophysical evidence for a suppressive annulus around an attended item. *Vision Research*, *43*(2), 205–219. https://doi.org/10.1016/S0042-6989(02)00491-1

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.

https://doi.org/10.1016/j.neuron.2008.07.001

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press. https://doi.org/10.1017/CBO9780511802843

de Haas, B., Schwarzkopf, D. S., Anderson, E. J., & Rees, G. (2014). Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. *Current Biology*, *24*(2), R66–R67. https://doi.org/10.1016/j.cub.2013.11.061

Dessing, J. C., Rey, F. P., & Beek, P. J. (2012). Gaze fixation improves the stability of expert juggling. *Experimental Brain Research*, *216*, 635–644. https://doi.org/10.1007/s00221-011-2967-6

Elizabeth Boudreau, C., Williford, T. H., & R Maunsell, J. H. (2006). Effects of Task Difficulty and Target Likelihood in Area V4 of Macaque Monkeys. *Journal of Neurophysiology*, *96*, 2377–2387. https://doi.org/10.1152/jn.01072.2005

Eriksen, C. W., & St James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/BF03211502

Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. J. (2010). When size matters: attention affects performance by contrast or response gain. *Nature Neuroscience*, *13*(12), 1554–1559. https://doi.org/10.1038/nn.2669

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*, 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Itthipuripat, S., Garcia, J. O., Rungratsameetaweemana, N., Sprague, T. C., & Serences, J. T. (2014). Changing the spatial scope of attention alters patterns of neural gain in human cortex. *Journal of Neuroscience*, *34*(1), 112–123. https://doi.org/10.1523/JNEUROSCI.3943-13.2014

Jans, B., Peters, J. C., & De Weerd, P. (2010). Visual Spatial Attention to Multiple Locations at Once: The Jury Is Still Out. *Psychological Review*, *117*(2), 637–682. https://doi.org/10.1037/a0019082

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/J.NEUROIMAGE.2011.09.015

Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology*, *25*(5), 595–600. https://doi.org/10.1016/j.cub.2014.12.050

Kelley, T. A., Serences, J. T., Giesbrecht, B., & Yantis, S. (2008). Cortical Mechanisms for

Shifting and Holding Visuospatial Attention. *Cerebral Cortex*, *18*(1), 114–125. https://doi.org/10.1093/cercor/bhm036

Klein, B. P., Harvey, B. M., & Dumoulin, S. O. (2014). Attraction of Position Preference by Spatial Attention throughout Human Visual Cortex. *Neuron*, 1–11. https://doi.org/10.1016/j.neuron.2014.08.047

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron*, *42*(4), 677–686. https://doi.org/10.1016/S0896-6273(04)00263-6

McMains, S. A., & Somers, D. C. (2005). Processing Efficiency of Divided Spatial Attention Mechanisms in Human Visual Cortex. *Journal of Neuroscience*, *25*(41), 9444–9448. https://doi.org/10.1523/jneurosci.2647-05.2005

Miconi, T., & VanRullen, R. (2016). A Feedback Model of Attention Explains the Diverse Effects of Attention on Neural Firing Rates and Receptive Field Structure. *PLOS Computational Biology*, *12*(2), e1004770. https://doi.org/10.1371/journal.pcbi.1004770

Moore, T., Armstrong, K. M., & Fallah, M. (2003). Visuomotor Origins of Covert Spatial Attention. *Neuron*, *40*(4), 671–683. https://doi.org/10.1016/S0896-6273(03)00716-5

Moore, T., & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology*, *91*(1), 152–162. https://doi.org/10.1152/jn.00741.2002

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*(6946), 309–312. https://doi.org/10.1038/nature01812

Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A Physiological Correlate of the Zoom Lens of Visual Attention. *Journal of Neuroscience*, *23*(9), 3561–3565. Retrieved from http://www.jneurosci.org/content/jneuro/23/9/3561.full.pdf

Navalpakkam, V., & Itti, L. (2007). Search Goal Tunes Visual Features Optimally. *Neuron*, *53*(4), 605–617. https://doi.org/10.1016/j.neuron.2007.01.018

Niebergall, R., Khayat, P. S., Treue, S., & Martinez-Trujillo, J. C. (2011a). Expansion of MT neurons excitatory receptive fields during covert attentive tracking. *The Journal of Neuroscience*, *31*(43), 15499–510. https://doi.org/10.1523/JNEUROSCI.2822-11.2011

Niebergall, R., Khayat, P. S., Treue, S., & Martinez-Trujillo, J. C. (2011b). Multifocal attention filters targets from distracters within and beyond primate MT neurons' receptive field boundaries. *Neuron*, *72*(6), 1067–79. https://doi.org/10.1016/j.neuron.2011.10.013

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*, 381–410. https://doi.org/10.1146/annurev.neuro.26.041002.131112

Pouget, A., Deneve, S., & Latham, P. E. (2001). The relevance of Fisher Information for theories of cortical computation and attention. In J. Braun, C. Koch, & J. L. Davis (Eds.), *Visual Attention and Cortical Circuits* (pp. 265–284). The MIT Press.

Puckett, X. A. M., & DeYoe, E. A. (2015). The Attentional Field Revealed by Single-Voxel Modeling of fMRI Time Courses. *Journal of Neuroscience*, *35*(12), 5030–5042. https://doi.org/10.1523/JNEUROSCI.3754-14.2015

Regan, D., & Beverley, K. I. (1985). Postadaptation orientation discrimination. *Journal of the Optical Society of America*, *2*(2), 147–155.

Roberts, M., Delicato, L. S., Herrero, J., Gieselmann, M. a, & Thiele, A. (2007). Attention alters spatial integration in macaque V1 in an eccentricity-dependent manner. *Nature Neuroscience*, *10*(11), 1483–1491. https://doi.org/10.1038/nn1967

Scolari, M., & Serences, J. T. (2010). Basing Perceptual Decisions on the Most Informative Sensory Neurons. *Journal of Neurophysiology*, *104*(4), 2266–2273. https://doi.org/10.1152/jn.00273.2010

Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences*, *90*(22), 10749–10753. https://doi.org/10.1073/pnas.90.22.10749

Sheremata, S., & Silver, M. A. (2015). Hemisphere-Dependent Attentional Modulation of Human Parietal Visual Field Representations. *Journal of Neuroscience*, *35*(2), 508–517. https://doi.org/10.1523/JNEUROSCI.2378-14.2015

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219. https://doi.org/10.1016/J.NEUROIMAGE.2004.07.051

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–87. https://doi.org/10.1038/nn.3574

Treue, S., & Martinez-Trujillo, J. C. (2012). The spotlight of attention: shifting, resizing and splitting receptive fields when processing visual motion. *E-Neuroforum*, *3*, 74–7i9. https://doi.org/10.1007/s13295-012-0034-9

Umiltà, C., Riggio, L., Dascola, I., & Rizzolatti, G. (1991). Differential effects of central and peripheral cues on the reorienting of spatial attention. *European Journal of Cognitive Psychology*, *3*(2). https://doi.org/10.1080/09541449108406228

van Es, D. M., Theeuwes, J., & Knapen, T. (2018). Spatial sampling in human visual cortex is modulated by both spatial and feature-based attention. *ELife*, *7*. https://doi.org/10.1101/147223

Vo, V. A., Sprague, T. C., & Serences, J. T. (2017). Spatial Tuning Shifts Increase the

Discriminability and Fidelity of Population Codes in Visual Cortex. *The Journal of Neuroscience*, *37*(12), 3386–3401. https://doi.org/10.1523/JNEUROSCI.3484-16.2017

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/BF03202828

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., & Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nature Neuroscience*, *9*(9), 1156–60. https://doi.org/10.1038/nn1748

Womelsdorf, T., Anton-Erxleben, K., & Treue, S. (2008). Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. *The Journal of Neuroscience*, *28*(36), 8934–44. https://doi.org/10.1523/JNEUROSCI.4030-07.2008

CHAPTER 3: Long-term and short-term memory representations in retinotopic visual cortex: cortical reinstatement and source information

Abstract

Current theories propose that the short-term retention of information in working memory (WM) and the recall of information from long-term memory (LTM) are supported by partly overlapping neural mechanisms. For instance, both are thought to rely on reinstating patterns of sensory activity associated with the initial perception of the remembered item. Although reinstatement has been demonstrated for both WM and LTM, it is not clear if these reinstated representations share a similar structure or if the cortical response patterns are uniquely tagged based on the origin of the sensory representation from WM or LTM. Here, we designed a behavioral paradigm that allowed us to directly compare reinstatement during WM and LTM while carefully matching the stimuli, timing, and behavioral recall precision. We then used fMRI and an inverted encoding model to show that visual memories were reinstated and represented to a similar degree across both tasks in all retinotopically organized regions of visual cortex that we analyzed. This suggests that the strength and anatomical distribution of cortical reinstatement in sensory regions is largely similar between WM and LTM. Second, we found that a support vector machine could successfully distinguish between WM trials and LTM trials based on activation patterns in retinotopically organized parietal regions, but not early visual areas, despite the well-matched stimuli and similar recall precision across tasks. We propose that these retinotopically organized parietal areas may be 'multiplexing' information for both the sensory content and the type of memory.

Introduction

It is well-established that performance on a working memory (WM) or a long-term memory (LTM) task relies on the partial reactivation, or reinstatement, of patterns of cortical activity that were present during the initial perception of the remembered item (for reviews, see D'Esposito & Postle, 2015; Serences, 2016; Xue, 2018). For example, viewing a blue couch elicits specific patterns of activity in several cortical areas, especially in visual regions selective for color and complex objects. Then, when trying to maintain the colored object in WM, or

retrieve it from a LTM formed the previous week, patterns of neural activity that resemble those at encoding are reinstated. That said, evidence for item-specific cortical reinstatement has typically been assessed separately in studies of LTM (Bosch, Jehee, Fernandez, & Doeller, 2014; Favila, Samide, Sweigart, & Kuhl, 2018; Kuhl & Chun, 2014; Ritchey, Wing, LaBar, & Cabeza, 2013; Xiao et al., 2017) and studies of WM (Emrich, Riggall, Larocque, & Postle, 2013; Ester, Anderson, Serences, & Awh, 2013; Ester, Sutterer, Serences, & Awh, 2016; Harrison & Tong, 2009; Lee, Kravitz, & Baker, 2013; Riggall & Postle, 2012; Serences, Ester, Vogel, Awh, & Serences, 2009; Sprague, Ester, & Serences, 2016). Thus, it is unknown whether LTM reinstatement and WM reinstatement operate via similar neural mechanisms, and whether reinstated patterns carry information about the source of the memory representation.

We had two motivations to design a study that directly compares reinstatement during WM and LTM tasks. First, several prominent theories suggest that both WM and LTM retrieval rely on shared neural substrates (Atkinson & Shriffin, 1968; Cowan, 1995; Kosslyn, 1980). Yet there have been few direct tests of this hypothesis, perhaps because it is challenging to match the WM and LTM tasks on relevant dimensions such as overall task performance. As a result, it is unclear if reinstated memories are represented in the same format, or if reinstated representations always carry some distinct information about the type of memory (e.g. maintained from a recent percept, as in a WM task, or recalled from LTM). Second, recent studies have reported that lateral parietal regions contain more feature-specific information during long-term memory tasks than during perceptual tasks (Favila et al., 2018; Xiao et al., 2017). This raises the possibility that some parts of visually responsive cortex may represent and reinstate feature-specific information to different degrees. A direct comparison of both memory tasks would enable us to investigate the relative strength of sensory reinstatement in different regions during WM and LTM. To investigate these issues, we designed a functional magnetic resonance imaging (fMRI) study to assess reinstated memory representations in 8 retinotopically-defined regions from V1 through dorsal parietal region IPS2. Participants viewed

a spatial position and held it in WM, or recalled the spatial position from a LTM cue (Figure 3.1). We carefully matched the performance and visual properties across memory tasks.

We first used an encoding model to quantify the fidelity of cortical reinstatement in both tasks. We found that spatial representations maintained in WM and retrieved from LTM were represented with a similar fidelity across all regions of interest (ROIs). Then we used a linear decoder to further investigate how WM and LTM representations might differ in retinotopic visual cortex. We found that while activation patterns from early visual cortex could not distinguish between memory tasks, data from retinotopic parietal regions successfully predicted the task type. Our findings suggest that the degree and the anatomical distribution of cortical reinstatement is largely similar between WM and LTM, providing evidence that memory recall in both tasks relies on similar sensory representations and neural substrates. However, parietal regions appear to multiplex sensory information from memories with information about the type of memory.

Methods

Participants

We trained 12 human participants (8 female) on both the LTM and the WM task. Two participants dropped out of the study during behavioral training. This left 10 participants (6 female) from which we collected both behavioral and fMRI data, 2 of which were co-authors on this paper. Six of these participants (participant IDs BI, BJ, BO, and BC) had previously completed a set of retinotopic mapping scans in the lab for other studies (M. M. Henderson & Serences, 2019; Rademaker, Chunharas, & Serences, 2018; unpublished data). For the remaining 4 participants, we collected new retinotopy data to define each region of interest (see *Retinotopic mapping details*). All participants provided written informed consent and were compensated for their time ($20/hour for fMRI, $10/hour for behavior) as approved by the local UC San Diego Institutional Review Board.

Pre-scan behavioral training

We trained participants to form long-term memory pairings between 24 unique clip art items and 24 unique spatial positions (Figure 3.1). After one day of LTM study and retrieval, we continued LTM training while also measuring their performance on an analogous working memory task. In both tasks, participants were asked to report the remembered spatial positions as precisely as possible. To minimize differences in effort and difficulty of recall between the two memory tasks at the time of scanning, participants were trained until their error on the LTM task was stable.

For each participant, the 24 spatial positions were selected randomly from 24 bins evenly spaced along an isoeccentric ring 3.9° visual angle from a central fixation point. As a result, the set of locations was roughly uniformly distributed around a circle for each participant, and each participant had a slightly different set of positions. Each spatial position was rendered as a dot subtending 0.2° visual angle, or 1.4° polar angle. The clip art items (diameter 1.5° visual angle) were filled outlines of real-world objects taken from a bank of 48 images that all subtended the same area of retinal space (Figure 3.1). These stimuli have been used in previous LTM studies (Sutterer & Awh, 2015).

The LTM study-retrieval task on the first day of training consisted of 6 alternating runs of a study phase and a retrieval phase. During the study phase, the participant was shown each clip art item at the center of the screen and its associated spatial position in a randomized order. This study phase was self-paced – while each pairing was always presented for 500 milliseconds (ms) followed by at least a 1000 ms blank period, participants chose when to advance to the next item by hitting the space bar. During the retrieval phase of the task, each clip art item was presented at the center in a randomized order, and participants were asked to report the associated spatial position as precisely as possible by clicking on a part of the visible isoeccentric ring. After each retrieval trial, participants received two forms of feedback about their performance: the true position was shown in dark gray on the ring, and the error was

shown as a signed integer at the center. This number indicated how far their report deviated from the true position (+ if report was too far clockwise, - for counter-clockwise).

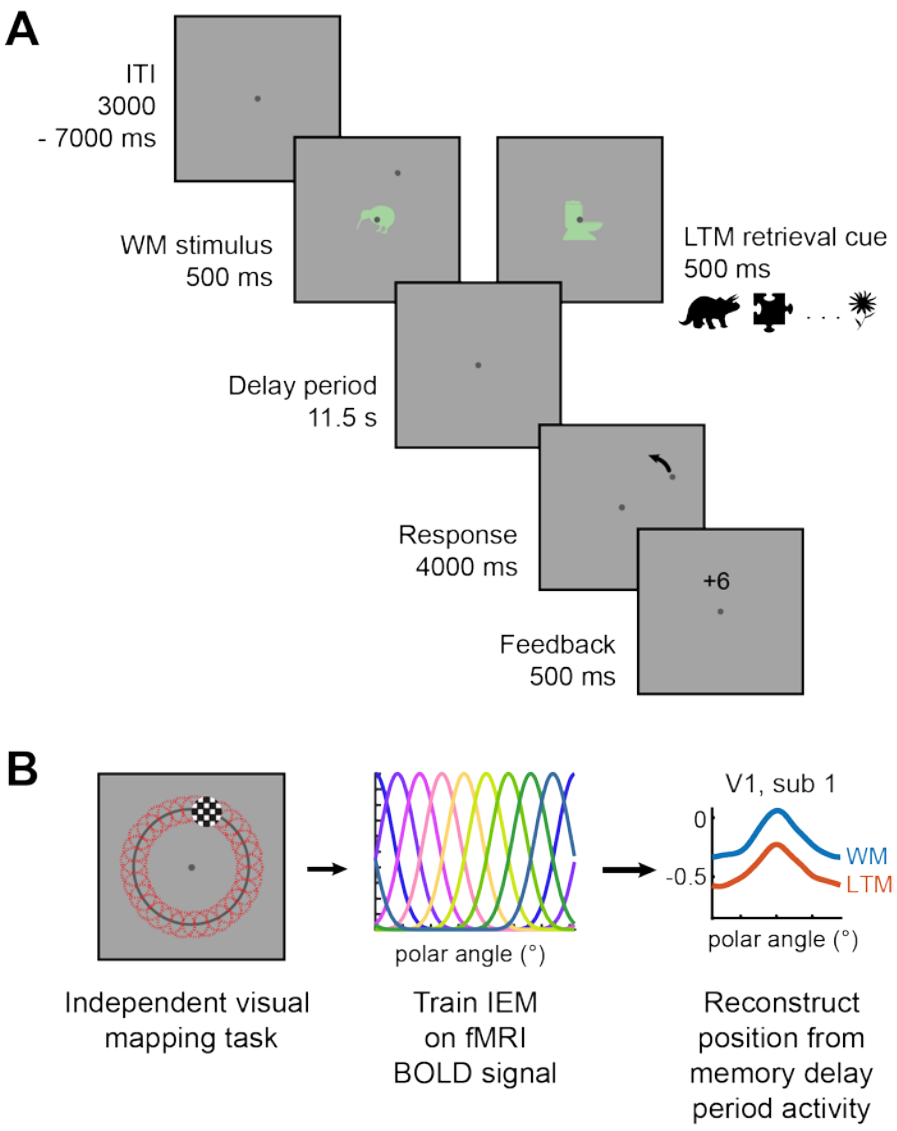


Figure 3.1. Memory tasks and mapping task presented to participants in the scanner. (A) The two memory tasks were identical except for the memory cue – in the WM task, participants saw a dot indicating the position along an annulus, along with an irrelevant clip art item; in the LTM task, participants instead saw a previously studied clip art item that they had learned to associate with a spatial position during behavioral training. After an 11.5s delay period, participants reported the remembered spatial position by rotating a randomly placed dot to the correct position. (B) An independent task was used to map the position selectivity of voxels in retinotopically defined regions of visual and parietal cortex. Participants detected an occasional dimming of a checkerboard stimulus that appeared at randomly ordered locations along a ring. This independent task was used to train the IEM, which was then tested on the memory tasks.

On all remaining days of pre-scan training, participants practiced both a LTM retrieval task and a similar WM maintenance task (similar to Figure 3.1A, with different timing). The order of the tasks was counterbalanced across participants. In the LTM retrieval task, the participant was shown one of the 24 clip art items at the center (500 ms). After a delay (1000 ms), they were presented with the response probe: a dot appeared at a random position along the invisible isoeccentric ring, and participants used four keyboard keys to rotate this dot along the ring until it matched the item's associated position as precisely as possible. The two outer buttons moved the dot along the ring at a fast rate, whereas the two inner buttons moved it along a slower rate. This allowed participants to make fast but precise responses. After the 4000 ms response period, participants received numerical feedback on their response error (Figure 3.1). The intertrial interval varied between 200 and 500 ms. Participants repeated this set of 24 trials per run, for a total of 6 runs in the LTM task block.

The WM maintenance task was nearly identical to the LTM task in timing and presentation. The one exception was that the initial memory array displayed a dot at a random spatial position on the ring, in addition to a clip art item presented at the center that was randomly selected from the remaining set of 24 images. Participants were asked to remember the position of the dot during the 1 second delay, and to use the keyboard keys to rotate the response dot to the position of the memorandum (Figure 3.1). The clip art item was a task-irrelevant sensory control, and there was no predictive relationship between WM position and clip art item. As with the LTM task, a set of 24 trials was repeated 6 times.

All participants performed at least 6 runs of each task on each training day. For slight deviations from this training schedule, see *Additional participant notes*.

Behavioral tasks performed in fMRI scanner

Participants performed both LTM and WM tasks in the scanner (Figure 3.1A). While these were very similar to the behavioral training task, they differed in several key respects: (1) instead of a 1000 ms delay between the cue and the response, we inserted a 11,500 ms delay

to accommodate the slow hemodynamic response; (2) LTM and WM runs were alternated instead of blocked, in the order preserved from the training; (3) participants viewed the experiment through a head-coil mounted mirror pointing at a gray rectangular screen (120x90 cm) at the foot of the scanner bore (~3.85 m viewing distance); (4) responses were made via a 4-key button box; and (5) the intertrial intervals were longer (uniformly distributed in 200 ms steps between 3000 and 7000 ms), to aid in estimating the stimulus-evoked BOLD response.

In addition to the memory tasks, participants also performed a mapping task which we used to train the spatial encoding model (see below) and to localize voxels responsive to the isoeccentric ring (as before, 3.9° from central fixation point). In this task, participants fixated at a central point as a flickering stimulus (6 Hz) appeared at some location on the screen for 3000 ms. They monitored the flickering stimulus for an occasional contrast change (8 of 54 total trials), and responded whether it was an increment or decrement on those trials (pedestal at 70% Michelson contrast). On most trials, the stimulus was a circular checkerboard (0.9° radius, 1.36 cycles/deg) centered at a point on the isoeccentric ring. Each location was pseudorandomly drawn from one of 30 evenly spaced polar bins. On 5 trials, participants instead saw a ring aperture (1° wide) with a similar flickering checkerboard pattern (3.9° from center). On the target trials, either the entire circular checkerboard changed contrast, or a 0.681° (36° polar angle) wide area of the ring changed contrast for 500 ms. We also included 6 null trials to aid in the estimation of the stimulus-evoked BOLD response. Trial order was randomized, with inter-trial intervals (ITIs) varying between 2000 and 5000 ms (mean 3,500 ms). The magnitude of the contrast change or the rotation was manually adjusted by the experimenter on each run to keep task performance for each participant near 75% (participant averaged mean and bootstrapped 95% CIs: 71.8% accuracy [61.12%, 81.28%]; Michelson contrast 0.50 [0.43, 0.57]).

Behavioral data analysis

To describe participant performance on each memory task, we examined each participant's trial-by-trial response error. This is the signed difference between the true location of the memorandum and the participant's response (between 0 and ±180°). Previous findings have shown that a histogram of response errors can be described as a mixture of a uniform distribution, representing guesses, and a circular Gaussian distribution centered near the correct response (Zhang & Luck, 2008). The two parameters of the circular Gaussian describe the mean response (μ), which describes any systematic bias, and the standard deviation of the responses (*SD*), which describe the variability of the memory reports. These parameters were fit using publicly available code (Schneegans & Bays, 2016).

Statistical analysis and confidence intervals

All confidence intervals (CIs) and statistical analyses reported in the paper were generated by bootstrapping. Unless otherwise noted, the bootstrapping was done by resampling with replacement across the 10 participants for 1000 iterations. Then the mean of the value of interest (e.g., response error) was computed on each resampling iteration to generate a distribution of means for the sampled population. The 95% CIs were simply the 2.5 and 97.5 percentiles of this distribution (Davison & Hinkley, 1997).

From the bootstrapped distributions, we computed p-values (two-tailed, unless otherwise noted) by a standard approach (Davison & Hinkley, 1997). First we found the proportion of values above and below the critical value, then we took the smaller of them and multiply by 2. If we repeated any statistical comparison across ROIs or tasks, we corrected across all tests using the false discovery rate procedure (q = 0.05 unless otherwise noted; Benjamini & Yekutieli, 2001).

Magnetic resonance imaging

We obtained all structural and functional MR images from participants using a GE 3T MR750 scanner at University of California, San Diego. We collected all functional images (19.2 cm^2 FOV, 64 x 64 acquisition matrix, 35 interleaved slices, 3 mm^3 voxels with 0 mm slice gap,

263 volumes per memory run and 179 volumes per mapping run) using a gradient echo planar pulse sequence (2000 ms TR, 30 ms TE, 90° flip angle) and a 32-channel head coil (Nova Medical, Wilmington, MA). Five dummy scans preceded each functional run. A high-resolution structural image was acquired at the end of each session using a FSPGR T1-weighted pulse sequence (25.6 cm^2 FOV, 256 x 192 acquisition matrix, 8136/3172 ms TR/TE, 9° flip angle, 1 mm^3 voxels, 172 volumes). All functional scans were co-registered to the anatomical images acquired during the same session, and this anatomical was in turn co-registered to the anatomical acquired during the retinotopy scan.

EPI images were unwarped with a custom script from UCSD's Center for Functional Magnetic Resonance Imaging that calls the FSL PRELUDE & FUGUE functions. All subsequent preprocessing was performed in BrainVoyager 2.6.1, including slice-time correction, affine motion correction, and temporal high-pass filtering to remove slow signal drifts over the course of each session. Data were then transformed into Talairach space and resampled to a 3x3x3 mm voxel size. Finally, the BOLD signal in each voxel was Z-transformed on a scan-by-scan basis. All subsequent analyses were performed in MATLAB using custom scripts (available online at osf.io upon publication).

For all region-based analyses, we applied a localizer mask to the retinotopically-defined occipital and parietal regions before extracting each voxel's timecourse. This mask was generated by defining different conditions for the circular checkerboard and ring aperture checkerboard trials during the mapping task. We estimated BOLD response activation to each condition by convolving the trial events with a canonical two-gamma HRF (peak at 5 s, undershoot peak at 15 s, response undershoot ratio 6, response dispersion 1, undershoot dispersion 1). We then solved a generalized linear model (GLM) on all runs of this mapping task for a single participant. This produced a statistical parametric map of voxels with a significant BOLD response change attributable to each condition (FDR q = 0.05, Benjamini and Yekutieli,

2001). The localizer mask only included voxels that responded with a significant BOLD response increase during the ring aperture condition.

For the encoding and decoding analyses, we balanced the data by equating the number of runs for each memory task. While we tried to acquire an equal number of WM and LTM runs for each participant, some data was discarded due to excessive movement (>3 mm in one run/scan) and required rebalancing.

Inverted encoding model for spatial position

Following previously published methods, we estimated an encoding model for spatial position using z-scored BOLD responses obtained from the independent mapping task (Brouwer & Heeger, 2009; Sprague & Serences, 2013). This allowed us to estimate the response of each voxel to any arbitrary spatial position within the mapped visual field.

We defined 9 evenly spaced spatial channels, or basis functions, along the polar angle axis (Figure 3.1). This was the only axis we varied in our mapping runs. This formed our basis set S (*c* channels x *p* points along the polar axis, here 9 x 360). For every training trial, we used the basis set to specify the predicted [modeled] activation on each of the *c* channels for the presented stimulus, resulting in a design matrix C (*c* channels x *n* trials).

The forward encoding model is then described in Equation 1, where B = BOLD responses (*v* voxels x *n* trials). As long as *n* > *c*, it is possible to solve for W using the pseudoinverse (Equation 2). We solved Equation 2 using *mldivide* in MATLAB.

(1) $B = WC$

(2) $W = (B^T B)^{-1} B^T C$

(3) $C_2 = (W^T W)^{-1} W^T B_2$

After training the model by solving for W, we inverted the encoding model (Equation 3) such that we could predict the spatial position of a stimulus given the known responses on a separate dataset and the independently trained model from the same participant. Here, we used the BOLD responses of voxels during the memory delay of either the WM or LTM task (Ester,

Sprague, & Serences, 2015; Sprague, Ester, & Serences, 2014). Notably, we are attempting to recover spatially-specific information in the patterns of voxel responses during each of two memory tasks when *no visual stimulus is shown on the screen*. Furthermore, we are using a 'fixed' encoding model estimated using a separate dataset and task, enabling us to directly compare reconstructed representations across memory tasks (Sprague, Adam, et al., 2018).

We estimated channel responses C_2 from two different types of BOLD data B_2. In Figure 3.3, B_2 consisted of the BOLD response at each TR (after motion-correction and z-scoring within each run), for a total of 8 TRs. In other analyses described below, we averaged the data across TRs for use in the inverted encoding model.

To create a continuous reconstruction over polar space, we multiply C_2 by the basis set S. We recenter all of the reconstructions such that the memory stimulus is at 0°, and average across all trials. This is the reconstruction plotted in Figure 3.3A.

Representational fidelity metric for spatial reconstructions

To quantify the reconstructions, we compute a metric which describes the fidelity of spatial reconstruction. This is a modified version of previously published metrics (Sprague et al., 2016; Wolff, Jochim, Akyurek, & Stokes, 2017). Here we consider each point of the stimulus reconstruction in polar coordinates as $[r_n, \theta_n]$, where r is the value of the reconstruction in arbitrary units (y-axis in Figure 3.3A) and θ is the polar angle (x-axis in Figure 3.3A). The representational fidelity is directly proportional to the average directional energy of the reconstruction pointing in the expected direction (0°). We computed this by finding the mean vector for across all n points using two metrics of circular data, the average preferred direction and the dispersion (Jammalamadaka & SenGupta, 2001). Since these metrics only exhibit the desired behavior when r is positive, and because a baseline offset has no theoretical impact on the fidelity of a representation, we forced the minimum of the reconstruction to be 0 by adding an offset to all reconstructions.

First we found the mean direction of the reconstruction (i.e. the angle of the vector) using the *circ_mean* function in the CircStats MATLAB toolbox (Berens, 2009). We then computed how closely the vector points to the expected value of 0 by taking the cosine of this value.

$$\text{fidelity} = \max(r) * |\text{circ_r}(\theta, r)| \ * cos(\text{circ_mean}(\theta, r))$$

We then calculated the dispersion of the reconstruction (i.e. the vector length), which scales monotonically with the width of the function. As the reconstruction becomes narrower and more precise, the vector length increases. This is computed with *circ_r* (Berens, 2009).

Finally, we multiplied the mean direction by the dispersion and scaled the result by the maximum value of the reconstruction. This yields a fidelity metric which monotonically increases as a given reconstruction increases in amplitude and in precision (Figure 3.3B). The metric is negative when the mean reconstruction vector points in the opposite direction (180°), and close to 0 if the reconstruction is flat and contains no information about the remembered location.

As noted in the *Statistical analysis* section, we resampled across participants to generate confidence intervals for the sampled population. To generate single-participant confidence intervals, we bootstrapped the data by resampling across trials. If the 95% CIs overlapped with 0, then there was no significant memory representation for that participant or for that region. When we compared fidelity across memory tasks, we subtracted the two bootstrapped distributions (e.g. WM – LTM) and compared this difference distribution against 0.

Correlation between fidelity values and behavioral precision

To investigate how the fidelity of reinstated representations might be related to the precision of memory recall, we computed a Pearson's correlation between representational fidelity and behavioral recall precision, or the estimated mixture model *SD*. First, we split each participant's data in half (which roughly corresponds to each fMRI session) and generated two fidelity values and two behavioral precision estimates for each participant. This procedure enhanced the stability of our Pearson's *ρ* estimates by changing N = 10 to N = 20, although each point was not independent of the other. Second, we averaged representational fidelity

values across time points where there was a significant memory representation for both tasks, and when any sensory-evoked BOLD response would have decayed (TRs 4-8; Figure 3.3A). And lastly, we transformed the behavioral precision values to correct for skew in the distribution. Like many other size metrics, the distribution of *SD* values is skewed leftward. Therefore we log-transformed the values before computing the Pearson's correlation coefficient ρ.

To assess the significance of the ρ coefficient, we resampled with replacement across participants to obtain 95% confidence intervals for each ROI and memory task. Note that by resampling across participants, we preserve the dependence between the two sessions of the data for each participant. We then applied FDR correction across all ρ values.

Even with N = 20, it can be difficult to detect a reliable correlation. We performed a power analysis for an exact bivariate normal correlation using G*Power 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009). This allowed us to estimate how many participants would be needed to establish the observed effect size with a type II error rate of 5%.

Support vector machine classification

The analyses described above allowed us to generate continuous estimates of remembered stimuli to precisely compare the quality of spatial memory representations across ROIs and memory tasks. We were also interested in whether activation patterns in these ROIs could discriminate between the two memory tasks. To test this, we trained a support vector machine (SVM) with the default linear kernel to discriminate between WM trials and LTM trials. This was done using MATLAB's *fitclinear* routine.

We first divided the data into two temporal epochs during the memory delay period, and averaged across the TRs in those epochs (early epoch: TRs 2 – 4, late epoch: TRs 5 – 7). Classifier training and testing were performed separately for each epoch, ROI, and participant. To ensure that classification performance was not driven by a global difference in univariate response between the memory tasks, we also computed the global mean across all trials and

voxels for each memory task. We then subtracted this constant from the data before we cross-validated or computed the null accuracy distribution (see below).

Mean classifier accuracy was determined by k-fold cross-validation (k = 12, which is the typical number of runs in each task condition) using MATLAB's *crossval* function. We then assessed statistical significance by randomizing the labels associated with each task and generated a null distribution of classifier accuracies over 1000 iterations.

To combine the data across participants, we took all classifier accuracies for a given ROI and temporal epoch (10 values, one for each participant) and calculated the mean. We then performed a similar operation on the permuted null distributions, concatenating all the distributions (10 participants x 1000 iterations) and calculating the mean across participants to generate a single null distribution for the mean.

We computed empirical one-tailed p-values by comparing when the mean classifier performance was higher than the permuted null values (1 – mean(classAcc > nullDistr)). Note that we used one-tailed tests to reflect the fact that the classifier accuracies were not predicted to drop significantly below chance. These p-values were subject to FDR-correction across temporal epochs and ROIs. We used a stricter FDR q = 0.025 to reflect the fact that we only used one-tailed p-values.

Results

Behavioral data

Outside of the scanner, participants were trained on both the LTM and WM tasks until their mean recall error on the LTM retrieval task was stable and no longer decreasing (mean number of training days 7.39 [5.90, 9.10]). This training target did not enforce similar precision between tasks (participant averaged error on the last day: WM 4.27° [3.60°, 5.05°]; LTM 6.51° [4.67 °, 8.40°]). To characterize each participant's recall success and precision, we plotted a histogram of their recall errors and fit the data with an additive mixture of a circular Gaussian and a uniform distribution (Harlow & Yonelinas, 2016; Zhang & Luck, 2008). The standard

deviation (*SD*) of the Gaussian is a measure of the spatial precision of their recall, with larger SDs corresponding to imprecise recall. The height of the uniform distribution is a measure of their rate of forgetting. Here we report the complement of this number, the probability of recall (*P*(recall)). The interval training procedure combined with precise feedback significantly increased both the precision and probability of recall in the LTM task (difference between first and last set of trials: 14.03° [3.03°,33.74°] *SD*, $p < .001$; 0.047 [0.094,0.002] *P*(recall), $p = 0.04$).

This analysis was repeated for the behavioral responses made in the scanner. Overall task performance was similar between task conditions (participant averaged WM error 7.40 [5.76, 9.43]; LTM error 7.78 [4.66, 11.61]). The probability of recall was high (*P*(recall): WM 0.98 [0.96, 0.99], LTM 0.97 [0.94, 1.00]; Figure 3.2), and similar between both memory tasks ($p = 0.87$).

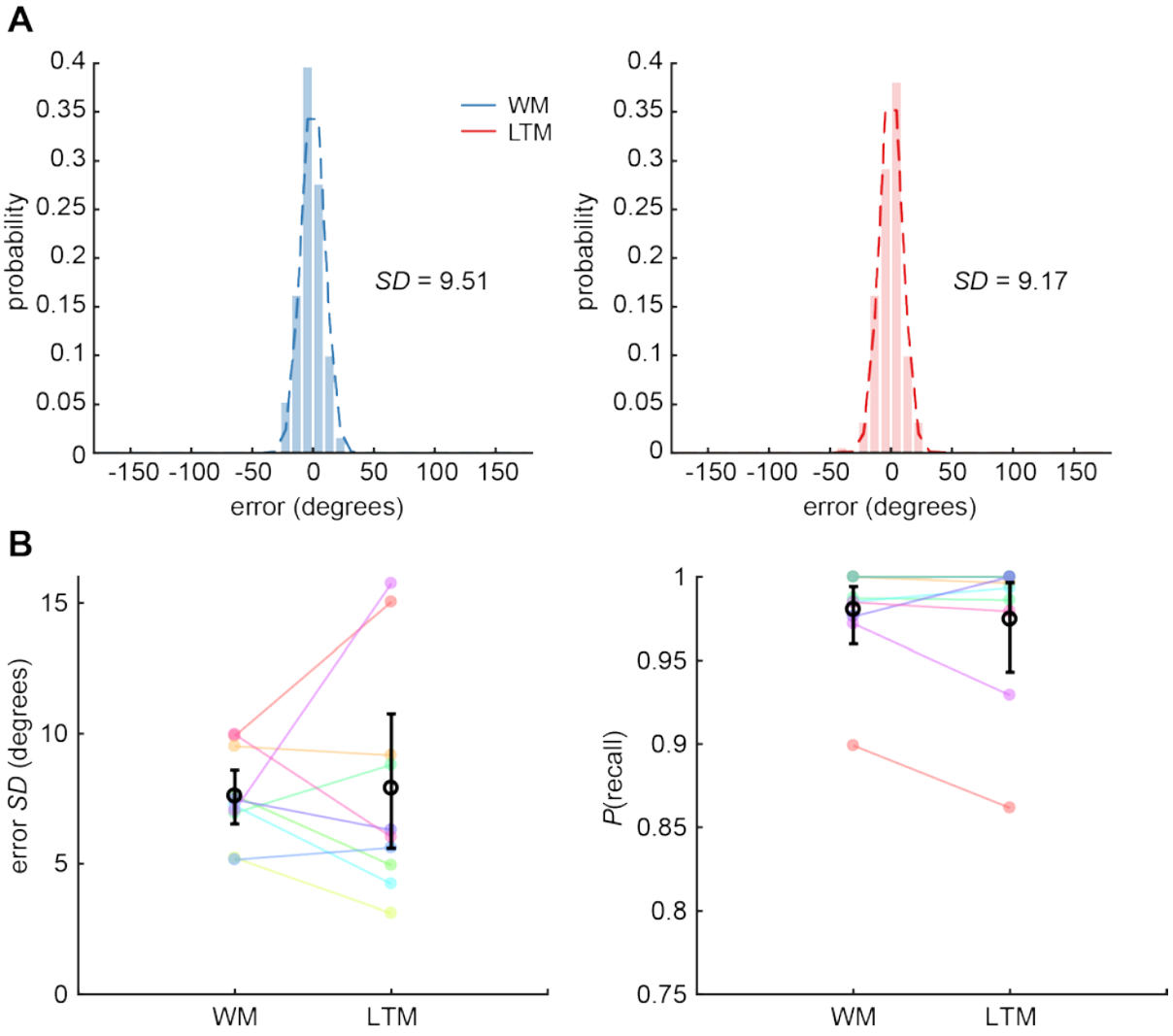


Figure 3.2. Memory recall across both scanner tasks is similar. (A) The distribution of memory recall errors for an example participant. We fit a mixture model to this distribution for every participant, where the Gaussian distribution characterized the variability of recall (standard deviation, *SD*) and the uniform distribution characterized the likelihood of recall (*P*(recall)). (B) The mixture model fit parameters for each subject and task. The mean across participants and 95% CIs are shown in black.

As a group, participants had similar recall precision between both tasks (mixture model *SD*: WM 7.60 [6.61, 8.62], LTM 7.87 [5.51, 10.57]; $p = 0.34$). However, recall precision varied more on the LTM task: some participants had worse precision in the LTM task than the WM task, while others had better precision (Figure 3.2). This variability between participants may reflect the quality of encoding during the initial study-retrieval training day.

Spatial reconstructions from the IEM

To assess how spatial information was encoded during the delay period of both memory tasks, we trained an encoding model using data from an independent perceptual task. The perceptual task allowed us to map the spatial selectivity of visually responsive voxels in several retinotopic regions of interest (ROIs).

We then inverted the trained encoding model to predict which spatial position was held in memory based on the pattern of voxel activations from the delay period of either the WM or LTM task. If this spatial 'reconstruction' is highly accurate, it suggests that the pattern of voxel activations during the perceptual task is very similar to the pattern of activations during the delay period of the memory tasks.

The spatial reconstruction (Figure 3.3B) peaks at the remembered position, and is parameterized as a curve in circular (polar) coordinate space since the eccentricity of the stimuli was fixed. To average across trials with different remembered positions, we recentered each trial so that the remembered position was at 0 degrees. In Figure 3.3A, we plot these averaged spatial reconstructions from area V1 at several timepoints during the memory delay of both tasks. As expected, spatial information during the delay period does not emerge until 4-6 seconds after stimulus onset, when the BOLD response evoked by a stimulus is expected to peak. Additionally, the WM task has a higher amplitude spatial reconstruction than the LTM task early in the delay period. This is likely due to fact that the remembered position is physically presented with the clip art item in the WM task, and therefore evokes a sensory response that occurs faster than a position retrieved from LTM. However, late in the delay period (~10 seconds after stimulus onset), we expect lingering sensory activation to have decayed. Here we find that spatial reconstructions in both the WM and LTM tasks are very similar.

To simplify our data and to quantify the reconstructions, we computed a metric which describes the fidelity of each recentered spatial reconstruction (*Methods*). This number monotonically increases as a reconstruction increases in amplitude and precision (i.e. as it

decreases in its width). We illustrate some properties of this metric and its computation in Figure 3.3B. First, each point in the reconstruction is replotted in polar space. We then find the mean vector of all of these points, which describes the average directional energy of the reconstruction. The fidelity, then, is proportional to the length of the vector along the expected direction (i.e. 0°, or the x-component of the vector; bold italic numbers in Figure 3.3B).

We found that spatial position was represented with high fidelity during the delay period of both memory tasks in several retinotopic ROIs. Representational fidelity tended to peak earlier for the WM task (mean timepoint across ROIs: WM TR 4.44, or 6.89 seconds after stimulus presentation; LTM 5.44, or 8.89 seconds), consistent with the physical presentation of the remembered position in the WM task. However, it also persisted well into the delay period. By 8 seconds, the representational fidelity between tasks was statistically indistinguishable (no pairwise tests survive FDR correction across ROIs and timepoints at $q = 0.05$; Figure 3.3C).

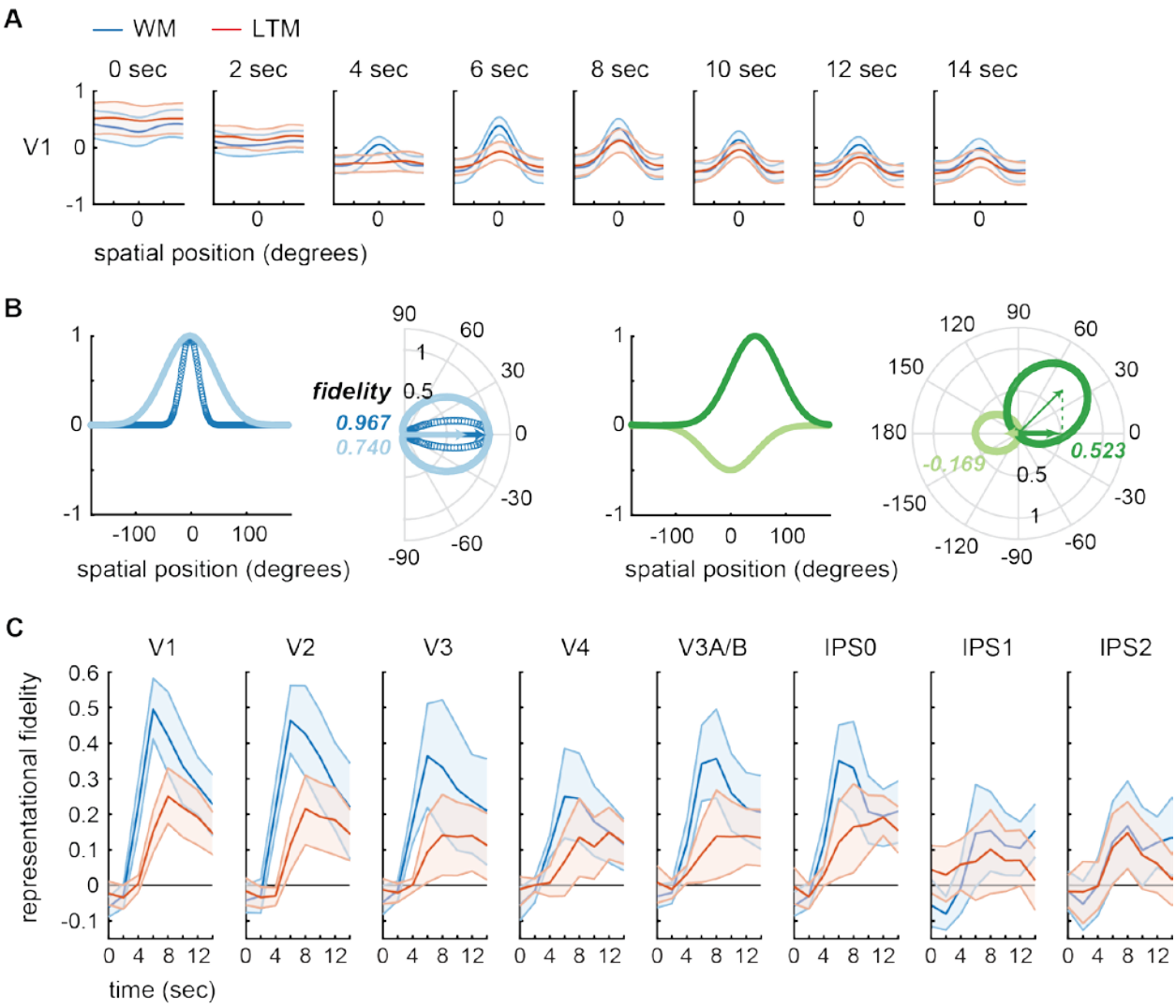


Figure 3.3. Reconstructions of remembered spatial positions over time for both memory tasks. Error bars are participant-resampled 95% CIs. (A) The timecourse of reconstructed spatial position using V1 data, averaged across participants (0 sec is stimulus onset). Remembered position was represented similarly for WM and LTM late in the delay period (8 – 14 seconds). (B) Example reconstructions and their corresponding fidelity values, a single number which characterizes the quality of tuning toward the remembered position (always set to 0° here). Each reconstruction was plotted in polar space. The fidelity metric is equivalent to the length of the bold horizontal vectors. Dark blue: the best reconstruction is narrow and centered exactly at 0°. Light blue: a broader reconstruction has a shorter mean vector, capturing the fact that less 'energy' is at 0°. Dark green: a reconstruction that is slightly offset from zero has a shorter x-component, and therefore a lower fidelity value. Light green: an inverted reconstruction has a mean vector that points in the opposite direction, resulting in a negative fidelity value. (C) Representational fidelity timecourses for all retinotopic areas we analyzed. CIs that intersect with 0 are not significant.

Furthermore, we find a modest correlation between the representational fidelity during the delay period (fidelity averaged over 6 - 14 seconds) and the precision of memory reports from both tasks (mixture model *SD*). This correlation is only present in V1 (mean correlation ρ and 95% CIs: WM -0.5263 [-0.7834, -0.1803], p = 0.010; LTM -0.5315 [-0.8272, -0.0105], p = 0.0480). However, this result does not survive FDR correction across task and all ROIs. While this trend is suggestive, our study is underpowered to detect a between-subject brain-behavior correlation of this magnitude. A power analysis suggests that a correlation of two normally distributed variables at $\rho = -0.5$ requires a sample size of at least 38 participants to achieve a 5% type II error rate.

Taken together, the evidence suggests that the LTM and WM tasks relied on similar sensory representations in retinotopic visual and parietal cortex. We assessed the representational fidelity of memory stimuli using a fixed encoding model that was trained on sensory-evoked responses from an independent task. We found no significant difference in representational fidelity between memory tasks in any ROI. Furthermore, we found weak evidence that the precision of memory reports in both tasks depended on the fidelity of visual memory representations in V1. These data imply that reinstated sensory information from both WM and LTM is encoded in a similar format.

Task classification

The encoding model analysis above only tested whether reinstated sensory memories were similar between WM and LTM. The reconstructed spatial positions only reflect the delay-period activity that resembles the visually-evoked activation patterns from the independent perceptual task. However, there may be additional information in these same retinotopic regions about the source of spatial information (WM or LTM). To test this possibility, we trained a binary support vector machine (SVM) to classify each memory trial as belonging to the WM or LTM task.

The SVM was trained and tested within two different temporal epochs of the memory task. We predicted that the SVM would easily discriminate between tasks when trained on data from the early temporal epoch (TRs 2 – 4, or 2 - 6 seconds), since (1) only the WM task contains an evoked sensory response at the remembered position and (2) we already observed a difference in representational fidelity between tasks early in the delay period (Figure 3.3). However, the sensory information in the late temporal epoch (TRs 5 – 7, or 8 – 12 seconds) is roughly equivalent. Thus, if the classifier could accurately distinguish between the memory tasks during the late epoch, it must reflect additional task information that is not well-captured by a sensory-driven model of remembered spatial positions.

We first tested this prediction by repeating the inverted encoding model (IEM) analysis in either temporal epoch. However, to parallel the SVM decoder, we trained the IEM using memory task data averaged over the TRs in either epoch. As expected, fidelity differed in the early temporal epoch in most ROIs (all but IPS1 & IPS2 survived FDR $q = 0.05$), but this difference disappeared by the late temporal epoch in all regions (FDR $q = 0.05$).

Having confirmed this prediction, we then trained and tested SVMs on data averaged across TRs in each temporal epoch. As predicted, we found that we could discriminate between the two tasks early in the delay period across all retinotopic ROIs (Figure 3.4, top row). However, early visual areas (V1 – V4) failed to discriminate between tasks in the late delay period. Instead, only parietal regions (V3A/B, IPS0, IPS1, IPS2) could discriminate between the WM and LTM task during the late temporal epoch (Figure 3.4, bottom row). Note that the successful SVM discrimination cannot be driven by differences in univariate delay period activity, because we removed the mean within each memory task before training and testing the classifier (*Methods*).

Our data suggest that while sensory information about the memoranda is largely similar between the WM and LTM tasks in several early retinotopic regions, information that differentiates the two tasks is selectively represented in retinotopic areas of parietal cortex.

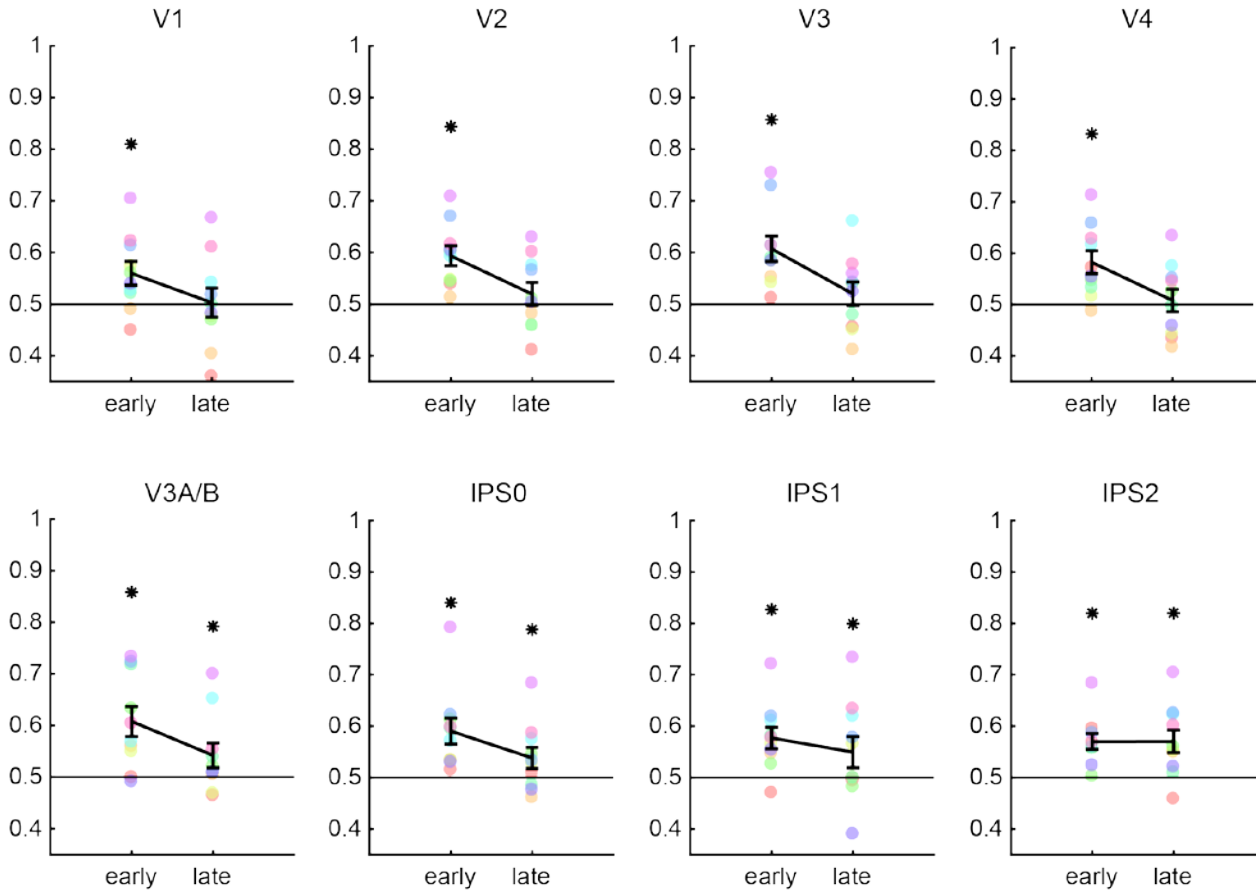


Figure 3.4. Decoding between memory tasks is successful only in parietal regions (bottom row) in the late delay period. The mean across participants and 95% CIs are shown in black. Individual participant colors are matched to Figure 3.2B. All p-values were FDR-corrected across ROI and temporal epoch.

Discussion

Several theories predict that WM and LTM rely on overlapping neural substrates, and a large body of evidence has shown that both tasks recruit sensory areas to represent remembered items (D'Esposito & Postle, 2015; Xue, 2018). Yet sensory recruitment and cortical reinstatement has typically been studied separately for each type of memory recall. While some researchers have investigated interactions between the two memory systems (Fukuda & Woodman, 2017; Lewis-Peacock & Postle, 2008), there has not been a systematic comparison between cortical memory representations in a WM task and a LTM task. Here we show that across a broad swath of retinotopic cortical regions, WM storage and LTM retrieval elicit graded,

spatially-specific representations that reinstate the patterns of activity observed during sensory encoding. Moreover, using training protocols that yielded similar behavioral precision during scanning, we show that late in the delay period, the fidelity of these neural representations was equivalent across the cortical hierarchy from V1 through dorsal parietal area IPS2 (Figure 3.3). We also investigated whether these same retinotopic regions contained information that distinguished between the WM and LTM tasks. Although we found no significant task difference in memory fidelity late in the delay period across all ROIs (Figure 3.3), only the retinotopic parietal regions (V3AB, IPS0, IPS1, IPS2) contained information that could distinguish between the WM task and the LTM task (Figure 3.4). Note that we carefully equated memory precision and visual properties to eliminate confounds that could have artifactually induced differences between WM and LTM. Thus, by focusing on a temporal window late in the delay period, when the impact of the peripheral stimulus in the WM condition had dissipated, we were able to obtain strong evidence for a signal that tracked memory type in retinotopic regions of parietal cortex.

To facilitate comparison between WM and LTM, we also minimized any differences in task performance. After several days of training with the LTM stimuli, task performance and memory precision in the scanner was equivalent (Figure 3.2). This was likely the result of the interval training procedure combined with precise numerical feedback. This procedure elicited an improvement in both the observers' ability to pair a location to a stimulus and an increase in the precision of their retrieved memories, such that performance was similar during the scanning sessions.

Our neuroimaging data has implications for theories about how WM and LTM are related. Many theories of how memory systems interact posit that LTM retrieval places remembered content directly into WM. In other words, both WM and actively retrieved LTMs operate under the same processes, as they are both in a capacity-limited state called the "focus of attention" (Atkinson & Shriffin, 1968; Cowan, 1995; D'Esposito & Postle, 2015; LaRocque, Lewis-Peacock, & Postle, 2014). Our findings fall directly in line with this account, and suggest

that regardless of the source of the spatial representation, a broad range of retinotopic cortical regions support the reinstatement of the same patterns that are observed during sensory processing (Serences, et al., 2009).

Our results also have some bearing on other theories of the role of attention and parietal regions in WM maintenance and LTM retrieval. Parietal regions that overlap with our retinotopic ROIs have been reported to be engaged during attention tasks, WM tasks, and LTM tasks (Awh & Jonides, 2001; Awh, Vogel, & Oh, 2006; Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Hutchinson et al., 2014; Sestieri, Shulman, & Corbetta, 2017). One hypothesis originating from the WM literature is that attention to an internal representation is the key to the maintenance or rehearsal of working memories, which can explain the anatomical overlap between fMRI and electroencephalography (EEG) studies of attention and WM (Awh & Jonides, 2001; Awh et al., 2006). A similar hypothesis has also been proposed in the LTM literature, suggesting that parietal regions mediate attention to episodic memories during retrieval (Cabeza et al., 2008; Sestieri et al., 2017). While we do find similar representations of the remembered stimuli with a sensory-driven encoding model, the ability of the SVM decoder to distinguish between WM and LTM tasks in parietal regions suggests that these attention-mediated memory process may not be identical. However, we note that the success of the SVM decoder may also be driven by representational differences, rather than by a distinct attention process (see discussion below). Overall, the data to date highlight the need to further delineate how attention modulates parietal activity and mnemonic representations during both types of memory tasks.

We found that retinotopic parietal regions represent both the remembered stimulus as well as information distinguishing the memory tasks themselves. We suggest that this is a form of representational ‘multiplexing’, although the specifics of how these representations are multiplexed are unknown. We discuss two possible forms of multiplexing here. One form of multiplexing holds that sensory representations are transformed differently according to the origin or source of the memory. This is consistent with a recently proposed hypothesis about

LTM that posits that stimulus representations in early sensory regions are partially transformed by the time they arrive at higher-order regions, such as dorsal or lateral parietal areas (Favila et al., 2018; Xiao et al., 2017). However, we do not have any positive evidence that stimulus representations are transformed differently between tasks. Our results show that the fixed encoding model trained on the independent sensory task generalizes to both WM and LTM, without any significant differences between memory task. We do find that representations of the remembered stimulus in both tasks degrade in quality in higher order regions, which could be consistent with representational transformations to both WM and LTM in these areas (Figure 3.3). An alternative form of multiplexing might be that sensory representations coexist simultaneously with an independent representation of the type of memory. In the context of our study, this would mean that the IEM and the SVM are operating on non-overlapping signals in parietal cortex. For example, separate neurons may be encoding the spatial memory representation and the type of memory (WM or LTM). Or the same neurons encode both representations, but in completely uncorrelated and independent ways.

Under either form of multiplexing, it is unclear what information the SVM decoder is using to classify the trials as WM or LTM. Again, we discuss two alternatives. One view based on prior work is that the parietal regions contain task-specific information. This is consistent with many reports that parietal regions prioritize task relevant representations (Bisley & Goldberg, 2010; Favila et al., 2018; Sprague, Itthipuripat, Vo, & Serences, 2018). Another view is that parietal regions contain specific representations of the source of memory representations. That is, a specific sensory memory could originate from recent perception, as in our WM task, or from a retrieved memory store, as in our LTM task. This would be a useful way to distinguish sensory memories from ongoing perception. Indeed, others have recently noted that this is a particular issue with cortical reinstatement. For example, if area V1 is representing detailed visual memories, concurrent sensory input may interfere with memory representations and subsequent memory recall (Bettencourt & Xu, 2015; Rademaker et al., 2018). In order to maintain separable

representations within sensory areas, parietal regions may additionally represent the source of sensory information. Further work is needed to test this hypothesis.

Finally, we note that our investigation of memory representations is limited to retinotopic visual and parietal cortex, and have no bearing on theories that suggest that WM and LTM are differentially dependent on the medial temporal lobe (MTL) (Squire & Wixted, 2011). A review of existing evidence suggests that the hippocampus is necessary for some types of WM tasks – perhaps specifically when they are complex (Yonelinas, 2013) or spatial tasks (Nadel & Hardt, 2011). And other work demonstrates that LTM retrieval is associated with the strength of hippocampal signals (Bosch et al., 2014; Richter, Cooper, Bays, & Simons, 2016; Ritchey et al., 2013). Future studies that are optimized for imaging and segmenting the MTL could use similar techniques to compare memory reinstatement and retrieval across WM and LTM in MTL regions. In particular, encoding models are powerful ways to test questions about how memories are represented. Here we focus on training a model based on sensory representations, but other models may implement specific models of episodic memory representations in the hippocampus. Indeed, some have advocated that a representational view of memory may provide a better understanding of the relationship between WM and LTM (Nadel & Hardt, 2011).

In conclusion, our findings suggest a striking correspondence between the patterns of activity elicited during the online maintenance of information encoded into WM or retrieved from LTM. In both cases, activity across a broad range of retinotopic regions recapitulated that which was observed during an independent sensory stimulation condition, showing that a sensory recruitment model provides a useful perspective for how visual details are maintained, regardless of the specific memory system. Moreover, during the later parts of the delay period, after sensory activity related to the WM sample had dissipated, we observed similar fidelity of the neural activity representing space even while activity in parietal regions distinguished between the two types of memory. Thus, while our broad findings fall in line with the hypothesis

that a common representational substrate is invoked during WM storage and LTM retrieval, they also highlight interesting differences in the sustained neural signals that support the two tasks.

Chapter 3 is being prepared for submission with colleagues either currently at or formerly at the University of Chicago and the University of California, San Diego. The author list and working title is: Vo, Vy A., Sutterer, David W., Foster, Joshua J., Sprague, Thomas C., Awh, Edward, Serences, John T. Long-term and short-term memory representations in retinotopic visual cortex: cortical reinstatement and source information. The dissertation author was the primary investigator and author of this paper.

References

Atkinson, R. C., & Shriffin, R. M. (1968). Human Memory: a Proposed System and Its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). New York, NY: Academic Press.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126. https://doi.org/10.1016/S1364-6613(00)01593-X

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and memory. *Neuroscience*, *139*, 201–208. https://doi.org/https://doi.org/10.1016/j.neuroscience.2005.08.023

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188.

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*(10). https://doi.org/10.18637/jss.v031.i10

Bettencourt, K. C., & Xu, Y. (2015). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*. https://doi.org/10.1038/nn.4174

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*(1), 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823

Bosch, S. E., Jehee, J. F. M., Fernandez, G., & Doeller, C. F. (2014). Reinstatement of Associative Memories in Early Visual Cortex Is Signaled by the Hippocampus. *Journal of Neuroscience*, *34*(22), 7493–7500. https://doi.org/10.1523/JNEUROSCI.0805-14.2014

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of Neuroscience*, *29*(44), 13992–4003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*(8), 613–625. https://doi.org/10.1038/nrn2459

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. New York: Oxford University Press.

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press. https://doi.org/10.1017/CBO9780511802843

Emrich, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *Journal of Neuroscience*, *33*(15), 6516–6523. https://doi.org/10.1523/JNEUROSCI.5732-12.2013

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience*, *25*, 754–761. https://doi.org/10.1162/jocn

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Ester, E. F., Sutterer, D. W., Serences, J. T., & Awh, E. (2016). Feature-selective attentional modulations in human frontoparietal cortex. *Journal of Neuroscience*, *36*(31), 8188–8199. https://doi.org/10.1523/JNEUROSCI.3935-15.2016

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*. https://doi.org/10.3758/BRM.41.4.1149

Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal representations of stimulus features are amplified during memory retrieval and flexibly aligned with top-down goals. *Journal of Neuroscience*. https://doi.org/10.1101/280974

Fukuda, K., & Woodman, G. F. (2017). Visual working memory buffers information retrieved from visual long-term memory. *Proceedings of the National Academy of Sciences*, (905). https://doi.org/10.1073/pnas.1617874114

Harlow, I. M., & Yonelinas, A. P. (2016). Distinguishing between the success and precision of recollection. *Memory*, *24*(1), 114–127. https://doi.org/10.1080/09658211.2014.988162

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Henderson, M. M., & Serences, J. T. (2019). Human frontoparietal cortex represents behaviorally-relevant target status based on abstract object features. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00015.2019

Hutchinson, J. B., Uncapher, M. R., Weiner, K. S., Bressler, D. W., Silver, M. A., Preston, A. R., & Wagner, A. D. (2014). Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. *Cerebral Cortex*, *24*(1), 49–66. https://doi.org/10.1093/cercor/bhs278

Jammalamadaka, S. R., & SenGupta, A. (2001). *Topics in Circular Statistics* (Vol. 5). World Scientific Publishing Co. https://doi.org/10.1142/4031

Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Kuhl, B. A., & Chun, M. M. (2014). Successful Remembering Elicits Event-Specific Activity Patterns in Lateral Parietal Cortex. *Journal of Neuroscience*, *34*(23), 8051–8060. https://doi.org/10.1523/JNEUROSCI.4328-13.2014

LaRocque, J. J., Lewis-Peacock, J. a, & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*, *8*(January), 5. https://doi.org/10.3389/fnhum.2014.00005

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience*, *16*(8), 997–9. https://doi.org/10.1038/nn.3452

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary Activation of Long-Term Memory Supports Working Memory. *Journal of Neuroscience*, *28*(35), 8765–8771. https://doi.org/10.1523/jneurosci.1953-08.2008

Nadel, L., & Hardt, O. (2011). Update on Memory Systems and Processes. *Neuropsychopharmacology Reviews*, *36*, 251–273. https://doi.org/10.1038/npp.2010.169

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2018). Simultaneous representation of sensory and mnemonic information in human visual cortex. *BioRxiv*. https://doi.org/10.1101/339200

Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *ELife*, *5*(e18260), 1–18. https://doi.org/10.7554/eLife.18260

Riggall, A. C., & Postle, B. R. (2012). The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, *32*(38), 12990–12998. https://doi.org/10.1523/JNEUROSCI.1892-12.2012

Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(12), 2818–2828. https://doi.org/10.1093/cercor/bhs258

Schneegans, S., & Bays, P. M. (2016). No fixed item limit in visuospatial working memory. *Cortex*, *83*, 181–193. https://doi.org/10.1016/j.cortex.2016.07.021

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, *128*, 53–67. https://doi.org/10.1016/J.VISRES.2016.09.010

Serences, J. T., Ester, E. F., Vogel, E. K., Awh, E., & Serences, J. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*, *20*(2), 207–214.

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience*, *18*, 183–192. https://doi.org/10.1038/nrn.2017.6

Sprague, T. C., Adam, K. C. S., Foster, J. J., Rahmati, M., Sutterer, D. W., & Vo, V. A. (2018).

Inverted Encoding Models Assay Population-Level Stimulus Representations, Not Single-Unit Neural Tuning. *ENeuro*, *5*(3), 1–5. https://doi.org/10.1523/ENEURO.0098-18.2018

Sprague, T. C., Ester, E. F., & Serences, J. T. (2014). Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology*, *24*(18), 2174–2180. https://doi.org/10.1016/j.cub.2014.07.066

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron*, *91*(3), 694–707. https://doi.org/10.1016/j.neuron.2016.07.006

Sprague, T. C., Itthipuripat, S., Vo, V. A., & Serences, J. T. (2018). Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex. *Journal of Neurophysiology*, *119*(6), 2153–2165. https://doi.org/doi:10.1152/jn.00059

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–87. https://doi.org/10.1038/nn.3574

Squire, L. R., & Wixted, J. T. (2011). The Cognitive Neuroscience of Human Memory Since H.M. *Annual Review of Neuroscience*, *34*, 259–88. https://doi.org/10.1146/annurev-neuro-061010-113720

Sutterer, D. W., & Awh, E. (2015). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-015-0937-x

Wolff, M. J., Jochim, J., Akyurek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546

Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., & Xue, G. (2017). Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. *The Journal of Neuroscience*, *37*(11), 2986–2998. https://doi.org/10.1523/JNEUROSCI.2324-16.2017

Xue, G. (2018). The Neural Representations Underlying Human Episodic Memory. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2018.03.004

Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behavioural Brain Research*, *254*, 34–44. https://doi.org/10.1016/j.bbr.2013.05.030

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

CONCLUSION

These three studies investigated how different cognitive processes—covert spatial attention, working memory, and long-term memory—affect the format of visual representations in human cortex. Together, they illustrate the utility of using explicit models to test hypotheses about neural representations. Here I discuss the positive contributions of this work to investigating neural representations and visual cognition, as well as some limitations of the general methods.

The findings from the studies in Chapters 1 and 2 illustrate that the allocation of spatial attention is flexibly supported by distinct patterns of attentional modulations. I found widespread and systematic changes in vRF position, size, amplitude, and baseline with covert spatial attention. However, many combinations of vRF modulations might support an increase in spatial information at the attended location to support behavior. For example, spatially selective increases in amplitude at the attended location can boost the population-level representation of the attended area. Alternatively, decreases in vRF size near the attended location and increases in vRF size adjacent to the attended location can increase the precision of the population-level representation. Indeed, the data in Chapter 2 illustrate that distinct patterns of attentional modulations support different task demands. These data show that changes in vRF response amplitude are more important when participants divide their attention across 3 peripheral targets, compared to when they focus attention at a single target. This underscores the importance of linking modulations of single encoding units to changes in the population-level code. It is long past time to abandon a ‘single-neuron doctrine’ in perception and cognition (Barlow, 1972). Work that bridges levels of analysis—i.e. across single-units and populations—may uncover general principles about population codes that could be extended to different domains. For example, it would be useful to perform a similar set of studies linking the modulation of single RFs to population-level codes, but with feature-based attention. Several studies have observed that feature-based attention and spatial attention operate by

fundamentally different mechanisms (e.g., Ibos & Freedman, 2016; Ling & Liu, 2009; Liu, Stevens, & Carrasco, 2007). Despite extensive psychophysical studies and several observations at either the level of single neurons or at a population level (Carrasco, 2011), there has been no empirical study linking the modulations of single RFs to a population-level representation under feature-based attention. Such linking studies might also help us better understand the interaction between spatial attention and feature-based attention.

In Chapter 3, I investigated two types of representations. One was a representation of spatial position, either maintained in WM or retrieved from LTM. Like the previous studies, the analysis of this type of representation relied on a spatial encoding model which exploited the retinotopic organization of visual cortex. The second type of representation was one that distinguished between the type of memory (WM or LTM). This analysis relied on a decoder that solved for a linear support vector which separated patterns of voxel activations into those belonging to a WM trial and those belonging to a LTM trial. I found that retinotopic parietal regions appear to 'multiplex' information about remembered spatial position and about the type of memory (WM or LTM). However, the format of the parietal representation which distinguishes between WM and LTM is unknown. Here an encoding model with an explicit form could be useful (Serences & Saproo, 2012). For a simple example, consider a population of voxels that either respond highly in the WM condition, or highly in the LTM condition. These may be scattered throughout the parietal regions and carry the representation that is decoded by the SVM. To test this, one could build a simple encoding model to estimate every voxel's preference for either memory condition. Then one could invert the model to predict a held out data set.

However, it must be noted that even models with an explicit format come with challenges in interpretation. The success of an inverted encoding model given a set of data does not prove that neural representations are in that exact format. One example to illustrate the difficulty of this problem is about the organization of visual cortex, which is known to contain several retinotopically organized, complete maps of visual space. However, it is not clear whether the

coordinates of those spatial maps are Cartesian or polar. Indeed, because these two different coordinate systems are mathematically isomorphic, this is a difficult question to answer (although see Zetzsche, Krieger, & Wegmann, 1999 for a detailed perspective). Even beyond completely isomorphic representations, a moderately successful encoding model may simply be accounting for representations that are in a similar or correlated format. This problem is not intractable, however. A manipulation which transforms the neural representation can provide useful diagnostic information. In the Cartesian vs. polar coordinate example, the authors argue that recording the results of a nonlinear transformation to spatially selective cells can distinguish between the two possible coordinate systems (Zetzsche et al., 1999). In the study described in Chapter 3, the data suggest that representations of reinstated memories are very similar between WM and LTM. One way to test whether these representations are truly similar, or simply appear correlated, is to transform the representations with an experimental manipulation. In a follow-up experiment, participants could first retrieve the remembered position, and then manipulate it somehow (e.g. by rotating it a constant amount). If the WM and LTM representations are indeed in the same format, this manipulation should result in the same type of modulations to the remembered representations across all areas of cortex we analyzed in Chapter 3. Furthermore, we can use this manipulation to test whether the linear decoder that distinguishes between WM and LTM trials relies on a signal that is independent from the representations of the remembered position. First, we can train the decoder to separate WM and LTM trials in trials without the manipulation. Then, we can use the same support vector to separate trials with the manipulation. If the decoder is using information independent of the spatial memory representation, there should be no difference in decoder performance.

In sum, this work illustrates how specific models can help us understand the structure of neural representations that support fundamental aspects of human cognition. It provides a basis for future investigations of sensory information processing and of representational structures in visual cognition.

References

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

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Ibos, G., & Freedman, D. J. (2016). Interaction between Spatial and Feature Attention in Posterior Parietal Cortex. *Neuron*, *91*, 931–943. https://doi.org/10.1016/j.neuron.2016.07.025

Ling, S., & Liu, T. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*(10), 1194–1204. https://doi.org/10.1016/J.VISRES.2008.05.025

Liu, T., Stevens, S. T., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, *47*(1), 108–13. https://doi.org/10.1016/j.visres.2006.09.017

Serences, J. T., & Saproo, S. (2012). Computational advances towards linking BOLD and behavior. *Neuropsychologia*, *50*(4), 435–46. https://doi.org/10.1016/j.neuropsychologia.2011.07.013

Zetzsche, C., Krieger, G., & Wegmann, B. (1999). The atoms of vision: Cartesian or polar? *Journal of the Optical Society of America A*, *16*(7), 1554–1565. https://doi.org/10.1364/josaa.16.001554