UNIVERSITY OF CALIFORNIA

Santa Barbara

Neural and Behavioral Prioritization of Attention Across Feature Dimension Maps

A dissertation submitted in partial satisfaction of the

Requirements for the degree Doctor of Philosophy

In Psychological and Brain Sciences

by

Daniel Dean Thayer

Committee in charge:

Professor Thomas C. Sprague, Chair

Professor Barry Giesbrecht

Professor Emily Jacobs

Professor Miguel Eckstein

June 2024

The dissertation of Daniel Dean Thayer is approved.

Barry Giesbrecht

Emily Jacobs

Miguel Eckstein

Thomas C. Sprague, Committee Chair

June 2024

Neural and Behavioral Prioritization of Attention Across Feature Dimension Maps

by

Daniel Dean Thayer

ACKNOWLEDGEMENTS

To Anne: Thank you for your constant support—from joining me on the 30-hour Santa Barbara drive, to helping me figure out my next steps, you have always been there to help. I love you mom.

To John: Thank you for always having something goofy to say during good times and keeping it real with me during tough times. Now I can *officially* tell you fun brain facts while you cook a gourmet meal—a fair exchange.

To Jo: Thank you for your kindness, creativity, and love. Whether it's an intricately planned walk through central park to visit a long-lost relative, an impromptu T-Swift concert, or a quick trip to the farmers market, you help make life full. I can't wait for our next adventure.

To Tommy: Thank you for being an amazing mentor. Your guidance helped me grow both as a scientist and a person. You've entertained my ideas (silly or otherwise), helped me be more confident with science, and retaught me the joy of Legos.

To my lab mates: Alison, Amelia, and Kelvin: Thank you for listening to me yap about science, and literally anything else that comes to mind. Because of you, I do not kill plants immediately, am comfortable on a motorcycle, and enjoy horror movies. The PhD was so much better because you were there for the ride.

To the Giesbrecht lab: Jordan, Henri, Carly, Caitlin, Dawa, Shibu, Emily, and Barry: Thank you for the lengthy chats, hiking expeditions, pizza competitions, and cursed flights (which may or may not be my fault). I'm fortunate that y'all gave me a lab away from lab.

To my research assistants: Thank you for all your hard work in helping me collect data for these projects and for bringing excitement to the BIC with conversations ranging from K-Pop to Nobel Prizes.

To my U of I mentors: Brett, Brad, Tobin, Jonathan, Eliot, Jan, and Andrew: Thank you for taking the time to teach me about research and the PhD. I wouldn't have known this career was possible without your help.

To my Iowa family: Josh, Araz, Trevin, and Nick: Thank you for the countless hours of laughter and fun. Y'all helped make life during the PhD easier by reminding me to relax sometimes. Here's to many more souls-like and VTM sessions.

VITA OF DANIEL DEAN THAYER
June 2024

EDUCATION

Bachelor of Science in Psychology, University of Iowa, May 2018 (honors in the major)
Doctor of Philosophy in Psychological and Brain Sciences, University of California, Santa Barbara, June 2024 (expected)

PROFESSIONAL EMPLOYMENT

2018-19: Research Assistant, University of Iowa
2019-24: Teaching Assistant, University of California, Santa Barbara
2019-24: Graduate Student Research, University of California, Santa Barbara

PUBLICATIONS

Bahle, B., **Thayer, D. D.**, Mordkoff, J. T., & Hollingworth, A. (2019). The architecture of working memory: Features from multiple remembered objects produce parallel, coactive guidance of attention in visual search. *Journal of Experimental Psychology: General*.

Thayer, D. D., Bahle, B., & Hollingworth, A. (2021). Guidance of attention from visual working memory is feature-based, not object-based: Implications for models of feature binding. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001116

Thayer, D. D., Miller, M., Giesbrecht, B., & Sprague, T. C. (2022). Learned feature regularities enable suppression of spatially overlapping stimuli. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-022-02612-1

Thayer, D. D., & Sprague, T. C. (2023). Feature-Specific Salience Maps in Human Cortex. *Journal of Neuroscience*, *43*(50), 8785–8800. https://doi.org/10.1523/JNEUROSCI.1104-23.2023

Thayer, D. D., & Sprague, T. C. (2023). Interactions between working memory and attention depend on remembered feature dimension. *PsyArXiv*. https://doi.org/10.31234/osf.io/sx6d5

AWARDS

2019 University of California, Santa Barbara Conference Travel Grant
2020 Society for Neuroscience Trainee Professional Development Award
2021 National Science Foundation Graduate Research Fellowship: Honorable Mention
2021 Rich Mayer Award
2022 Kavli Summer Institute Fellowship
2023 Graduate Dissertation Fellowship
2024 Rice Academy Postdoctoral Fellowship
2024 Vision Sciences Society Travel Grant

ABSTRACT

Neural and Behavioral Prioritization of Attention Across Feature Dimension Maps

by

Daniel Dean Thayer

Attention is a cognitive mechanism used to select important information. Here, importance is defined as one of three factors: image salience, or something that stands out based on local feature contrast; goal relevance, or items that are related to accomplishing a task; and selection history, or implicitly learned stimulus properties. Features of an item, like color and motion, are prioritized for attentional selection, but the exact mechanisms that allow for feature-specific attentional selection are unclear. Across three studies, evidence is presented on how information is attended based on feature-specific image salience, goal relevance, and selection history. Chapters 2 and 3 describe two fMRI studies investigating how feature-selective regions of visual cortex are recruited to track the location of color and motion stimuli throughout the visual field. Stimuli were either made salient because they were a different feature value from the rest of the display (Chapter 2), or relevant because participants were instructed to attend a certain feature dimension (Chapter 3). Inverted

encoding models were computed to reconstruct spatial maps using neural activation patterns from retinotopic color (hV4/VO1/VO2) and motion (TO1/TO2) regions. Neural responses to the stimuli were strongest when the preferred feature of each region was important through manipulations of salience or relevance. These findings implicate color and motion areas as 'feature dimension maps'—regions that are critical for indexing important color and motion information for attention. Chapter 4 describes a study concretely demonstrating that feature values learned over the course of an experiment can modulate what is prioritized. Specifically, the color of a regularly presented distracting stimulus was learned, and eventually ignored to improve target detection. Suppression was potent enough to interfere with ongoing goals, as participants took longer to select the target when it was shown in the learned distractor color. This rules out alternative accounts suggesting that only spatial information can be suppressed through selection history. Together, this body of work highlights how stimulus features are leveraged to successfully navigate our environment.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter I: Prioritizing Important Visual Information

Priority map theory

The environment is filled with an overwhelming amount of sensory information that is impossible to completely process at a given moment. Yet, we are seemingly able to manage this input with ease and successfully interact with our surroundings. How is it that irrelevant information is ignored, while important objects are processed? Attention. Attention is how we choose stimuli that are important for behavior and is generally thought to operate by processing a limited amount of available sensory information. This can occur through a vast array of mechanisms (Duncan, 1984; Posner, 1980; Treue & Trujillo, 1999a) that can all be categorized as some form of selection or filtering operation (Bundesen, 1990; Carrasco, 2011; Duncan & Humphreys, 1989; Eckstein, 2011). One architectural framework that has helped scaffold the numerous mechanisms of attention is priority map theory (Awh et al., 2012; Itti & Koch, 2001; Luck et al., 2020; Serences & Yantis, 2006; Treisman & Gelade, 1980; Wolfe, 1994, 2021).

At the broadest level, priority map theory describes a series of spatial maps that index locations in the environment as 'important'. Spatial information is extracted from retinal input to generate these maps, and locations that are important have a high level of activation. Locations with the highest activation are selected by attention for further processing in what is called the *priority map*. The priority map is a feature agnostic representation of the visual field, meaning that it indexes information regardless of the specific properties that make it important. While it is ambiguous within the priority map what makes something important, there are several maps that compute importance for specific feature dimensions. Like the

priority map, these *feature dimension maps* are spatially organized and index important locations. The only difference is that they compute importance for a single feature dimension. Various feature dimensions are leveraged to determine what locations should be prioritized (Wolfe & Horowitz, 2004, 2017). This includes color (D'Zmura, 1991), motion (Mendoza et al., 2011), orientation (Treisman, 1985), size (Hollingworth & Luck, 2009), and shape (Theeuwes, 1992).

Even though individual feature dimension maps are necessary to select feature-specific information, ultimately, activation profiles across feature dimension maps are integrated to form the priority map. It is the activation profiles within the priority map that determine what is selected by attention. Selection can occur in a variety of different ways, such as through: winner-take-all (Koch & Ullman, 1985), where the location with the highest level of importance is always selected; diffusion (Wolfe, 2021), where one or more locations are processed at a time and evidence accumulates to attend a specific location; or probabilistic computations (Eckstein, 2017), where greater importance increases the chance a location is selected, while not guaranteeing that it will be selected over another location.

Several modulatory inputs contribute to sculpt importance across these maps. Broadly, importance at a location is determined by image salience, observer goals, and selection history. Each input can increase the importance of a location, and some can decrease the importance of a location under certain circumstances. I will first define each factor and will then provide empirical examples of how each modulates importance.

Modulations of priority

As described above, priority and feature dimension maps are representations of the visual field derived from retinal input. As such, information that stands out based on physical, image-computable salience, will be prioritized within these maps. Salience is defined as locations with high local feature contrast (Nothdurft, 1993a, 1993b, 2000), which can occur for any core feature dimension that is processed by the visual system (e.g., color, motion, and orientation). For example, a red strawberry in a bed of green leaves would stand out and capture attention because of the color contrast. A task commonly used to investigate how salience is prioritized is the additional singleton paradigm (Theeuwes, 1992). In this task, participants are shown display comprised of several colored shapes along an invisible ring in the periphery. Typically, they search for a single target item that is a different shape from all other display items (e.g., diamond among circles), which results in a salient shape target. On some trials, all items in the display are the same color, but occasionally one non-target item is a different color from everything else (e.g., red item among green items). This item is now salient through color and is considered a 'singleton distractor'. On trials where a color singleton distractor is present, response times are slower than when there was no singleton in the display (Theeuwes, 1991, 1992; Theeuwes et al., 2000, 2003), suggesting that it captured attention as participants were searching for the target.

Neural studies confirm the role of salience in directing attention. Usually, a salient stimulus evokes a stronger response relative to isoluminant non-salient stimuli as indexed by electrophysiological recordings in monkeys (Bisley & Goldberg, 2006; Chen et al., 2020; Gottlieb et al., 1998; Klink et al., 2023; Lin et al., 2024), EEG (Gaspelin & Luck, 2018; Li, 2002; Liesefeld et al., 2017; Sawaki & Luck, 2010), and fMRI (Bogler et al., 2011, 2013;

Mulckhuyse et al., 2011; Poltoratski et al., 2017; Serences et al., 2005). Many of these studies employ the additional singleton paradigm, but similar results are observed when using different stimulus displays (e.g, Li, 2002). Thus, stimuli that are salient in the visual field are considered important within the priority map framework.

Observer goals, or relevance, refers to the intentions of an individual to find items for accomplishing a task. This could be the goal of finding a car in the parking lot or a red shirt in the closet. Relevance can modulate what is attended by increasing the activation of specific spatial locations or features within the priority and feature dimension maps (Cave & Wolfe, 1990). Behaviorally, this has often been observed using visual search tasks (Bahle et al., 2019; Becker et al., 2010; Folk et al., 1992; Leber & Egeth, 2006). In one example, Bahle et al. (2019) instructed participants to search for a specific color and shape. They were told that any item containing either the cued color or shape was the target. On some trials, only a single task-relevant feature was present (e.g., a red circle when looking for either blue or circle), but on other trials both relevant features were present in the same item (e.g., blue circle). When the target matched both features, response times were faster than when only a single relevant feature was present. This is due to priority signals integrating across both the shape and color dimension maps, increasing activation at the corresponding location within the feature-agnostic priority map more than what would occur if only the color or shape map was prioritized. Furthermore, goals modulate priority independent of salience, as attentional benefits are still observed when there is no salient information in the display, such as with a heterogenous search display (Bacon & Egeth, 1994).

Goals also enhance neural responses to a stimulus. Using electrophysiological measures, greater spike rates are observed from receptive fields (RFs) when a target is present than when a non-target item is present (Bisley & Goldberg, 2003; Bruce & Goldberg, 1985). Similar results have been observed in fMRI (Poltoratski et al., 2017; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013) and EEG (Gaspar & McDonald, 2014; Stilwell et al., 2022). Furthermore, knowing what is task irrelevant can result in neural and behavioral suppression of non-target stimuli (Arita et al., 2012; Carlisle, 2023; Cosman et al., 2018; Zhang et al., 2022), demonstrating the flexibility of goal-based prioritization. Thus, goals greatly influence what is important in the visual field.

Selection history is a third factor influencing visual importance, which is how previous encounters with a stimulus influence what is currently prioritized (Awh et al., 2012; Luck et al., 2020). For instance, if a kitten leaps across the kitchen, that may capture attention because of the sudden salient motion. However, if the kitten leaps across the kitchen every day, one may learn to ignore the salient motion because it is a common occurrence that was encountered in the past. Selection history can both increase (Chun & Jiang, 1998, 2003) and decrease (Stilwell et al., 2019; Wang & Theeuwes, 2018) stimulus priority. One example is when a target has a high probability of being in a quadrant of the screen, then responses are faster when the target is in the high-probability quadrant relative to a low-probability quadrant (Chun & Jiang, 1998; Jiang et al., 2013). In another study, Stilwell and colleagues (2019) had participants perform the additional singleton task, where a color singleton was occasionally present. On most trials, the singleton distractor was a specific color, but could still be other colors (e.g., 65% red, 35% other). When the high-probability color singleton

was present, participants found the target much faster than when a low-probability color distractor was shown. They argue that the learned distractor color was suppressed, making it easier to find the target.

While selection history can occur after learning a statistical regularity over several encounters, it can also emerge through immediate experience (Wiggs & Martin, 1998; Wöstmann et al., 2022). One example of this is priming (Maljkovic & Nakayama, 1994). Priming is when target properties on the current trial match what was observed in the previous trial, which often improves the response to the current trial target. An example would be if a target on trial N-1 is red, and the current trial target is also red. When this occurs, individuals are faster and more accurate at reporting the current trial target. A similar effect can occur when distracting information is repeated, where responses to a target are faster when a distractor stimulus is shown in the same color on subsequent trials (Thayer et al., 2022). Note, that modulations of priority through selection history are distinct from observer goals because they are implicitly acquired, while goals are explicit (Gao & Theeuwes, 2022).

Burgeoning neural findings corroborate behavioral selection history results. Adam & Serences (2020) had participants perform the additional singleton paradigm in the scanner, where four stimuli were shown on each trial. On some blocks of trials, the singleton was always the same color, while other blocks the singleton color was randomized. In V1, neural responses to the color distractor were much weaker when the color was the same, indicating that subjects learned to suppress the irrelevant distractor. Evidence for neural modulation through selection history has similarly been observed in EEG (Gaspelin & Luck, 2018) and

electrophysiological recordings (Klink et al., 2023), further supporting the role of selection history in computing importance.

In sum, the priority map framework as described has clearly defined components such as the priority and feature dimension maps, that index importance at specific visual field locations. Importance in these maps is determined through image salience, observer goals, and selection history. This is supported by a large body of behavioral and neural empirical work, explaining how these factors direct attention to select important visual information.

Neural correlates of priority maps

Factors that compute attentional priority have a direct impact on both behavioral and neural responses. Because of this, much work has been dedicated to identifying neural regions that may represent a priority map. For a region to be considered a priority map, it needs to be spatial selectivity, preferentially responsive to important stimuli, and track important stimuli across various feature dimensions. One study that identified a region satisfying all these criteria was conducted by Klink and colleagues (2023). The researchers recorded activity from RFs in monkey V4 while they performed the additional singleton task. Responses to a shape-defined target and a color-defined salient distractor were stronger than to a non-salient distractor, demonstrating that monkey V4 tracked image-salient distractors and goal-relevant targets. Since increased responses were observed from neural RFs, there must be some degree of spatial selectivity, and because stimuli were important through shape or color, V4 must track importance across feature dimensions. An additional detail of this study was that the color of the salient distractor was always the same. This encouraged learned suppression of distractors through selection history (Stilwell et al., 2019). Consistent with this, the

initially strong response to the salient distractor quickly decreased below baseline approximately 150 ms after stimulus presentation. Monkeys likely learned to reactively suppress the salient distractors after it initially captured attention (Geng, 2014). This study is compelling evidence supporting monkey V4 as a candidate neural priority map.

Many other regions have been identified as neural priority maps, including: lateral geniculate nucleus (Kastner et al., 2006), intraparietal cortex (Bisley & Goldberg, 2003, 2006; Chen et al., 2020), frontal eye fields (Moore & Armstrong, 2003), V1 (Li, 2002; Poltoratski et al., 2017; Zhaoping & May, 2007, 2007), extrastriate visual cortex (Adam & Serences, 2020; Bogler et al., 2011, 2013; Burrows & Moore, 2009; Klink et al., 2023; Mazer & Gallant, 2003; Sprague & Serences, 2013), substantia nigra (Basso & Wurtz, 2002), superior colliculus (Basso & Wurtz, 1998; Fecteau & Munoz, 2006; White et al., 2017), and pulvinar (Shipp, 2003). All these regions track the location of important stimuli at specific locations for at least one feature dimension and are modulated by salience and relevance. Although, it is worth highlighting that very few studies have shown how these regions are modulated by selection history (but see: Adam & Serences, 2020). Due to there being so many candidate neural priority maps, it can be difficult to identify which region ultimately directs attention, especially when the cognitive framework describes a single map. However, when implemented by the brain, there is no requirement that a single region should represent the priority map, as redundant coding may help ensure that the appropriate information is selected by attention (Zelinsky & Bisley, 2015). A similar architecture has been proposed for working memory, where multiple redundant representations can help ensure a robust memory signal in case any single representation is compromised (Christophel et al., 2017;

Sreenivasan & D'Esposito, 2019), adding plausibility to this account. Overall, there is overwhelming evidence that priority maps are reflected in neural architecture.

Candidate neural feature dimension maps

While there is much empirical support for various candidate neural priority maps, there has been little work identifying the correlates of feature dimension maps. This is particularly surprising since they are needed to determine what is prioritized by attention. Like a priority map, there are several essential criteria for a region to be considered a neural feature dimension map. They must be spatially selective and respond to important information, but unlike the priority map, the prioritized information needs to be feature specific. Nearly all studies that support a given region as a neural priority map have either utilized single feature dimension stimuli, measured responses from a limited portion of the visual field, or studied feature-agnostic salience. This diminishes the ability to interpret any region as a feature dimension map.

Even though previous work has not identified specific regions as neural feature dimension maps, some have speculated that visual cortex could contain such maps (Katsuki & Constantinidis, 2014). This is a reasonable assumption because much of visual cortex contains several spatially organized retinotopic maps (Wandell et al., 2007; Wandell & Winawer, 2015) and many of these retinotopic regions are modulated by goal-relevance (McMains et al., 2007; Runeson et al., 2013), image-salience (Sprague, Itthipuripat, et al., 2018; Wang et al., 2022; Zhang et al., 2012), and selection history (Adam & Serences, 2020). Furthermore, some retinotopic regions happen to be feature-selective (Huk et al., 2002; Wade et al., 2002). For instance, Huk & Heeger (2002) showed participants either static or moving

dot fields. Univariate BOLD responses in MT and MST, which align with retinotopically defined TO1 and TO2 (Amano et al., 2009), were stronger when moving dot fields were presented. This demonstrates that TO1/TO2 are motion selective. A similar study was conducted identifying color-selective regions. Using grayscale and colorful stimuli, Wade and colleagues (2002) found that hV4, VO1, and VO2 preferentially responded to color. Thus, for color- and motion-selective regions of visual cortex, there is some evidence to suggest they represent feature dimension maps.

Each of these findings are independent components implicating hV4/VO1/VO2 and TO1/TO2 as feature dimension maps. However, for a region to be considered a feature dimension map, all properties need to converge—even though there are motion- and color-selective regions of visual cortex that are modulated by priority signals, it is unknown if these areas track the exact location of important feature-specific stimuli throughout the visual field. These regions need to prioritize the location of important feature-specific stimuli over non-prioritized locations for them to be considered neural feature dimension maps. Chapters 2 and 3 provide the necessary evidence to identify color-selective regions as a neural color dimension map, and motion-selective regions as a neural motion dimension map. Chapter 2 describes a study where stimuli were important based on feature-specific image salience and the location of the stimulus varied on each trial. Feature selective regions were able to track the location of the important stimulus throughout the visual field and had a stronger response to stimuli based on region preferences. Similarly, Chapter 3 describes a study where the location of a colorful moving dot stimulus varied on each trial. The key manipulation was that the relevant feature dimension of that stimulus (color or motion) varied across trials.

Feature-selective areas had a stronger response to the stimulus whenever the goal-relevant feature matched region preferences. Thus, these two studies identify color- and motion-selective regions as neural feature dimension maps.

Even though selection history is a factor that contributes to attentional priority, there are few studies investigating how previous encounters with a stimulus modulate neural responses. Research leveraging neural adaptation with fMRI provide some degree of support that visual cortex is influenced by stimulus history (Grill-Spector & Malach, 2001). In these studies, a specific stimulus is repeated several times (e.g., green and red lines; Engel & Furmanski, 2001) and occasionally a novel stimulus is presented (e.g., dark and light gray oriented lines). If neural responses to the novel stimulus are greater than during the stimulus repetition period, neural populations adapted to the repeated stimulus. This indicates that a region has some degree of selectivity for the repeated visual features and that responses are influenced by stimulus history (Benda, 2021). There are many proposed functions for neural adaptation (Webster, 2015). It could be a mechanism for optimizing information (Wark et al., 2007) or building predictions about the environment (Chopin & Mamassian, 2012). However, neural adaptation needs to be directly linked to attentional priority for it to be useful in identifying candidate feature dimension maps. Behavioral evidence implies that adaptation can alter perceived salience (McDermott et al., 2010), which would be a modulation of attentional priority, but this link needs to be demonstrated neurally.

An additional complication is that some models claim selection history can only suppress priority via spatial mechanisms (Theeuwes, 2010; Theeuwes et al., 2003). If true, then there would be no need for selection history to act within feature dimension maps at all since

suppression would be better instantiated within the priority map, making it unclear what level of the priority map framework should reflect selection history prioritization. This makes it difficult to know if selection history should be a defining characteristic of neural feature dimension maps. Chapter 4 presents a study investigating whether specific features can be suppressed independent of location. Based on those results, specific features can be suppressed, indicating that feature dimension maps must be prioritized by selection history. Together, this body of work characterizes key neural and behavioral evidence for feature dimension maps.

Chapter II: Neural Prioritization of Feature-Salient Stimuli

Introduction

Often, we search for items that are relevant to ongoing goals, such as the coffee maker in the morning. However, objects within a given scene are constantly vying for our attention. A salient but task-irrelevant object, like a bright yellow banana on the counter or a cat leaping across the kitchen, may distract our attention and slow the search for coffee. One prominent model that highlights the competition between task-irrelevant salience and task-relevant goals in guiding attention is priority map theory (Awh et al., 2012; Fecteau & Munoz, 2006; Itti & Koch, 2001; Serences & Yantis, 2006; Treisman & Gelade, 1980; Wolfe, 1994). Per this theory, the activity profile across a priority map reflects a combination of task-relevant locations and salient, but irrelevant, locations, and is used to direct attention to the highest priority locations (Carrasco, 2011; Eckstein, 2011; Yu et al., 2023).

To compute the bottom-up salience associated with a given location, priority map theory posits that information about individual feature dimensions (e.g., color, motion, etc) is independently extracted from retinal input into a series of 'feature dimension maps'. For a given feature dimension map, salient regions of space are defined based on within-dimension local feature contrast, such as the aberrant color of the yellow banana or motion direction of the leaping cat (Itti & Koch, 2000, 2001). A conspicuous location defined by a given feature dimension is given high activation in the corresponding feature dimension map (Fig. 1). Activity profiles across various feature dimension maps are then integrated into a unified feature-agnostic priority map, which indexes the most important locations within the visual field, regardless of the source of their importance.

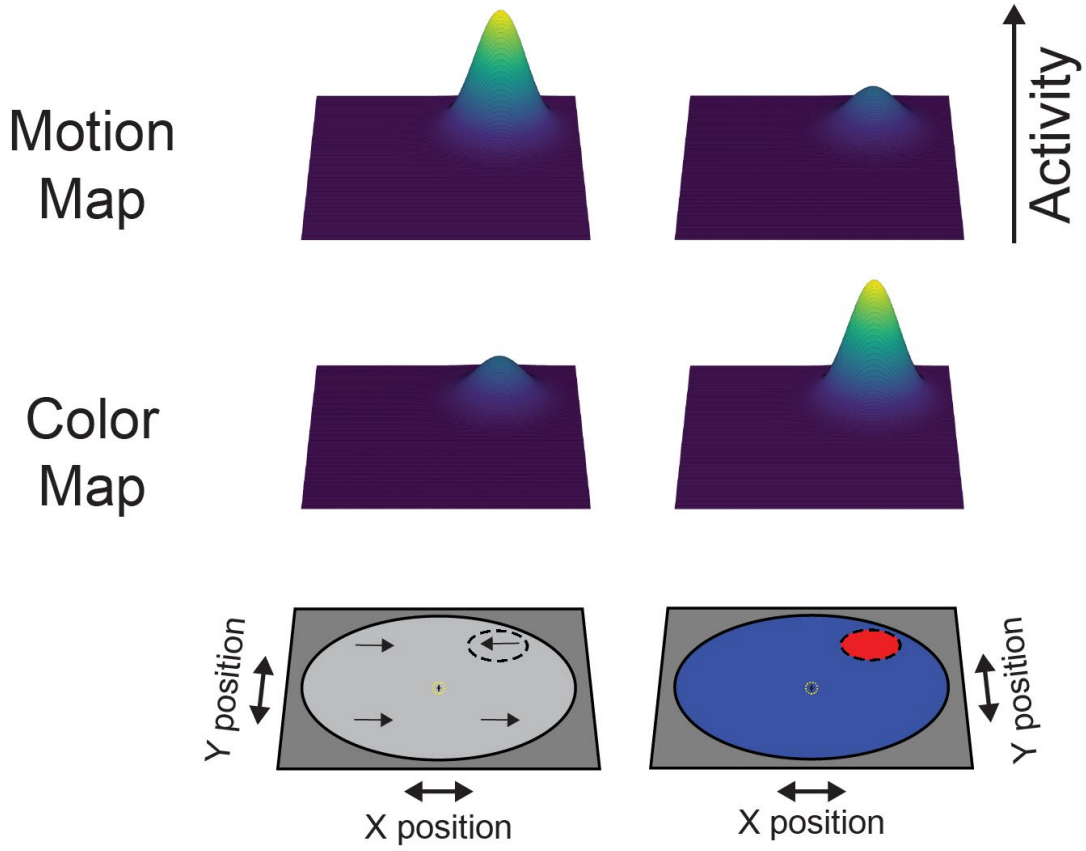

Figure 1. Feature dimension maps index salient locations based on their preferred feature dimension. Priority map theory invokes 'feature dimension maps' to compute salient location(s) based on local feature contrast within each feature dimension (e.g., color, motion). Accordingly, when a location in a stimulus display is made salient based on local differences in motion direction, activation profiles over a 'motion map' should track the salient location, while a 'color map' would not. Similarly, map activation corresponding to a salient motion stimulus should be stronger in a 'motion map' than when a location is made salient based on local differences in color. Complementary results would be predicted for a 'color map', with stronger activation at the location of a salient color stimulus as compared to a salient motion stimulus. While these feature dimension maps productively account for behavioral results in visual search tasks, it remains unknown whether salient locations are independently indexed in different feature-selective regions in visual cortex.

Studies in humans and nonhuman primates have identified stronger neural responses associated with salient stimulus locations than those associated with non-salient locations throughout the brain (Bogler et al., 2011, 2013; Bisley & Goldberg, 2006; Bichot & Schall, 1999; White et al., 2017), consistent with a neural instantiation of a priority map (Bisley &

Goldberg, 2010; Fecteau & Munoz, 2006; Katsuki & Constantinidis, 2014; Serences & Yantis, 2006). However, despite this converging neural evidence offering strong support for the implementation of feature-agnostic priority maps, support for feature dimension maps is primarily based on behavioral studies measuring visual search response times (Bacon & Egeth, 1994; Folk et al., 1992; Folk & Anderson, 2010; Huang & Pashler, 2007; Müller et al., 1995; Theeuwes, 1992; Treisman, 1998; Wolfe & Horowitz, 2004). Indeed, neural studies have all either focused on stimulus displays involving a single salient feature dimension (Beck & Kastner, 2005; Cook & Maunsell, 2002; Moran & Desimone, 1985; X. Zhang et al., 2012), measured from a single feature-selective visual region (Bichot et al., 2005; Burrows & Moore, 2009; Martínez-Trujillo & Treue, 2002; Mazer & Gallant, 2003; Ogawa & Komatsu, 2004, 2006; Klink et al., 2023; Reynolds & Desimone, 2003), or studied feature-agnostic salience (Bisley & Goldberg, 2006; Bogler et al., 2011, 2013; Gottlieb et al., 1998; Sprague, Itthipuripat, et al., 2018). Thus, despite the key theoretical role specific stimulus feature dimensions are believed to play in computing representations of stimulus salience, it remains unknown whether the brain implements this compartmentalized computational architecture.

Here, we sought to resolve this question by testing the hypothesis that feature-selective retinotopic regions of visual cortex preferentially index salient locations based on their preferred feature dimension. Within retinotopic ROIs, we characterized feature selectivity and spatial salience computations for task-irrelevant visual stimuli defined by different feature dimensions using a spatial inverted encoding model. Participants attended a central fixation point while viewing stimuli typically containing one salient location. Across trials,

we varied the salience-defining stimulus feature (color, motion, or a single salient stimulus in isolation). If color-selective retinotopic regions hV4/VO1/VO2 (Brewer et al., 2005; Conway et al., 2007; Mullen, 2019) and motion-selective retinotopic regions TO1/TO2 (Albright, 1993; Amano et al., 2009; Huk et al., 2002) act as neural feature dimension maps, salient stimuli will result in patterns of multivariate BOLD responses containing strong activation at the salient location that are strongest when defined by a region's preferred feature dimension (Fig. 1). Consistent with these predictions, we found that reconstructed spatial maps in color-selective regions indexed color-based salience, and motion-selective regions indexed motion-based salience, with each region preferentially representing salient locations based on their feature dimension.

Materials & Methods

Participants.

8 subjects recruited from the University of California, Santa Barbara (UCSB) community participated in the primary fMRI study (6 female, 18-27 years old). Pilot data (n = 3) confirmed that this sample size allowed for adequate power to detect our effects of interest ($d_z = 3.10$). We opted to collect a large number of measurements from each subject to minimize within-subject variance, which often benefits statistical power more than increased sample sizes (Baker et al., 2021). All subjects reported normal or corrected-to-normal vision and did not report neurological conditions. Procedures were approved by the UCSB Institutional Review Board (IRB# 2-20-0012). All subjects gave written informed consent before participating and were compensated for their time ($20/h for scanning sessions, $10/h for behavioral familiarization/training).

Stimuli and Procedure.

Participants performed a 30-minute training session before scanning so that they were familiarized with the instructions. We used this session to establish the initial behavioral performance thresholds used in the first run of the scanning session. In the main task session, we scanned participants for a single two-hour period consisting of at least 4 mapping task runs, which we used to independently estimate encoding models for each voxel, and 8 experimental feature-salience task runs. All participants also underwent additional anatomical and retinotopic mapping scanning sessions (1-2x 1.5-2 hr sessions) to identify regions of interest (ROI; see Region of interest definition). Additionally, most participants (n = 6) underwent an independent functional localizer session which we used to verify retinotopically defined ROIs were feature selective.

Stimuli were presented using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) for MATLAB (The MathWorks, Natick, MA). Visual stimuli were rear-projected onto a screen placed ~110 cm from the participant's eyes at the head of the scanner bore using a contrast-linearized LCD projector (1,920×1,080, 60 Hz) during the scanning session. In the behavioral familiarization session, we presented stimuli on a contrast-linearized LCD monitor (2,560×1,440, 60 Hz) 62 cm from participants, who were seated in a dimmed room and positioned using a chin rest. For all sessions and tasks (main tasks, localizers, and mapping task), we presented stimuli on a neutral gray circular aperture (9.5° radius), surrounded by black (only aperture shown in Fig. 1).

Feature-salience task.

For the main task (Fig. 2) and functional localizers (see below), participants attended a flashing cross within the fixation circle and ignored any other stimuli presented throughout the scanning session. This task localized goal-directed attention to fixation and was equivalent across all stimulus conditions, allowing us to isolate signals associated with bottom-up salience processing of our peripheral stimuli. Participants monitored the fixation cross throughout the whole run for any increase in length in either the vertical or horizontal bar of the cross and responded to changes with a button press (the left button for a horizontal target, the right button for a vertical target). The vertical and horizontal lines of the fixation cross were 0.25° of visual angle long and flickered at 3 Hz (10 frames on, 10 frames off at 60 Hz). Whenever the cross was visible, there was a 22.5% chance either line had a small change in length. When a change was detected, participants reported which line increased in length (horizontal or vertical). To ensure participants maintained vigilant attention at fixation throughout the entire experiment, they performed an initial behavioral training session where they practiced the fixation task several times. Between runs of the practice session, we adjusted the degree of size change for the vertical/horizontal lines until they consistently achieved ~80% accuracy. We further adjusted the difficulty of the fixation task between runs of the scanning session by altering the degree of size change for vertical/horizontal lines based on behavioral accuracy (range: 0.05° to 0.125°). Participants performed the fixation task continuously throughout both stimulus presentation periods and ITIs to ensure salient events were temporally decoupled from fixation task performance and/or target detection. Feedback for each response to the fixation task was given via the aperture around fixation

changing color for 0.5 s, with green indicating a correct button press response, red indicating an incorrect button press response, and yellow indicating no response (missed a target). Feedback was given 1 s after a target was presented. A fixation target was never present for the first or last 2 s of a trial, or for 2 s after the presentation of a previous target.

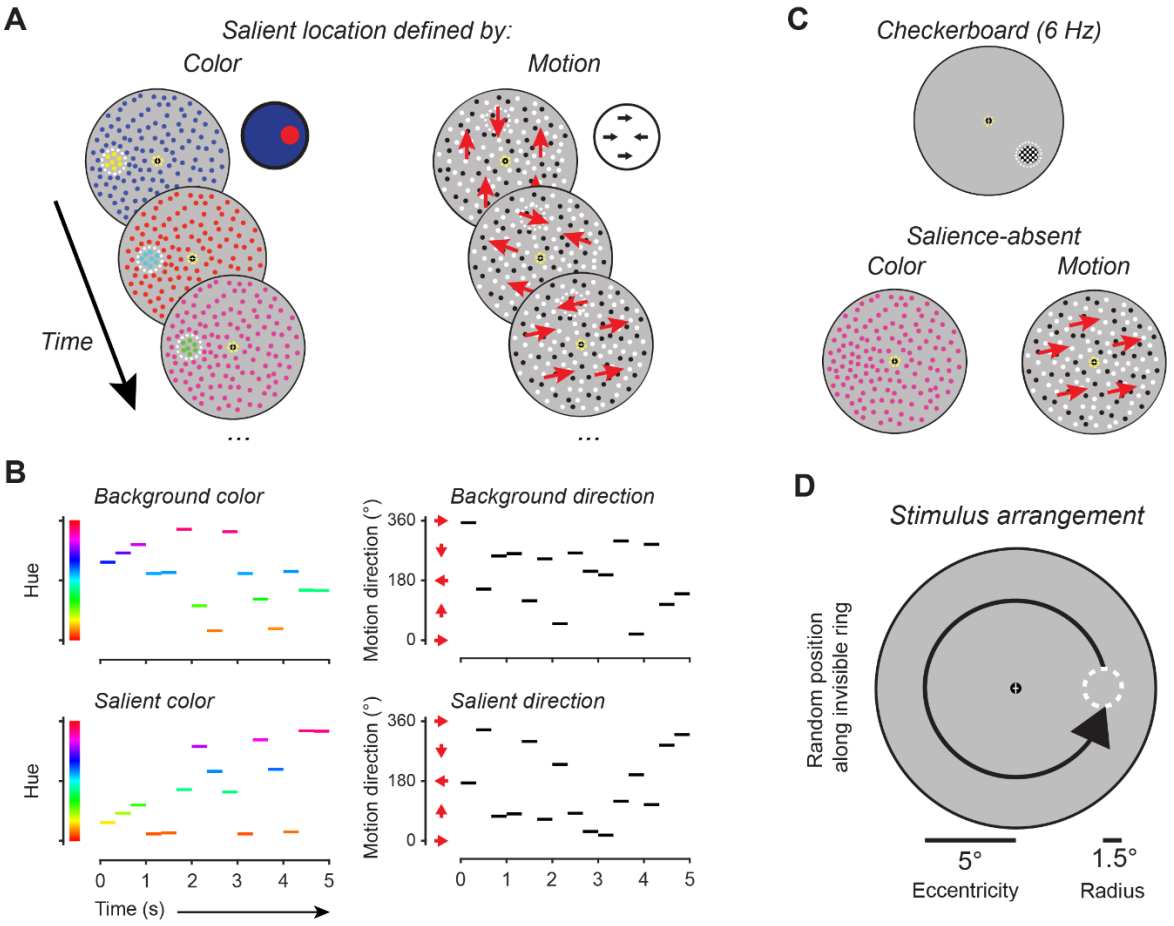

Figure 2. Feature-salience task. A: On each fMRI scanning run, participants continuously performed an attention-demanding task at fixation where they reported changes in length of either the horizontal or vertical bar of the fixation cross. While attention was directed to the demanding fixation task, we measured how feature-selective retinotopic ROIs encode task-irrelevant salient stimulus locations by presenting various types of visual stimuli. On most stimulus presentation trials, the visual stimulus consisted of dots spanning the entire screen. The dot stimuli could either be presented as static colored dots, or grayscale (black/white) moving dots. Subjects received feedback after each target presentation via color changes in the ring around fixation (green = correct button press; red = incorrect button press; yellow = no button press). **B:** The features of all dots were updated at 3 Hz such that, on average, the overall feature value (color/motion) presented across each trial was neutral. For example, on 'color' trials, the color and location of each static dot was updated every 333 ms with a new randomly selected hue and randomly drawn location. On most dot array trials (66.6%; 'salience-present'), a circular portion of the stimulus display was made salient by

presenting dots in the opposite feature value as presented in the background. For example, during a 333 ms period, if the background dots were moving at 45°, the salient foreground dots would be moving 225°. These salient stimulus regions were never relevant for participant behavior, and the challenging fixation task ensured attention was withdrawn from peripheral stimuli. After stimulus presentation, there was a 6-9 s blank ITI during which time the fixation task continued. Example trial for each condition shown. **C:** As control conditions, we also included trials with the salient location defined by a flickering checkerboard (6 Hz full-field flicker) on a blank background, and trials with colored static or moving black and white dots with no salient location. **D:** On salience-present trials, the salient stimulus was 1.5° in radius, and was presented at a location randomly chosen from an invisible ring centered 5° from fixation.

The critical, ignored, stimuli were either a color- or motion-defined salient location presented as a circular disc within random dot arrays spanning the entire stimulus aperture except for a region around fixation (0.75°). On color trials, static dots within a disc were presented in the 'opposite' color (HSV colorspace) as compared to the background dots; on motion trials, moving black and white dots within a disc were presented in the opposite motion direction as compared to the background dots. For example, if the dot array contained dots moving at 0° (to the right), the motion-defined salient location would contain dots moving at 180° (to the left). Similarly, if the dot array contained static colored dots with a red hue (H = 0°), then the color-defined salient location would contain dots with a green hue (H = 180°). Individual dots occupied 0.05° of visual angle, and dot density was 15 dots/deg^2. In the motion array, dots moved at a speed of 9.5° /s in a randomly selected planar direction and each dot was randomly colored black or white (100% contrast). Dots were randomly replotted every 50 ms or when they exceeded the stimulus bounds. For the color array, all dots remained static and were assigned a random hue value. Dot locations were updated every 333 ms. Both arrays updated every 333 ms during the 5 s presentation period, such that a new color or motion value was applied to every dot in the updated array three times per second. Trials started with the onset of the peripheral dot array while participants were attending fixation. The salient location appeared throughout the entire stimulus interval, centered 5° from fixation at a

random location along an invisible ring from 0°-359° and had a radius of 1.5°. While the location of the salient stimulus remained constant on a given trial, it randomly varied between trials along the invisible ring.

We included 3 additional control conditions intermixed with salience-present trials. First, to ensure spatially localized activation is due to the presence of the salient location, we presented colored static dots and moving black and white dots with no salient location defined ('salience-absent' trials). Second, as a positive control to ensure our image reconstruction procedure was effective in each retinotopic ROI, we presented a flickering checkerboard disc (spatial frequency 0.679 cycles/°) on a gray background at the same size, eccentricity, and duration as the salient discs ('checkerboard' trials; similar to previous reports; Sprague & Serences, 2013). The checkerboard stimulus flickered at a rate of 6 Hz and was considered to be feature-agnostic with respect to the key manipulations in the study (i.e., color/motion). All trials were separated by a randomly selected ITI ranging from 6-9 s with an average ITI of 7.5 s. Each run had 24 trials, and during each run, there were 6 trials of each salience-present condition (based on color, based on motion, checkerboard) and 3 trials each of the salience-absent color/motion conditions. Trial order was shuffled within run. Each run started with a 3 s blank period and ended with a 10.5 s blank period, for a total run duration of 313.5 s (for one participant, we acquired 416 TRs for one run instead of 418, resulting in 312 s of data for this run). Eye position was monitored throughout the experiment using an Eyelink 1000 eyetracker (SR Research).

Additionally, we performed a control eyetracking experiment (n = 10, 9 female, age 18-27 years old) outside of the scanner to ensure that our stimuli were sufficiently salient to capture

attention as indexed by saccades directed to salient stimulus locations. Subjects viewed the same displays as the MRI version of the experiment with the following exceptions: each stimulus was presented for 1 s, ITIs were reduced to 3 s, each subject viewed a total of 180 stimuli (36 occurrences of each stimulus condition), and subjects were encouraged to freely view the display with no instructions to perform a fixation task (the fixation task stimulus appeared on-screen, but participants were never instructed to report aspects of the stimulus). The eyetracker sampled right eye gaze position at 500 Hz. Participants performed a 9-point calibration procedure before each run with a viewing distance of 58 cm while seated in a chin and forehead rest.

Spatial mapping task.

We also acquired several runs of a spatial mapping task used to independently estimate a spatial encoding model for each voxel, following previous studies (Sprague et al., 2016; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013). On each trial of the mapping task, we presented a flickering checkerboard at different positions selected from a hexagonal grid spanning the screen. Participants viewed these stimuli and responded whenever a rare contrast change occurred (10 out 47 trials, 21.3%), evenly split between contrast increments and decrements. The checkerboard stimulus was the same size as the salient locations in the feature salience task (1.5° radius) and was presented at 70% contrast and 6-Hz full-field flicker. All stimuli appeared within a gray circular aperture with a 9.5° radius, as in the feature-salience task. For each trial, the location of the stimulus was selected from a triangular grid of 37 possible locations with an added random uniform circular jitter (0.5° radius). The base position of the triangular grid was rotated by 30° on every other scanner

run to increase spatial sampling density. As a result, every mapping trial was unique, which enabled robust spatial encoding model estimation.

Each trial started with a 3000 ms stimulus presentation period. If a target was present, then the stimulus would be dimmed/brightened for 500 ms with the stipulation that the contrast change would not occur in either the first or last 500 ms of the stimulus presentation period. Finally, there was an ITI ranging from 2 to 6 s (uniformly sampled using the linspace command in MATLAB [linspace(2, 6, 47)]). All target-present trials were discarded when estimating the spatial encoding model. Each run consisted of 47 trials (10 of which included targets). We also included a 3 s blank period at the beginning of the run and a 10.5-s blank period at the end of the run. Each run totaled 432 s.

Retinotopic mapping task.

We used a previously reported task (Mackey et al., 2017) to identify retinotopic regions of interest (ROIs) via the voxel receptive field (vRF) method (Dumoulin & Wandell, 2008). Each run of the retinotopy task required participants attend several random dot kinematograms (RDK) within bars that would sweep across the visual field in 2.25 s (or, for one participant, 2.6 s) steps. Three equally sized bars were presented on each step and the participants had to determine which of the two peripheral bars the motion in the central bar matched with a button press. Participants received feedback via a red or green color change at fixation. We used a three-down/one-up staircase to maintain ~80% accuracy throughout each run so that participants would continue to attend the RDK bars. RDK bars swept 17.5° of the visual field. Bar width and sweep direction was pseudo-randomly selected from

several different widths (ranging from 2.0° to 7.5°) and four directions (left-to-right, right-to-left, bottom-to-top, and top-to-bottom).

Functional localizer tasks.

To independently identify color- and motion-selective voxels, we scanned participants while they performed 3 runs each of a color and motion localizer task (Bartels & Zeki, 2000; Huk et al., 2002) using a blocked design. During both tasks, the participant performed the same fixation task from the feature-salience attention task, where they monitored a central cross for changes in the size of the horizontal and vertical lines. In the color localizer task, participants viewed colored or greyscale rectangles of various sizes within the same aperture dimensions described in the spatial mapping task (see spatial mapping task). Stimuli were presented spanning the entire aperture. Rectangle colors were individually sampled from the entire RGB colorspace (uniform independent random distribution of R, G, and B). Similarly, each greyscale rectangle had a randomly sampled contrast (identical R, G, and B value, randomly sampled for each rectangle). During the motion localizer, participants viewed either static or moving black and white dots. For the moving dots, motion could be CW, CCW, or planar (20 evenly spaced steps from 18 to 360°). Dots were redrawn every 100 ms or when they exceeded the stimulus boundary. When the array contained planar motion, dots moved at 1.2°/s; if CW/CCW motion, 0.6°/s. Within each block, each stimulus (static/motion or color/greyscale) was shown for 400 ms followed by a 100 ms blank period before the next stimulus presentation. Each block lasted 18 s (36 updates of stimulus feature values), with feature values randomly selected each presentation. During each scanning run, we presented 6 total blocks, alternating between grayscale rectangles (static dots) and colored rectangles

(moving dots) for the color (motion) localizer runs. At the end of each run, participants viewed a blank screen while performing the fixation task for 18 s. Runs started with a 3 s blank period and ended with a 10.5 s blank period. There was no fixation task during the start and end blank periods. Each run lasted 229.5 s.

We acquired localizer data for 6 of 8 participants (the other 2 participants were unable to return to complete the localizer session).

fMRI acquisition.

fMRI data acquisition and preprocessing pipelines in the current study closely followed a previous report (Hallenbeck et al., 2021) but with slight modifications. We acquired all functional and anatomical images at the UCSB Brain Imaging Center using a 3T Siemens Prisma scanner. fMRI scans for experimental, model estimation, retinotopic mapping, and functional localizers were acquired using the CMRR MultiBand Accelerated EPI pulse sequences. We acquired all images with the Siemens 64 channel head/neck coil with all elements enabled. We acquired both T1- and T2-weighted anatomical scans using the Siemens product MPRAGE and Turbo Spin-Echo sequences (both 3D) with 0.8 mm isotropic voxels, 256 x 240 mm slice FOV, and TE/TR of 2.24/2400 ms (T1w) and 564/3200 ms (T2w). We collected 192 and 224 slices for the T1w and T2w, respectively. We acquired three T1 images, which were aligned and averaged to improve signal-to-noise ratio.

For all functional scans, we used a Multiband (MB) 2D GE-EPI scanning sequence with MB factor of 4, acquiring 44 2.5 mm interleaved slices with no gap, isotropic voxel size 2.5 mm and TE/TR: 30/750ms, and P-to-A phase encoded direction to measure BOLD contrast images. For retinotopic mapping of one participant (sub004), we used a MB 2D GE-EPI

scanning sequence acquired 56 2 mm interleaved slices with isotropic voxel size 2 mm and TE/TR: 42/1300 ms. We measured field inhomogeneities by acquiring spin echo images with normal and reversed phase encoding (3 volumes each), using a 2D SE-EPI with readout matching that of the GE-EPI and the same number of slices, no slice acceleration, TE/TR: 45.6/3537ms (TE/TR: 71.8/6690ms for sub004's retinotopic mapping session).

MRI preprocessing.

Our approach for preprocessing was to coregister all functional images to each participant's native anatomical space. First, we used all intensity-normalized high-resolution anatomical scans (3 T1 images and 1 T2 image for each participant) as input to the 'hi-res' mode of Freesurfer's recon-all script (version 6.0) to identify pial and white matter surfaces. Processed anatomical data for each participant was used as the alignment target for all functional datasets which were kept within each participant's native space. We used AFNI's afni_proc.py to preprocess functional images, including motion correction (6-parameter affine transform), unwarping (using the forward/reverse phase-encode spin echo images), and coregistration (using the unwarped spin-echo images to compute alignment parameters to the anatomical target images). We projected data to the cortical surface, then back into volume space, which incurs a modest amount of smoothing perpendicular to the cortical surface. To optimize distortion correction, we divided functional sessions into 3-5 sub-sessions which consisted of 1-4 fMRI runs and a pair of forward/reverse phase encode direction spin echo images each, which were used to compute that sub-session's distortion correction field. For the feature salience and mapping task, we did not perform any spatial smoothing beyond the smoothing introduced by resampling during coregistration and motion correction. For

retinotopic mapping and functional localizer scans, we smoothed data by 5 mm FWHM on the surface before projecting back into native volume space.

Region of interest definition.

We identified 15 ROIs using independent retinotopic mapping data. We fit a vRF model for each voxel in the cortical surface (in volume space) using averaged and spatially smoothed (on the cortical surface; 5 mm FWHM) time series data across all retinotopy runs (8-12 per participant). We used a compressive spatial summation isotropic Gaussian model (Kay, Winawer, Mezer, et al., 2013; Mackey et al., 2017) as implemented in a customized, GPU-optimized version of mrVista (see Mackey et al., 2017) for detailed description of the model). High-resolution stimulus masks were created (270 x 270 pixels) to ensure similar predicted responses within each bar size across all visual field positions. Model fitting began with an initial high-density grid search, followed by subsequent nonlinear optimization. We visualized retinotopic maps by projecting vRF best-fit polar angle and eccentricity parameters with variance explained ≥10% onto each participant's inflated cortical surfaces via AFNI and SUMA (Fig. 4). We drew retinotopic ROIs (V1, V2, V3, V3AB, hV4, LO1, LO2, VO1, VO2, TO1, TO2, IPS0-3) on each hemisphere's cortical surface based on previously established polar angle reversal and foveal representation criteria (Amano et al., 2009; Mackey et al., 2017; Swisher et al., 2007; Wandell et al., 2007; Winawer & Witthoft, 2015). Finally, ROIs were projected back into volume space to select voxels for analysis. Retinotopically defined ROIs were used for all analyses in the current study.

For primary analyses (Figs. 4 & 6), we aggregated across color-selective maps (hV4, VO1, VO2) and motion-selective maps (TO1, TO2) as reported in previous literature (Albright,

1993; Amano et al., 2009; Brewer et al., 2005; Conway et al., 2007; Huk et al., 2002; Mullen, 2019) by concatenating all voxels for which the best-fit vRF model explained at least 10% of the signal variance prior to univariate or multivariate analyses. For completeness, we also conducted all analyses for each individual ROI using the same voxel selection threshold (see Figs. 5 & 7). We verified that our ROIs based on retinotopic mapping exhibited typical color- and motion-selective responses during our localizer tasks. The motion localizer revealed significant motion-related activation (permuted one-sample T-test comparing activation in response to moving dots to activation in response to static dots) in retinotopic ROIs: TO1/TO2 ($p < 0.001$), hV4/VO1/VO2 ($p = 0.011$). The color localizer identified significant color-related activation (permuted one-sample T-test comparing activation in response to colored rectangles to activation in response to grayscale rectangles) in color-selective retinotopic ROIS hV4/VO1/VO2 (permuted one-sample T-test, $p < 0.001$), but not motion-selective retinotopic ROIs TO1/TO2 ($p = 0.154$).

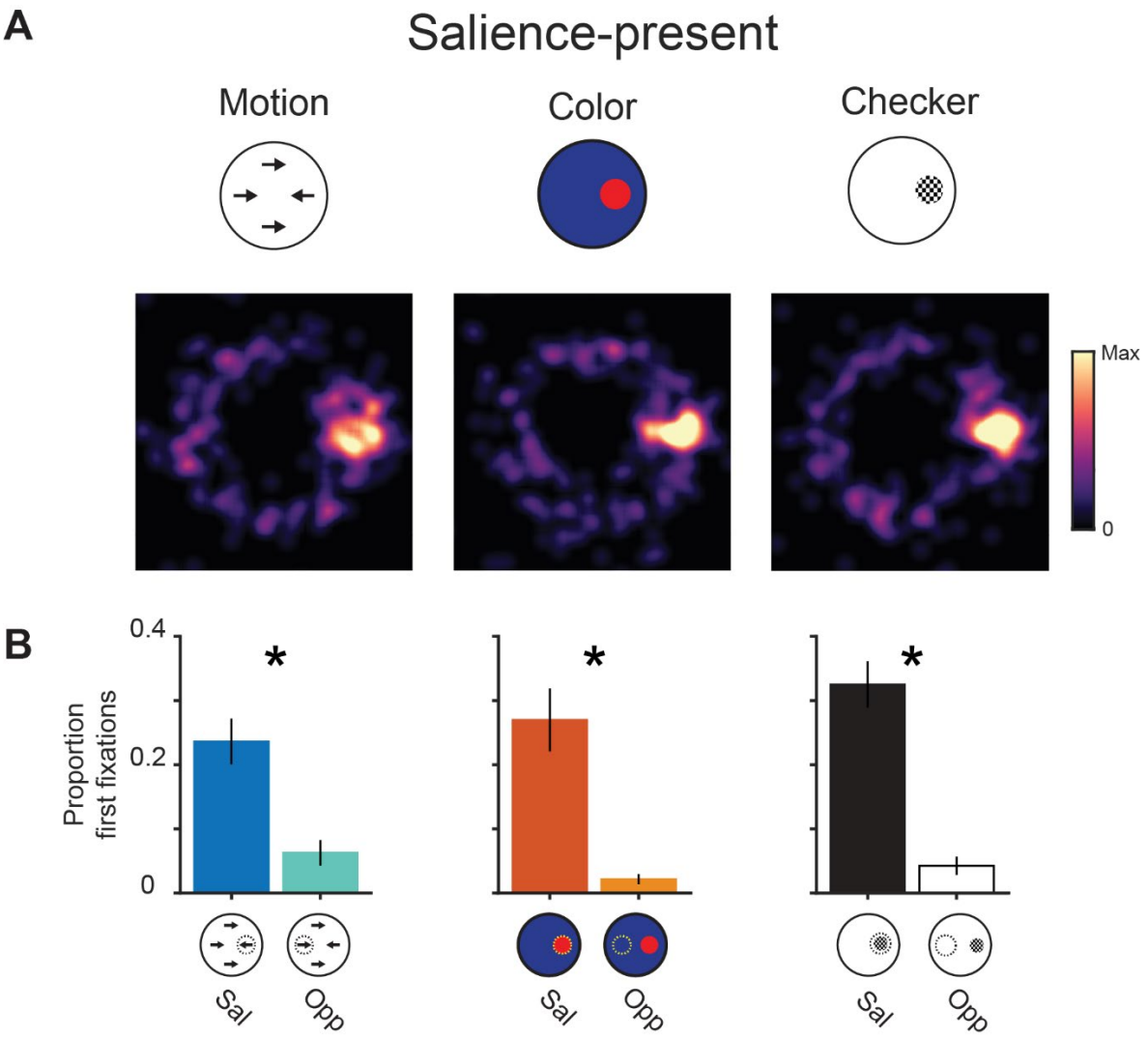

Figure 3. Salient locations in task stimuli are fixated during free-viewing (salience control experiment). **A:** Participants viewed the same stimuli that were presented in the MRI version of the experiment with slight modifications (see Materials & Methods). We plotted the first fixation from each trial of the salient stimulus conditions (salient motion, salient color, and checkerboard). We rotated fixation coordinates on each trial based on the known salient stimulus location (all trials aligned such that the stimulus is at the 0° position). We generated heatmaps by smoothing the aligned 2D fixation histogram (summed across trials) with a 2-D gaussian after removing eye movements near fixation (using Matlab function imgaussfilt: kernel sigma = 0.33°). Fixations within 2° radius of the screen center were excluded from the heatmaps. **B:** We quantified fixation heatmaps by computing the proportion of first fixations directed to the aligned stimulus location and compared this value to the proportion of fixations to the opposite location. Across conditions, the salient stimulus was fixated more than the opposite location. Errors bars reflect SEM across participants. * indicates significant difference based on permuted paired-samples T-test, $p < 0.001$.

Inverted encoding model.

We used a spatial inverted encoding model (IEM) to reconstruct images based on stimulus-related activation patterns measured across entire ROIs (Sprague & Serences, 2013; Fig. 4A). To do this, we first estimated an encoding model, which describes the sensitivity profile over the relevant feature dimension for each voxel in a region. This requires using data set aside for this purpose, referred to as the "training set". Here, we used data from the spatial mapping task as the independent training set. The encoding model across all voxels within a given region is then inverted to estimate a mapping used to transform novel activation patterns from a "test set" (runs from the feature salience task) and reconstruct the spatial representation of the stimulus at each timepoint.

We built an encoding model for spatial position based on a linear combination of 37 spatial filters (Sprague et al., 2014; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013). Each voxel's response was modeled as a weighted sum of each identically shaped spatial filter arrayed in a triangular gird (Fig. 4A). The centers of each filter were spaced by 2.83° and were Cosine functions raised to the 7th power:

$$f(r) = \left(0.5 + 0.5\cos\frac{\pi r}{s}\right)^7 \quad (1)$$

for $r < s$; 0 otherwise

where r is the distance from the filter center and s is a size constant. The size constant reflects the distance from the center of each spatial filter at which the filter returns to 0. This triangular grid of filters forms the basis set, or information channels for our analysis. For each stimulus used in our mapping task, we converted from a contrast mask to a set of filter

activation levels by taking the dot product of the vectorized stimulus mask (*n* trials × *p* pixels) and the sensitivity profile of each filter (*p* pixels × *k* channels). We then normalized the estimated filter activation levels such that the maximum activation was 1 and used this output as $\boldsymbol{C_1}$ in the following equation, which acts as the forward model of our measured fMRI signals:

$$B_1 = C_1 W \tag{2}$$

$\boldsymbol{B_1}$ (*n* trials × *m* voxels) in this equation is the measured fMRI activity of each voxel during the visuospatial mapping task and $\boldsymbol{W}$ is a weight matrix (*k* channels × *m* voxels) which quantifies the contribution of each information channel to each voxel. $\widehat{\boldsymbol{W}}$ can be estimated using ordinary least-squares linear regression to find the weights that minimize the differences between predicted values of $\boldsymbol{B}$ and the observed $\boldsymbol{B_1}$:

$$\widehat{W} = (C_1^T C_1)^{-1} C_1^T B_1 \tag{3}$$

This is computed for each voxel within a region independently, making this step univariate. The resulting $\widehat{W}$ represents all estimated voxel sensitivity profiles. We then used $\widehat{W}$ and the measured fMRI activity of each voxel (i.e., BOLD response) during each trial (using each TR from each trial, in turn) of the feature salience task using the following equation:

$$\hat{C}_2 = B_2 \widehat{W}^T (\widehat{W} \widehat{W}^T)^{-1} \tag{4}$$

Here, $\hat{C}_2$ represents the estimated activation of each information channel (*n* trials × *k* channels) which gave rise to that observed activation pattern across all voxels within a given ROI ($\boldsymbol{B_2}$; *n* trials × *m* voxels). To aid with visualization, quantification, and coregistration of trials across stimulus positions, we computed spatial reconstructions using the output of Equation 4. To do this, we weighted each filter's spatial profile by the corresponding

channel's reconstructed activation level and then summed all weighted filters together (Fig. 4A).

Since stimuli in the feature-selective attention task were randomly positioned on every trial, we rotated the center position of spatial filters such that the resulting 2D reconstructions of the stimuli were aligned across trials and participants (Fig. 4B). We then sorted trials based on condition (salience-present: color, salience-present: motion, checkerboard, salience-absent: color, salience-absent: motion. Finally, we averaged the 2D reconstructions across trials within the same condition for individual participants, then across all participants for our grand-average spatial reconstructions (Fig. 4B-C). Individual values within the 2D reconstructed spatial maps correspond to visual field coordinates. To visualize feature-selectivity within reconstructed spatial maps, we computed the difference in map activation between the salience-present: color and motion conditions (Fig. 6A). We used these difference maps to assess whether feature-selective ROIs had the same feature preferences throughout the visual field, or if they were localized to the position of the salient stimulus when present.

Critically, because we reconstructed all trials from all conditions of the feature-selective attention task using an identical spatial encoding model estimated with an independent spatial mapping task, we can compare reconstructions across conditions on the same footing (Sprague, Adam, et al., 2018; Sprague et al., 2019; Sprague, Itthipuripat, et al., 2018). Moreover, because we were not interested in decoding precision, but instead in the activation profile across the entire reconstructed map, we did not employ any feature decoding approaches and instead opted to directly characterize the resulting model-based

reconstructions (e.g., correlation table; Scotti et al., 2021). Finally, the resulting model-based reconstructions are necessarily based on the modeling choices used here and should not be used to infer any features of single-neuron tuning properties (which we do not claim in this report; Sprague, Adam, et al., 2018; Sprague et al., 2019). Should readers be interested in testing the impact these modeling choices have on results, all analysis code and data are freely available (see below).

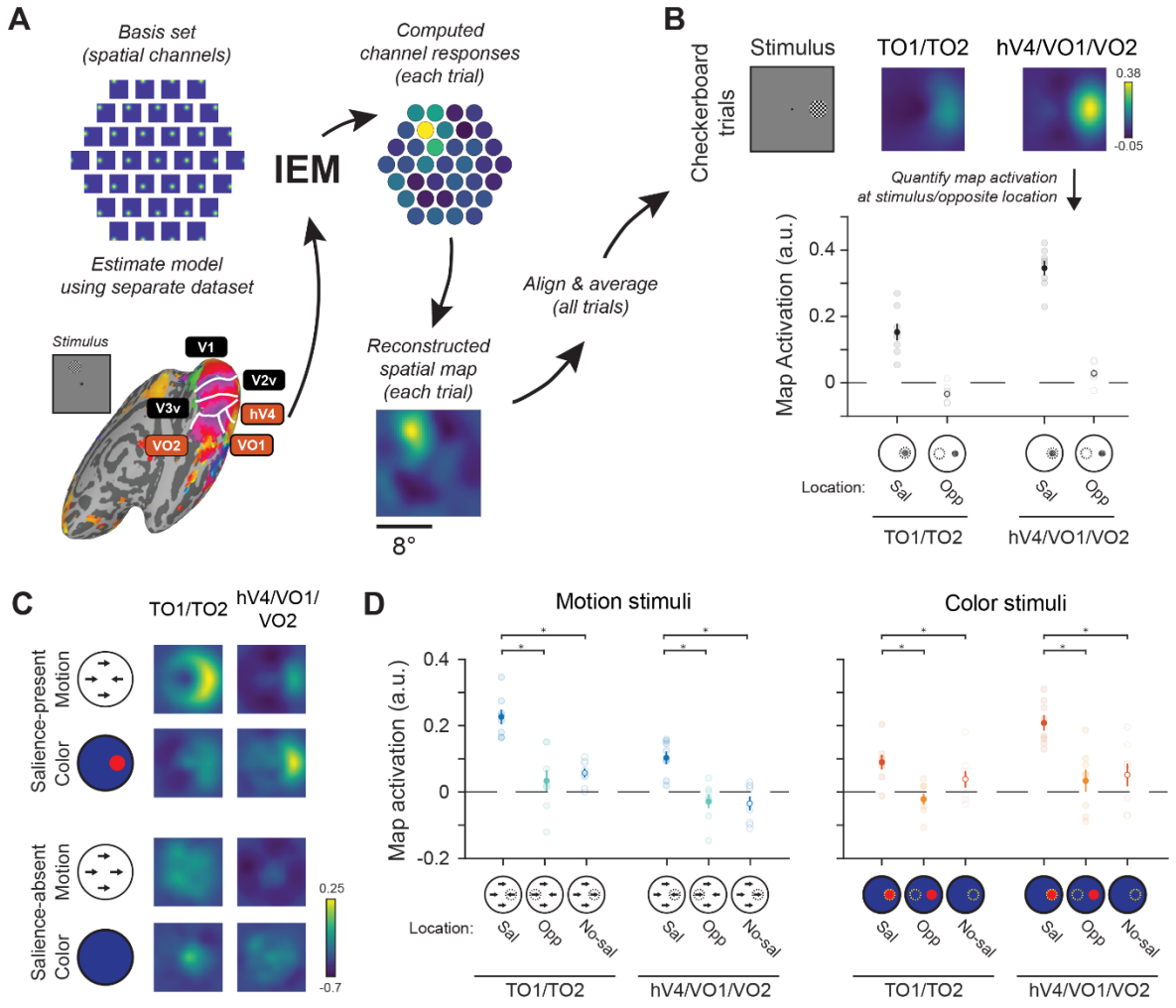

Figure 4. Reconstructed spatial maps track salient stimulus location. A: We estimated a spatial inverted encoding model (IEM) for each ROI using an independent spatial mapping task (see Methods for details). Using this spatial encoding model, which maps activation patterns to activation of spatial channels that can be summed to produce reconstructed spatial maps, we were able to generate image reconstructions of the visual field on each trial and directly compare map activation across conditions. For each condition, we

averaged trial-wise reconstructions computed using activation patterns from 5-8 s after stimulus onset after we rotated and aligned them to the known position of the salient stimulus, if present. **B:** To validate the utility of our method, we computed reconstructions of checkerboard trials from each ROI. Qualitatively, there was a strong response to the checkerboard stimulus across both aggregate ROIs. To quantify reconstructions, we computed the mean map activation at the aligned stimulus location and at the location on the opposite side of fixation within each ROI. In each ROI's reconstruction, activation at the stimulus location was greater than at the opposite location. Two-way permuted repeated-measures ANOVA (ROI; location) identified a significant main effect of location ($p < 0.001$). **C:** Qualitatively, the salient location was highlighted in the aggregate motion ROI TO1/TO2 when the salient location was defined by motion, but not by color, with the converse result observed in the aggregate color ROI hV4/VO1/VO2. On salience-absent trials, no location is reliably highlighted in average reconstructions. **D:** Using data from each stimulus condition, we identified whether enhanced reconstruction responses were localized to the salient stimulus by comparing mean salient location activation ('Sal') to the mean activation of the position opposite of the salient location ('Opp'), as well as the mean activation of the 'aligned' position of the salience-absent condition ('No-sal'). On trials with motion-defined stimuli, activation in the motion-selective ROI was greatest at the location of the salient motion stimulus. When stimuli were defined by static colorful dots, activation in the color-selective ROI was greatest at the location of the salient color stimulus. In all panels, error bars are SEM across participants (n = 8). * indicates significant difference based on permuted paired T-test, corrected for multiple comparisons with FDR. three-way permuted repeated-measures ANOVA (ROI; location; stimulus feature) identified a significant main effect of location ($p < 0.001$), two-way interaction between feature and ROI ($p = 0.007$), and a three-way interaction ($p = 0.001$). For all statistical comparisons, see Table 1.

Figures 4/5	Motion						Color						Checkerboard		
	Salient Location	Opposite Location	Salience-absent	Salient loc vs Salience-absent	Salient loc vs opposite loc	Salience-absent vs opposite loc	Salient Location	Opposite location	Salience-absent	Salient loc vs Salience-absent	Salient loc vs opposite loc	Salience-absent vs opposite loc	Checker location	Opposite location	Checker loc vs opposite loc
V1	0.195	0.290	0.309	**0.002**	**0**	0.768	0.057	0.940	0.869	**0**	0.004	0.656	**0**	0.200	**0**
V2	**0.005**	0.966	0.899	**0**	**0.002**	0.928	*0.018*	0.493	0.277	**0**	**0**	0.268	**0**	0.052	**0**
V3	**0**	0.941	0.828	**0**	**0**	0.674	**0**	0.391	0.093	**0**	**0**	**0.002**	**0**	0.278	**0**
V3ab	**0.002**	0.828	0.999	**0**	**0**	0.842	**0**	0.274	0.186	**0**	**0**	0.718	**0**	0.443	**0**
hV4	*0.028*	0.283	0.235	0.008	**0.008**	0.678	**0**	0.195	0.107	**0**	**0**	0.392	**0**	0.180	**0**
VO1	**0.002**	0.485	0.305	**0**	**0**	0.672	**0**	0.198	0.043	**0**	**0**	0.944	**0**	0.259	**0**
VO2	*0.010*	0.071	0.413	**0.006**	**0**	0.084	**0**	0.313	0.628	**0.01**	**0**	0.022	**0**	0.512	**0**
LO1	*0.014*	0.941	0.546	*0.040*	**0.004**	0.422	*0.034*	0.556	0.437	**0.030**	**0**	0.458	**0.001**	0.079	**0.002**
LO2	**0**	0.652	0.747	**0.002**	**0**	0.542	0.056	0.407	0.227	0.284	0.004	0.068	**0.001**	0.588	**0.004**
TO1	**0**	0.575	*0.038*	**0**	**0**	0.738	*0.013*	0.082	0.614	**0.012**	**0**	0.076	**0.001**	*0.046*	**0**
TO2	**0**	0.253	*0.040*	**0**	**0**	0.532	**0.002**	0.592	0.055	**0.262**	0.066	0.206	**0.001**	**0.017**	**0**
TO1/TO2	**0**	0.325	**0.004**	**0**	**0**	0.408	**0.004**	0.237	0.164	**0.006**	**0.002**	0.054	**0**	**0.009**	**0**
hV4/VO1/VO2	**0.001**	0.194	0.151	**0**	**0**	0.796	**0**	0.345	0.176	**0.008**	**0**	0.546	**0**	*0.024*	**0**
IPS0/1	**0**	*0.028*	0.617	**0.004**	**0**	0.210	**0**	0.575	0.486	**0**	**0**	0.146	**0**	0.075	**0**
IPS2/3	**0.003**	0.215	0.202	**0.004**	**0.002**	0.678	**0**	0.853	0.114	**0**	**0**	0.076	**0**	0.087	**0.002**

Table 1. Statistical tests on multivariate reconstructed spatial maps (related to Figures 4/5). P-values for comparisons at specific locations in reconstructed spatial maps for each ROI. Comparisons include one-sample T-tests for each motion, color, and checkerboard stimulus at the location of the salient stimulus, the opposite location (loc), and the salience-absent condition, and two-sample T-tests between all condition combinations (salience loc vs opposite loc, salience loc vs salience absent, opposite loc vs salience absent). Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$). 0 indicates comparisons where $p < 0.001$. Italicized numbers indicate significance before FDR corrections using $\alpha = 0.05$.

Quantifying stimulus representations.

To quantify the strength of stimulus representations within each reconstruction, we computed the mean map activation of pixels located within a 1.5° radius disk centered at the known position of each stimulus (matching the stimulus radius of 1.5°; see Sprague, Itthipuripat, et al., 2018). This provides a single value corresponding to the activation of the salient stimulus

location for a given condition, within each retinotopic ROI. To assess the spatial selectivity of reconstructed spatial maps, we compared the mean map activation at the location of salient stimuli to map activation at the location opposite fixation using a 1.5° radius disk (Fig. 4B, 4D). Previous studies using a similar IEM approach have used other methods to quantify stimulus reconstructions, such as 'fidelity' (e.g., Sprague et al., 2016). Conclusions using fidelity in the current study were qualitatively and quantitatively consistent with mean map activation results, so we opted to quantify our findings with map activation as it is more intuitive.

To compare values across conditions, we computed a difference in extracted map activation across conditions (e.g., subtracted color map activation from motion map activation, Fig. 6). As an exploratory analysis, we computed these differences at every pixel in reconstructed maps. For quantification, we focused on activation averaged over the discs aligned to the salient stimulus location within each map or the opposite location (Fig. 6C).

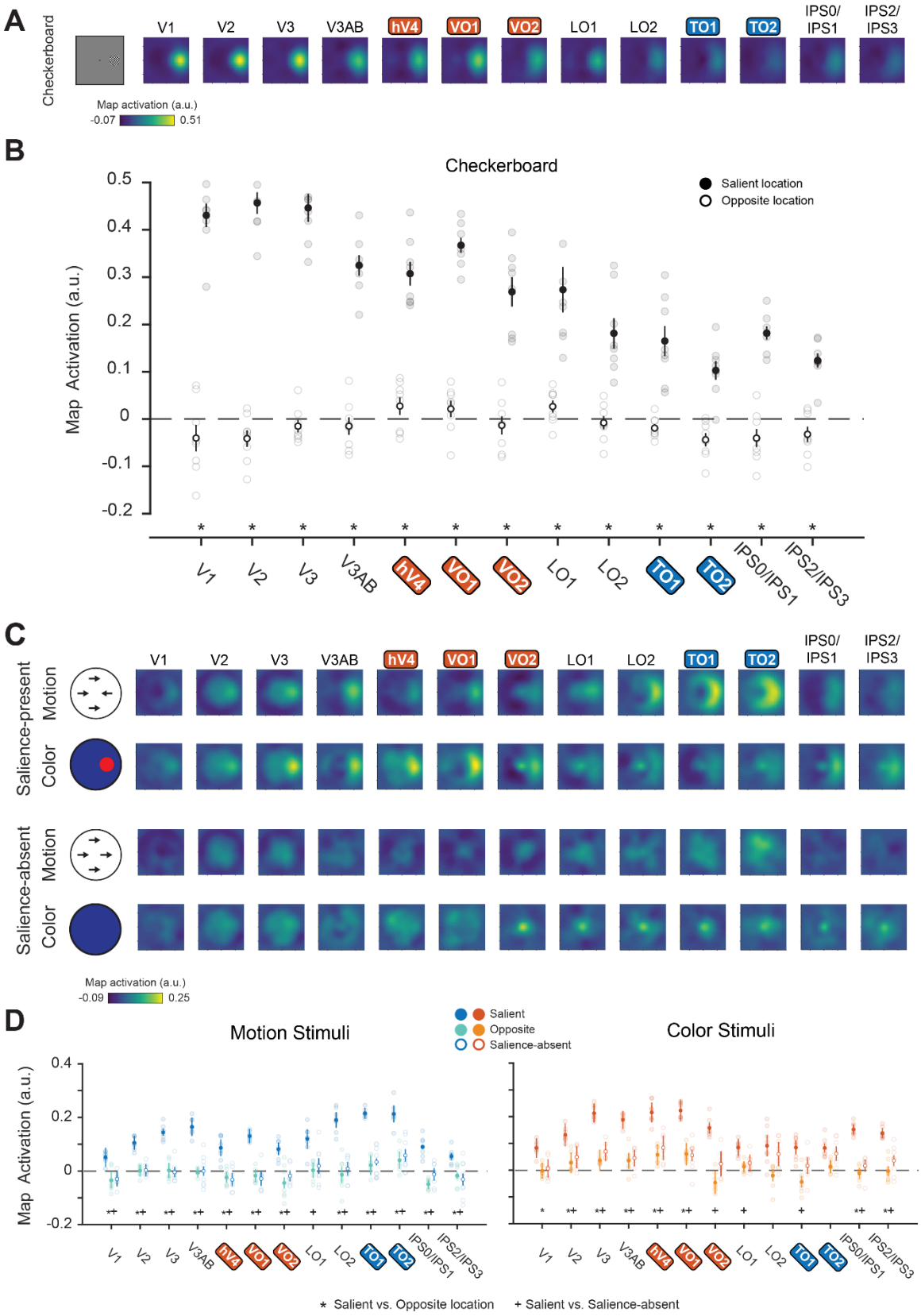
A
Checkerboard
V1
V2
V3
V3AB
hV4
VO1
VO2
LO1
LO2
TO1
TO2
IPS0/
IPS1
IPS2/
IPS3
Map activation (a.u.)
-0.07
0.51
B
Checkerboard
Salient location
Opposite location
Map Activation (a.u.)
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
IPS0/IPS1
IPS2/IPS3
C
Salience-present
Motion
Color
Salience-absent
Motion
Color
Map activation (a.u.)
-0.09
0.25
D
Motion Stimuli
Salient
Opposite
Salience-absent
Color Stimuli
Map Activation (a.u.)
0.4
0.2
0
-0.2
* Salient vs. Opposite location
+ Salient vs. Salience-absent

Figure 5. Reconstructed spatial maps track salient stimulus location across all retinotopic ROIs. A: Reconstructions of checkerboard stimuli from each individual retinotopic ROI. Qualitatively, there was a strong response to the checkerboard stimulus across all ROIs. **B:** We quantified reconstructions by computing the mean map activation at the aligned checkerboard stimulus location and at the location on the opposite side of fixation within each ROI. Permuted 2-way repeated measures ANOVA (ROI and location) showed a significant main effect of ROI ($p < 0.001$), location ($p < 0.001$) and interaction ($p < 0.001$). Map activation was stronger at the location of the checkerboard as compared to the opposite location in all ROIs. * indicates significant difference based on permuted paired T-test, corrected for multiple comparisons with FDR. **C:** Reconstructions were computed for the salience-present and -absent conditions for both features (color/motion; as in Fig. 4). Qualitatively, the salient location was highlighted in feature-selective ROIs when the salient location was defined by motion, but not by color, with the converse result observed in color-selective ROIs. On salience-absent trials, no location is reliably highlighted in average reconstructions. **D:** Using data from each stimulus condition, we identified whether enhanced reconstruction responses were localized to the salient location by comparing mean salient location activation to the mean activation of the position opposite of the salient location, as well as the mean activation of the 'aligned' position of the salience-absent condition. On trials with motion-defined stimuli, activation in motion-selective ROIs was greatest at the location of the salient motion patch. When stimuli were defined by static colorful dots, activation in the color-selective ROIs was greatest at the location of the salient color stimulus. * indicates significant difference between salient and opposite locations, and + indicates significant difference between the salient location and salience-absent reconstructions. Both sets of tests are based on permuted paired T-test and corrected for multiple comparisons with FDR. 3-way permuted repeated-measures ANOVA (ROI; location; stimulus feature) identified a significant main effect of location ($p < 0.001$), 2-way interaction between feature and ROI ($p < 0.001$), and a 3-way interaction ($p < 0.001$). Individual feature-selective ROIs are highlighted (color-selective regions = red; motion-selective regions = blue). For all statistical comparisons, see Table 1.

Visualizing and quantifying stimulus salience using gaze position.

To ensure that our stimuli were able to capture attention in the absence of an instructed fixation task, we analyzed the eye position data from the salience control experiment. We first generated gaze heatmaps for each of the salience conditions using gaze fixation data extracted using a velocity threshold of 22°/s and an acceleration threshold of 3800°/s^{2}s. We plotted the *x* and *y* positions of each fixation, rotated fixations based on the known location of the salient stimulus on each trial, and then smoothed the maps with a 2-D gaussian kernel using the Matlab function imgaussfilt (Fig. 3A; kernel sigma = 0.33°). Fixations that were within 2° of the central fixation point were excluded from the heatmaps. A behavioral index of stimulus salience was quantified by computing the proportion of first fixations on each trial that landed at the salient stimulus location (1.5° radius disk at 5.0° eccentricity) to

fixations to the opposite location (1.5° radius disk at -5.0° eccentricity, where 0° was center; Fig. 3B).

Statistical analysis.

We used parametric statistical tests for all comparisons (repeated-measures ANOVAs and T-tests). To account for possible non-normalities in our data, we generated null distributions for each test using a permutation procedure (see below) to derive p-values.

First, we used a one-way repeated-measures ANOVA (factor: stimulus condition; 5 levels: motion salience, color salience, checkerboard, non-salient motion, and non-salient color) to determine whether behavioral performance on the fixation task depended on the type of ignored peripheral stimulus presented on each trial. All behavioral analyses only used fixation target trials that occurred during the stimulus presentation period, as this was the trial period of interest in neuroimaging analyses. For the salience control experiment (Fig. 3), we compared the proportion of first fixations with a two-way repeated measures ANOVA with factors of stimulus condition (3 levels: motion salience, color salience, checkerboard) and location (2 levels: salient location, opposite location). To confirm that the salient location captured attention, we then performed follow-up paired samples T-Tests between proportion of fixations to the aligned salient and opposite locations (Fig. 3B). Note that the sum of first fixations across salient and opposite locations does not sum to 1 because participants could fixate other locations on the screen.

For all primary fMRI analyses, we focused on sets of retinotopically defined regions selected *a priori* based on previous reports establishing feature selectivity for color or motion (see above). Additionally, for completeness, we repeated these tests across each individual

retinotopic ROI (Figs. 5 & 7). To determine the spatial selectivity of reconstructed spatial maps based on fMRI activation patterns, we computed a three-way repeated measures ANOVA to determine the spatial selectivity of neural modulations with location activation (salient location, opposite location, and aligned position in salience-absent conditions), ROI, and feature (motion/color) as factors (Figs. 4D & 5D). To directly test whether feature-selective ROIs represent salient locations more strongly when salience is defined by their preferred feature value, we computed a paired-samples T-Test on the difference between map activation on color-salience and motion-salience trials between color-selective and motion-selective ROIs (Fig. 6). We compared the same difference across all individual ROIs using a one-way repeated-measures ANOVA with ROI as factor (Fig. 7C). Finally, we assessed the spatial selectivity of feature-selective responses by computing a two-way ANOVA with ROI and location (salient location, location opposite to salient stimulus, and aligned position in salience absent condition) as factors (Fig. 6B; Fig. 7B).

For our shuffling procedure, we used a random number generator that was seeded with a single value for all analyses. The seed number was randomly selected using an online random number generator (https://numbergenerator.org/random-8-digit-number-generator). Within each participant, averaged data within each condition were shuffled across conditions for each participant individually, and once shuffled, the statistical test of interest was recomputed over 1000 iterations. *P* values were derived by computing the percentage of shuffled test statistics that were greater than or equal to the measured test statistic. We controlled for multiple comparisons using the false discovery rate (Benjamini & Yekutieli,

2001) across all comparisons within an analysis when necessary. Error bars are standard error, unless noted otherwise.

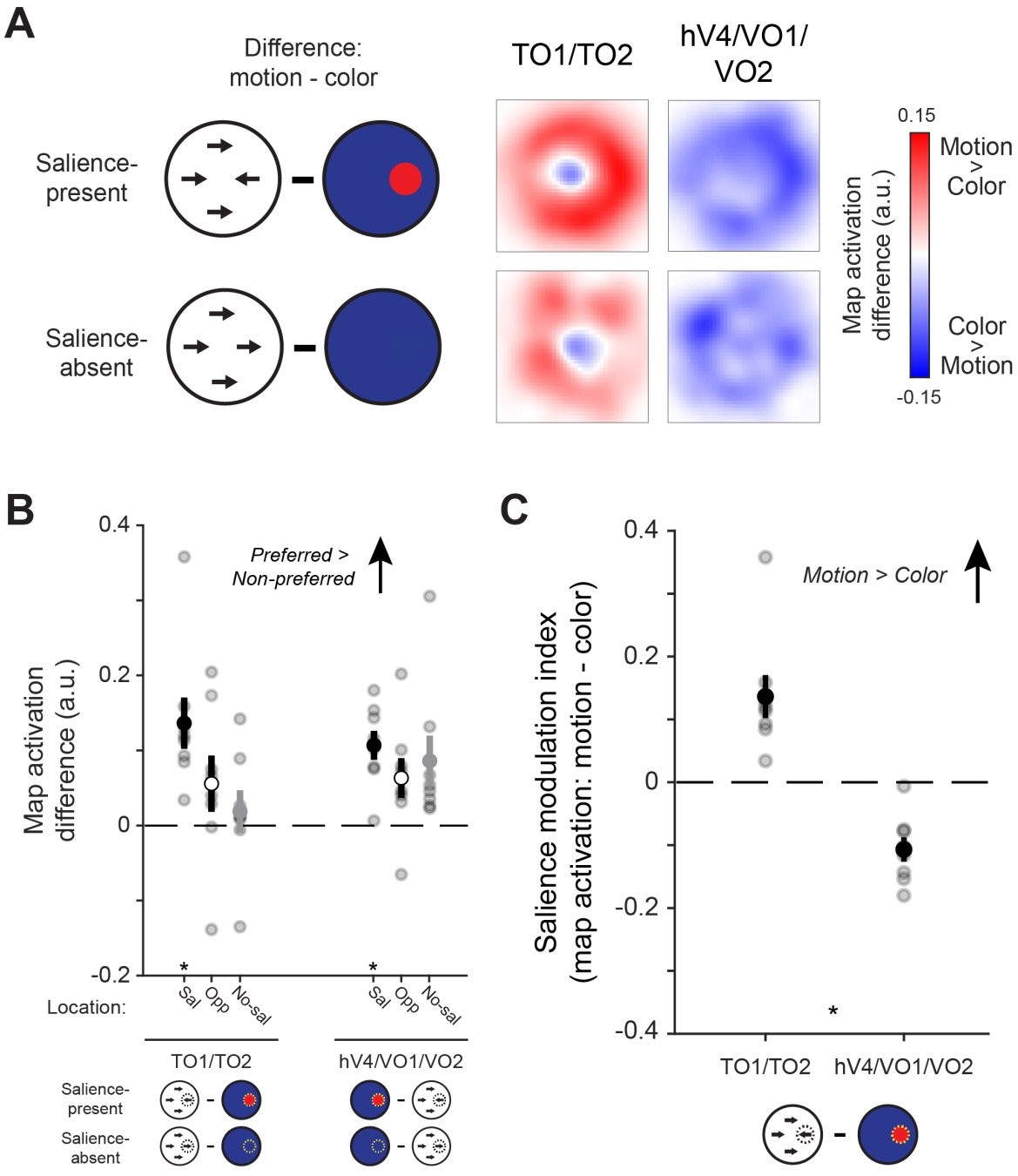

Figure 6. Neural dimension maps selectively index salience based on their preferred feature. A: We directly compared reconstructed spatial maps for each ROI between trials where the salient location was defined by motion to those where the salient location was defined by color by computing their pixelwise difference (motion – color). For comparison, we also computed these maps for salience-absent trials.

Positive values indicate a region more strongly represents a location based on motion stimuli over color stimuli, and negative values indicate the opposite. **B:** To compare feature selectivity across spatial locations and salience-presence conditions, we extracted values of each ROI's difference map computed between the preferred and non-preferred feature dimension for that ROI (see cartoons). Difference map activation was more positive (preferred > non-preferred) at the salient location than the opposite location on salience-present trials, and more positive than a random location on salience-absent trials (two-way permuted repeated-measures ANOVA with factors of location and ROI; significant main effect of location $p = 0.002$). Additionally, difference map values were only reliably greater than zero at the aligned position on salience-present trials (one-sample T-tests against zero, FDR corrected, $p \leq 0.005$), indicating that these ROIs preferentially encode salient locations based on their preferred feature dimension. * indicates significant difference from zero. **C:** Difference map activation (A) computed at the salient location reliably differed between ROIs, such that motion-selective TO1/TO2 indexed salience more strongly when it was defined by motion than by color, and vice versa for color-selective hV4/VO1/VO2. Asterisk indicates significant difference based on permuted paired-samples T-test, $p < 0.001$. Error bars reflect SEM across participants.

Data & Code Availability.

All data supporting the conclusions of this report and all associated analysis scripts available on Open Science Framework (https://osf.io/wkb67/). To protect participant privacy, and in accordance with IRB-approved procedures, freely available data is limited to extracted timeseries for each voxel of each ROI for all scans of the study. Whole-brain 'raw' data will be made available from the authors upon reasonable request.

Results

Behavior.

Participants continuously monitored a central fixation cross for brief changes in line segment length while viewing stimuli in the periphery (Fig. 2). Stimuli could either be full-screen arrays of colored static or grayscale moving dots, or flickering checkerboards. When stimuli were dot arrays, they typically contained a salient region (colored dots: a disc appeared in a different color, 180° away in HSV colorspace; moving dots: a disc contained dots moving in the opposing motion direction). This design requires attention to be maintained at fixation, and allows for bottom-up salience to be isolated and evaluated in response to the ignored,

task-irrelevant peripheral stimulus as a function of the salience-defining feature. Across runs we adjusted fixation task difficulty to keep behavioral performance above chance and below ceiling and to maximize participant engagement (average response accuracy across conditions: 81.07% % ± 2.13%, mean ± SEM across participants; average miss rate across conditions: 26.49% ± 9.12%). Importantly, we observed no difference in response accuracy ($p = 0.7$; one-way permuted repeated-measures ANOVA) or miss rate ($p = 0.29$; one-way permuted repeated-measures ANOVA) as a function of peripheral stimulus type. Thus, any differences observed in multivariate activation patterns between conditions cannot be driven by differences in behavioral performance.

To verify that our stimuli were behaviorally salient in the absence of a demanding fixation task, we acquired eyetracking data outside the scanner from naïve participants who were encouraged to freely view the display while viewing the same dot arrays that appeared in the MRI version of the task. The first fixation after the appearance of the stimulus array was most commonly directed to the salient location (Fig. 3). A two-way permuted repeated-measures ANOVA with salience condition (salient motion, salient color, and checkerboard) and activation location (salient location and opposite location) as factors showed only a significant main effect of location ($p < 0.001$). Follow-up comparisons show that the first fixation of each trial was more likely to be directed to the salient location than the opposite location for the salient motion ($p = 0.006$), color ($p = 0.002$), and checkerboard ($p < 0.001$) conditions. These results verify that our stimuli were sufficiently salient to capture attention even though no behavioral differences were observed in the demanding fixation task conducted during the MRI sessions.

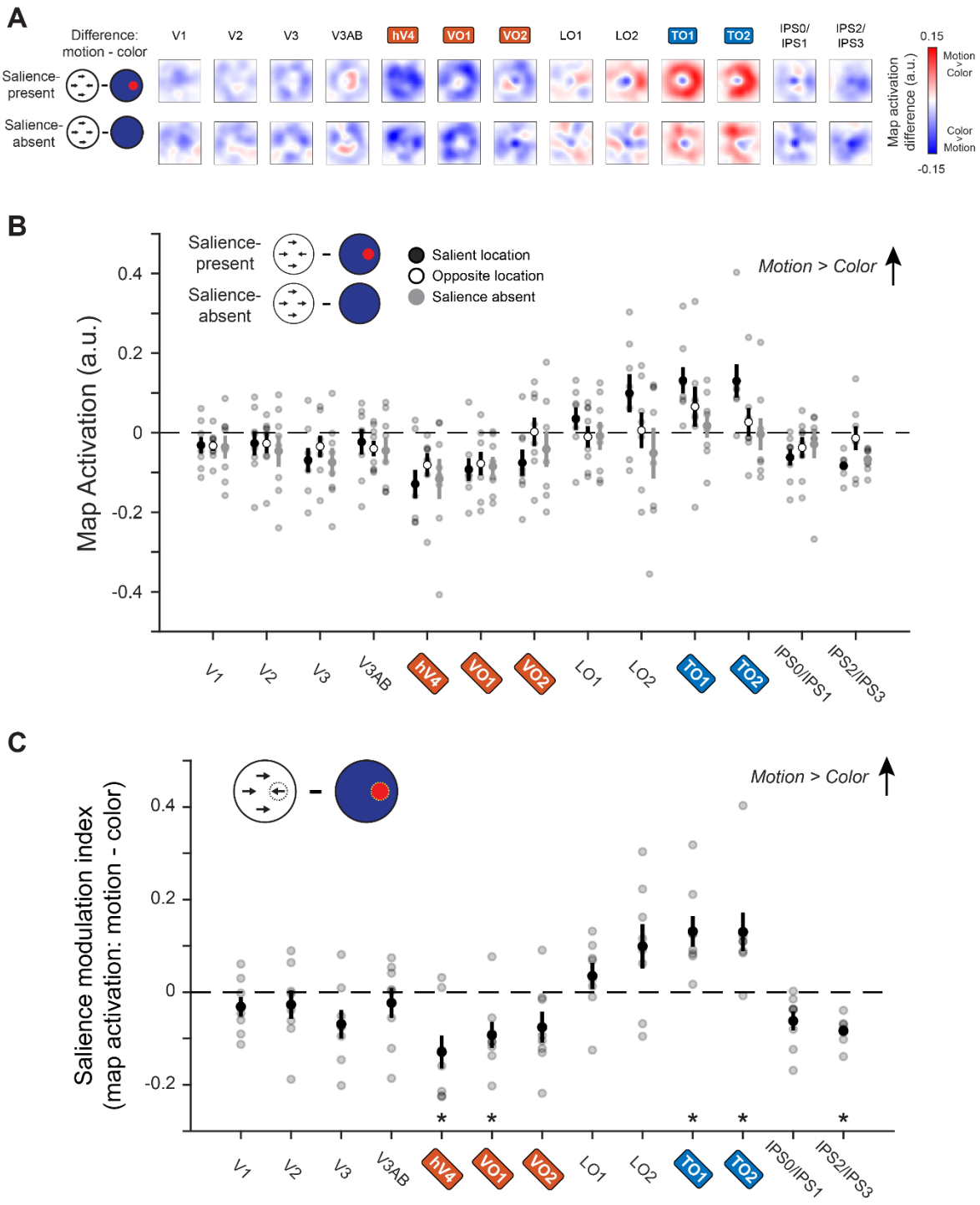

Figure 7. Individual feature-selective ROIs index salience based on their preferred feature. A: We directly compared reconstructed spatial maps for each ROI between trials where the salient location was defined by motion to those where the salient location was defined by color by computing the pixelwise difference (motion – color). For comparison, we also computed these maps for salience-absent trials. Positive values indicate a region more strongly represents a salient location based on motion over color, and negative values indicate the opposite. Data presented as in Fig. 6A. **B:** To compare feature selectivity across

spatial locations and salience-presence conditions, we extracted values of each ROI's difference map at the salient location, opposite location, and the salience-absent condition. Absolute difference map activation was greater at the salient location than the opposite location on salience-present trials than at a random location on salience-absent trials, and this effect depended on ROI (2-way permuted repeated-measures ANOVA with factors of location and ROI; significant main effect of ROI $p < 0.001$ and interaction $p < 0.001$). **C:** Difference map activation (A) computed at the salient location was only reliably different from 0 in feature-selective ROIs, such that motion-selective regions indexed salience more strongly when it was defined by motion than by color, and vice versa for color-selective regions. Asterisk indicates significant difference based on permuted one-sample T-test, $p < 0.05$. * indicates significance difference from zero. Error bars reflect SEM across participants. Individual feature-selective ROIs are highlighted (color-selective regions = red; motion-selective regions = blue). See Table 2 for all statistical tests.

Figures 6/7	Salient Location	Opposite Location	Salience absent	Salient loc vs opposite loc	Salient loc vs salience-absent	Opposite loc vs salience-absent
V1	0.176	*0.041*	0.252	0.938	0.788	0.894
V2	0.415	0.318	0.282	0.948	0.472	0.556
V3	0.061	0.246	0.079	0.244	0.866	*0.046*
V3ab	0.496	0.081	0.219	0.480	0.590	0.83
hV4	*0.009*	*0.029*	0.055	0.178	0.820	0.322
VO1	*0.014*	*0.034*	*0.010*	0.508	0.736	0.548
VO2	0.057	0.954	0.385	0.096	0.554	0.176
LO1	0.253	0.713	0.824	0.062	0.208	0.944
LO2	0.077	0.902	0.444	**0.004**	*0.016*	0.316
TO1	*0.005*	0.233	0.576	0.106	**0.006**	0.350
TO2	*0.017*	0.476	0.927	**0.002**	**0.002**	0.260
TO1/TO2	**0.005**	0.180	0.529	**0.016**	**< 0.001**	0.174
hV4/VO1 /VO2	**0.001**	*0.049*	*0.038*	0.074	0.556	0.388
IPS0/1	*0.020*	0.194	0.434	0.318	0.188	0.81
IPS2/3	**< 0.001**	0.667	**< 0.001**	*0.030*	0.354	0.106

Table 2. Feature-selective map activation statistical tests (related to Figures 6/7). P-values for comparisons at specific locations within feature-difference spatial maps for each ROI. Comparisons include one-sample T-tests at the location of the salient stimulus, the opposite location, and the salience-absent maps, and paired-sample T-tests for all location combinations (salience vs opposite, salience vs absent, opposite vs absent). Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$). Italicized numbers indicate significance before FDR corrections using $\alpha = 0.05$.

Multivariate spatial representations.

Next, we used a spatial inverted encoding model (IEM) to reconstruct spatial maps based on measured activation patterns from each ROI on each trial. We used data from an independent 'mapping' task (see methods; Sprague et al., 2018) to estimate a spatial encoding model for each voxel parameterized as a set of weights on smooth, overlapping spatial channels. Then, we inverted the set of encoding models across all voxels in each cluster of regions to reconstruct spatial maps based on activation profiles from the feature-salience task (Fig. 4A). This procedure generates a reconstructed image for each timepoint, which we then averaged within condition and across timepoints corresponding to 5 to 8 s after stimulus onset. The resulting images are well-established to show strong activation at locations corresponding to visual stimulation (e.g., where a checkerboard was presented; Sprague et al., 2019; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013). Indeed, when the 'checkerboard' trials were used for stimulus reconstruction, we observed strong representations of the salient stimulus location in all ROIs (Fig. 4B).

Having validated our method, we asked: does activation within these reconstructions additionally track locations made salient based on local differences in feature values? In our study, the entire visual field is equivalently stimulated (e.g., equal amount of motion energy or density of colored dots), so any nonuniform activation must be due to salience-related activation patterns within each ROI. Indeed, model-based reconstructions were able to track the salient location throughout the visual field on both motion- and color-salient trials (Fig. 4C). When the salient location was defined by dots moving in a different motion direction,

the reconstructed spatial map from TO1/TO2 showed a stronger representation of the salient location than when the salient location was defined by static dots presented in a different color, and the converse result is apparent when examining spatial maps reconstructed from hV4/VO1/VO2 (Fig. 4C). Critically, in these trials, the feature values at each location are updated at 3 Hz, minimizing the possibility that reconstructions of salient locations emerge from a serendipitous selection of a specific local feature value.

To quantify the condition-specific stimulus representation within model-based reconstructions for each ROI, we computed the mean activation at the known position of the salient stimulus (see Sprague et al, 2018; Fig. 4B). For ROIs which compute spatial maps of salient location(s), we predict that reconstructions will show an enhanced representation of the salient stimulus position when compared to non-salient locations. Our design allows for two important comparisons to establish whether these salience computations occur. First, we can directly compare the activation in reconstructed spatial maps at the salient location to activation in the location on the opposite side of the fixation point (which contains non-salient 'background' dots with an equal amount of color/motion energy as the salient location). This comparison allows us to demonstrate that spatial maps highlight salient locations within each salient stimulus condition. Second, we can compare the mean activation of the salient location on salience-present trials to a randomly selected location of the reconstructed spatial map on salience-absent trials. This allows us to see if the map activation at the salient location was greater than map activation at an equivalent spatial location when viewing a uniform dot array with no salient position(s).

Map activation values were strongest at the salient location when a salient stimulus was presented, and weaker at non-salient locations (both when a salient stimulus was presented elsewhere, and when no salient stimulus was present at all; Fig. 4D). We compared map activation values across conditions, map locations, and ROIs using a three-way repeated-measures permuted ANOVA with stimulus feature (motion/color), activation location (salience-absent; salience-present: salient location; salience-present: opposite location), and ROI (TO1/TO2; hV4/VO1/VO2) as factors. This analysis indicated that there was a main effect of activation location ($p < 0.001$), a two-way interaction between stimulus feature and ROI ($p = 0.007$), and a three-way interaction between all three factors ($p = 0.001$). All other comparisons were non-significant ($p > 0.05$).

Within hV4/VO1/VO2, we observed a significant difference between map activation at the salient location and opposite location on color salience-present trials ($p < 0.001$, permuted paired samples T-test; Fig. 4D) as well as a significant difference between the color-defined salient location and map activation on salience-absent colored dot trials ($p = 0.008$, permuted paired samples T-test). Map activation at the salient location was significantly greater than zero ($p < 0.001$; permuted one-sample T-test).

We found complementary results in TO1/TO2 when using data from the motion stimulus conditions (Fig. 4D). We observed a significant difference in map activation between the salient location defined by motion and the opposite location on salience-present trials ($p < 0.001$, permuted paired samples T-test) in addition to a significant difference between the map activation at the salient location and map activation from similar locations on trials when no salient location was defined ($p < 0.001$, permuted paired samples T-test). Map

activation at salient locations defined by motion was also greater than zero ($p < 0.001$; permuted one-sample T-test). Altogether, these results suggest that activation patterns in these regions reflect the image-computable salience of the corresponding location in the visual field.

Comparing salience computations across feature dimensions.

Thus far, we have shown that color- and motion-selective regions each compute a representation of the location of a salient stimulus defined by feature contrast. If these regions act as neural dimension maps which each individually compute representations of salient locations defined by their preferred feature value, we expect to observe a more efficient extraction of salient locations when the salience-defining feature matches the region's preferred feature value. We tested this by computing a pixelwise difference between reconstructed spatial maps from each ROI when salience was defined based on color and when salience was defined based on motion (Fig. 6A), along with the same difference between the salience-absent control conditions. Values near zero (white) indicate that the map activation is equal between stimulus features, while positive (red)/negative (blue) values near the salient location indicate that the map preferentially extracts salient locations when the salience defining feature is motion or color, respectively. If each ROI selectively identifies salient locations based on its preferred feature value, then these difference maps will show greater absolute differences at the salient location than other locations. However, if instead feature selectivity and salience computations each independently and additively impact spatial maps, then these difference maps should show no spatial structure (particularly, no difference between salient and non-salient locations).

Qualitatively comparing these difference maps (Fig. 6A), it is apparent that both ROIs show a stronger difference at the salient location than the opposite location, consistent with a local and specialized computation of salient locations based on each region's preferred feature values. The localized differences at the salient location are in opposite directions (TO1/TO2: positive/red; hV4/VO1/VO2: negative/blue), as expected if motion (color)-selective ROIs more efficiently extract salient locations defined by motion (color) feature contrast than color (motion) feature contrast.

We quantified the degree to which each ROI selectively computes salience based on its preferred feature value by extracting activation values from these difference maps at the salient location, the opposite location, and the 'aligned' location in the salience-absent condition and computing map activation difference scores based on the regions' feature preferences (TO1/TO2: motion – color; hV4/VO1/VO2: color – motion; Fig. 6B). A two-way repeated measures ANOVA with location (salient location, opposite location, salience-absent) and ROI as factors revealed a significant main effect of location ($p = 0.002$). Follow-up comparisons in TO1/TO2 demonstrate that map activation was greater at the salient location than the opposite location ($p = 0.016$, permuted paired samples T-test) and the salience-absent location ($p < 0.001$, permuted paired samples T-test) and was the only location with an activation difference greater than zero ($p = 0.005$, permuted one-sample T-test). In hV4/VO1/VO2, the same tests did not reveal significant differences between the salient location and the opposite location ($p = 0.074$, permuted paired samples T-test) or the salient-absent position ($p = 0.556$, permuted paired samples T-test). However, only the salient location had map activation differences greater than zero after FDR corrections ($p =$

0.001, permuted paired samples T-test). Together, these results suggest that each ROI selectively indexes a salient location when salience is defined based on its preferred feature value.

Finally, to directly compare salience computations between regions, we computed a salience modulation index (SMI). This was defined as the difference between map activation at the salient location between the motion and color conditions, where positive values indicate a stronger response to the motion-defined salient location, negative values indicate a stronger response to the color-defined salient location, and zero indicates no difference between conditions. SMI reliably differed between motion- and color-selective ROIs (Fig. 6C; $p < 0.001$; permuted paired- samples T-test). This indicates that feature-selective ROIs preferentially compute salience based on their preferred feature dimension, and further supports the proposal that these retinotopically- defined regions act as neural dimension maps within priority map theory.

Discussion

In the present study, our goal was to determine whether visual cortex computes spatial maps representing salient locations based on specific feature dimensions within feature-selective retinotopic regions (Fig. 1)—a key prediction of priority map theory. We probed this question by reconstructing spatial maps based on fMRI activation patterns measured while participants viewed, but ignored, stimuli containing salient locations based on different feature dimensions (Fig. 2). Our results show that salient location representations in color-selective regions hV4/VO1/VO2 and motion-selective regions TO1/TO2 are modulated by bottom-up feature salience even when top-down attention is kept at fixation (Fig. 4).

Representations were selectively enhanced when the salience-defining feature matched the preferred feature of a given region and these enhancements occurred at the *salient location* (Fig. 6). These results provide strong evidence that these retinotopic cortical regions act as 'neural feature dimension maps', confirming an important aspect of priority map theory.

In previous studies in humans and nonhuman primates, neural correlates of salience and/or priority maps have been identified in several regions, including: LGN (Kastner et al., 2006; Poltoratski et al., 2017), V1 (Li, 2002; Poltoratski et al., 2017; L. Wang et al., 2022; X. Zhang et al., 2012), extrastriate visual cortex (including V4/hV4; Adam & Serences, 2021; Bogler et al., 2011, 2013; Burrows & Moore, 2009; Mazer & Gallant, 2003; Poltoratski et al., 2017; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013), LIP/IPS (Adam & Serences, 2021; Bisley & Goldberg, 2003, 2006; Chen et al., 2020; Gottlieb et al., 1998; Jerde et al., 2012a; Sprague, Itthipuripat, et al., 2018), FEF (Bichot & Schall, 1999; Schall et al., 1995; Schall & Hanes, 1993), substantia nigra (Basso & Wurtz, 2002), pulvinar (Shipp, 2003), and SC (Basso & Wurtz, 1998; Fecteau & Munoz, 2006; B. J. White et al., 2017). Across these studies, activity in neurons/voxels tuned for salient and/or relevant locations is greater than activity in neurons/voxels tuned towards nonsalient and/or nonrelevant locations. However, many of these previous studies are limited by focusing on a single salience-defining feature and/or by relying on sparse single-unit recordings from one or a handful of cells within one or a small number of brain regions in non-human primates. Such studies necessarily face difficulty comparing across different brain regions and assessing activation profiles across the entire region (and thus, the entire visual field). Here, we overcame these limitations by implementing a multivariate IEM which allowed us to reconstruct activation

profiles across a map of the entire visual field from each timepoint's measured activation pattern in each ROI (Fig. 3A). Additionally, by manipulating the salience-defining stimulus feature dimension across trials (Fig. 2) while simultaneously measuring fMRI activation patterns across multiple feature-selective ROIs (Figs. 4C-D & 5C-D), we established that the region best representing a salient location depends on the salience-defining feature (Figs. 6-7).

As mentioned above, there is extensive, and often conflicting, evidence for priority maps implemented in different structures throughout the brain. With the seemingly redundant computations of priority across regions, how is information across maps ultimately leveraged to guide attention? We expect measurements of feature dimension maps like those identified in the current study can be used to disentangle these various accounts by establishing which regions act to integrate information about salient locations based on combinations of features. One testable prediction of the priority map framework is that activation profiles in a feature-agnostic priority map should reflect some computation over the activation profiles measured across individual feature dimension maps, such as: linear combination (Wolfe, 1984), winner-take-all processes (Itti & Koch, 2001), and probabilistic integration (Eckstein, 2017). While such a test is not possible in our current study, future work incorporating stimuli with multiple salient locations with different degrees of stimulus-defined salience may better disentangle the roles of various putative priority maps in guiding visual attention based on stimulus properties.

Aspects of priority map theory are invoked by foundational models of behavioral performance on visual search tasks (Duncan & Humphreys, 1989; Folk et al., 1992; Müller et

al., 1995; Theeuwes, 2010). For example, when a subject is tasked with searching for one shape among a homogeneous array of other shapes (e.g., a circle among squares), search performance is slower if one of the distracting stimuli is presented in a different color (Theeuwes, 1992), and slowing is greater for larger target/distractor color discrepancies (Duncan & Humphreys, 1989; Theeuwes, 1992; Wolfe & Horowitz, 2017). Influential cognitive models posit that the distractor results in greater activation in a priority map, which slows search for the shape-defined target stimulus. Indeed, reconstructed spatial representations of target and distractor stimuli measured from extrastriate visual and parietal cortex during an adapted version of this task show enhanced neural representations for color-defined distracting stimuli (Adam & Serences, 2021). While cognitive priority map models have offered parsimonious explanations of changes in discrimination performance, RTs, and gaze trajectories, they often disagree on when and how salient items capture attention (Luck et al., 2021).

One central issue limiting the ability to adjudicate among competing models is that there is no well-established method for quantitatively measuring how neural activity indexes the relative salience of different aspects of stimulus displays (Chang et al., 2021; Gaspelin & Luck, 2021; Pearson et al., 2021). Our findings may offer a practical solution to this challenge. Here, we demonstrate that neural representations of feature-based salience can simultaneously be tracked across multiple feature-selective regions, and previous work has identified similar salience representations using stimuli varying in luminance contrast across multiple retinotopic regions (Sprague et al., 2018). Together, this work establishes a framework for empirically estimating neural representations of visual salience across cortical

processing stages, including feature dimension maps. Using these methods, future work can investigate how competing stimuli made salient by different feature dimensions are represented within retinotopic maps and how the relative strength of those representations – and how they interact with one another within and between maps – may explain aspects of how salient stimuli capture attention measured using behavioral methods.

An example of a cognitive model that can be informed by these neural observations is the 'attentional window' account of attentional capture (Theeuwes, 2010, 2023a, 2023b). This model hypothesizes that difficult tasks requiring attention directed to a narrow region of the screen result in less attentional capture by salient distracting stimuli appearing outside the attended 'window', because the distracting stimuli are not processed to a sufficient degree to guide attention (as compared to when attention is directed to a larger area of the screen, which allows for the distracting stimuli to capture attention). Our study offers an important new constraint on this model: even when attention was narrowly directed to a challenging task at fixation, and when no behavioral impacts of the salient peripheral stimuli could be observed inside the scanner, we were able to identify representations of salient locations in feature dimension maps (Figs. 4 & 6). Thus, if variations in the size of an attentional window account for differences in attentional capture between task designs, we predict this window is likely to operate at a later stage, after feature-specific salience is computed in neural feature dimension maps.

Our results show that feature-selective retinotopic ROIs compute a stronger representation of the salient location defined based on their preferred feature dimension as compared to a non-preferred feature dimension. However, each ROI still represents the salient location, even

when made salient by the non-preferred feature value (Fig. 4). We speculate that this is due to feedback from higher-order regions (e.g., parietal or frontal cortex) that aggregate salience maps across individual feature dimensions to guide attention to important locations in the scene. Because the observers' task inside the scanner required careful fixation and the stimulus was always irrelevant, such automatic extraction of salient locations was never used by the participant to guide covert or overt attention (though overt attention was guided to these locations when participants were allowed to free-view, Fig. 3). However, it may be the case that the automatic identification of salient scene locations results in feedback signals across retinotopic cortex, similar to widespread retinotopic effects of cued spatial attention observed previously (e.g., Gandhi et al., 1999; Itthipuripat et al., 2019; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013; Tootell et al., 1998). Indeed, reconstructions based on parietal cortex activation patterns show representations of the salient location, as do those from all other retinotopic regions studied (Fig. 5C-D). Importantly, only feature-selective regions TO1/TO2 and hV4/VO1/VO2 show a systematic change in the representation of the salient location as a function of the salience-defining feature, supporting their role as neural feature dimension maps despite their weaker representation of a salient location based on a non-preferred feature (Fig. 7).

While this study provides evidence that specialized computations support the identification of salient locations based on different feature values, there remain some important limitations to this work. First, to maximize our ability to detect representations of salient locations in this study, we used stimuli of fixed size but random location defined by 100% feature contrast (color: opposite hues in HSV colorspace; motion: opposite motion directions). Future

research which parametrically manipulates the size, number, and feature contrast of salient stimulus locations in similar stimulus displays (e.g., Bogler et al, 2013; Zhang et al, 2012; Burrows & Moore, 2009) could enable both comparison of reconstructed spatial maps across various levels of stimulus salience and in-depth forward modeling of salience-related computations and their associated nonlinearities based on local feature contrast input to each voxel's receptive field (e.g., Hughes et al., 2019; Kay, Winawer, Rokem, et al., 2013; Yildirim et al., 2018). Second, future studies are required to test whether activation profiles in these neural feature dimension maps are equivalently sensitive to screen regions made salient by increases and decreases in feature intensity (e.g., motion speed, color saturation), which is a manipulation that has previously been effectively used to dissociate location-specific activation driven by stimulus intensity and local salience computations (Betz et al, 2013).

In summary, we found that feature-selective retinotopic ROIs compute maps of stimulus salience primarily based on feature contrast within their preferred feature dimension, confirming a key untested prediction of priority map theory. These results identify feature-selective retinotopic regions as the neural correlates of feature dimension maps within the priority map framework and support a new approach for probing the neural computations supporting visual cognition.

Chapter III: Neural Prioritization of Feature-Relevant Stimuli

Introduction

The environment is full of objects competing for our attention. While gardening, your goal may be to find green weeds in a strawberry plant, so attention will be directed to green objects. However, task-irrelevant objects that are salient may capture your attention, such as a hummingbird darting in the sky or a red strawberry. Furthermore, upon seeing the strawberry, you may be more likely to look for other strawberries. The numerous factors which contribute to how attention is allocated in the visual field—task-relevant goals, task-irrelevant salience, and stimulus history—are described by priority map theory (Awh et al., 2012; Itti & Koch, 2001; Luck et al., 2020; Serences & Yantis, 2006; Treisman & Gelade, 1980; Wolfe, 1994), where these priority signals are integrated to index the most important locations in a scene. Locations with the highest level of priority are then selected by attention (Carrasco, 2011; Eckstein, 2011).

The priority map itself indexes important information throughout the scene, but to generate this feature-agnostic map, it is necessary to evaluate importance related to specific feature dimensions (e.g., color or motion Wolfe & Horowitz, 2004, 2017). To this end, priority map theory states that there are several 'feature dimension maps' that prioritize information based on goals, salience, and stimulus history within particular feature dimensions (e.g., color or motion map). For example, when a salient moving hummingbird is present, the motion dimension map would have a high level of activation corresponding to the location of the bird (Klink et al., 2023; Liesefeld et al., 2017; Thayer & Sprague, 2023; Theeuwes, 1992). Similarly, goal-relevant locations are prioritized within the appropriate feature dimension

map—the color dimension map would prioritize locations containing green objects when searching for green weeds (Bahle et al., 2019; Corbetta & Shulman, 2002; Folk et al., 1992; Leber & Egeth, 2006; Poltoratski et al., 2017). Activation profiles across all feature dimension maps are integrated to create the feature-agnostic priority map, which ultimately directs attention.

Certain priority map models emphasize the role of goal relevance in modulating priority (Wolfe, 1994, 2021), because they are particularly viable for flexible deployment of attention (Jiang, 2018). Consistent with cognitive priority map models, goals modulate neural responses based on currently relevant spatial and feature information (Kastner et al., 1999; Maunsell & Treue, 2006; Schall & Hanes, 1993) and can guide attention to target items when they match active sensory representations (Desimone & Duncan, 1995; van Loon et al., 2018; Witkowski & Geng, 2023). When distracting information is known in advance, goals can be used to ignore specific stimuli (Arita et al., 2012; Carlisle, 2023; Cosman et al., 2018; Folk et al., 1992; Zhang et al., 2022). This shows that goals are pivotal for prioritizing attention.

Several studies leveraged goal-based attentional prioritization to identify the neural correlates of the feature-agnostic priority map (Bisley & Goldberg, 2010; Moore & Armstrong, 2003; Serences & Yantis, 2006). Superior colliculus (SC), posterior parietal cortex (PPC), and frontal eye fields (FEF) all show neural responses biased towards goal-relevant locations during covert and overt attention tasks (Bisley & Goldberg, 2006; Bruce & Goldberg, 1985; Moore & Armstrong, 2003), and these modulations occur across many different feature dimensions such as color (Burrows & Moore, 2009), shape (Klink et al., 2023), and luminance (Sprague, Itthipuripat, et al., 2018). Since these areas track the location of

important stimuli throughout the visual field across various feature dimensions, they could all be considered priority map correlates.

While there are many candidate neural priority maps, there is minimal evidence for neural feature dimension maps. Throughout visual cortex, certain areas have been identified as feature selective. Various studies have shown that hV4, VO1, and VO2 are preferentially responsive to color information through single-unit recording (Conway et al., 2007) and fMRI (Bartels & Zeki, 2000; Brewer et al., 2005). Similarly, regions TO1 and TO2 have an enhanced response when motion stimuli are present in the visual field (Amano et al., 2009; Huk et al., 2002). These regions are specialized in processing specific feature properties; however, to isolate them as feature dimension maps, activation within these regions need to be preferentially modulated by stimulus salience and task goals at the specific location(s) of presented stimuli. Work from our lab provides critical evidence identifying these feature-selective regions as dimension maps, as they track the location of feature-selective salient stimuli throughout the visual field (Thayer & Sprague, 2023). Responses from these regions were strongest when salience was defined by each region's preferred feature. However, in this study, the intent was to isolate the influence of salience. As such, it is unclear how top-down goals modulate activation profiles within these candidate neural dimension maps.

Some reports have found feature-selective attention modulates activity in color and motion regions with fMRI (McMains et al., 2007; Runeson et al., 2013). McMains and colleagues (2007) presented a colorful moving dot stimulus in the upper right visual field. At the start of each trial, they instructed participants to attend one feature dimension (e.g., color or motion) of the stimulus and respond whenever a target feature was shown (e.g., report when red

appears). Using univariate BOLD responses, they found that color- and motion-selective areas had a greater response when attending to the preferred feature dimension of each region.

While this provides evidence that feature-selective areas are modulated by goals, they fall short in demonstrating the spatial profiles of feature-selective responses, which is a necessary aspect of a feature dimension map (Fig. 8). Specifically, the location of feature-specific relevant information needs to be prioritized over other visual field locations and the univariate methods implemented in these studies do not have the spatial sensitivity of multivariate approaches (Kriegeskorte & Bandettini, 2007). This makes it ambiguous as to whether modulations occurred across the entire visual field or were localized to the location of the presented stimuli. Often, whenever a specific feature dimension is goal-relevant, neural modulations occur globally throughout the visual field (Hayden & Gallant, 2005; Saenz et al., 2002; Serences & Boynton, 2007; Treue & Trujillo, 1999b), which makes it possible that the actual stimulus location had activation similar to other locations. Thus, identifying the spatial extent of neural modulations in response to relevancy signals is key for considering feature-selective extrastriate regions as dimension maps.

We used a multivariate approach to scrutinize whether feature-selective cortical areas are representative of feature dimension maps. Specifically, we wanted to know whether top-down attention modulated activity at the *location of the stimulus* during a feature-selective attention task in color-selective regions hV4/VO1/VO2 and motion-selective regions TO1/TO2, and if responses were stronger when cued to attend the preferred feature of each region. If so, this would be evidence that feature-selective regions in visual cortex act as

feature dimension maps within priority map theory. To circumvent the limited spatial resolution inherent with voxelwise univariate averaging (Kriegeskorte & Bandettini, 2007; Serences & Saproo, 2012), we utilized a multivariate inverted encoding model (IEM) which can reconstruct spatial maps from neural response patterns (Sprague et al., 2014; Sprague, Itthipuripat, et al., 2018; Sprague et al., 2019; Sprague & Serences, 2013; Thayer & Sprague, 2023). To preview our results, we found that the stimulus location in both regions had the greatest activation, and it was modulated by task demands, such that it was strongest when attending to each region's preferred feature. These findings were only observable for both regions after implementing IEM, demonstrating the usefulness of this approach in investigating neural feature dimension maps.

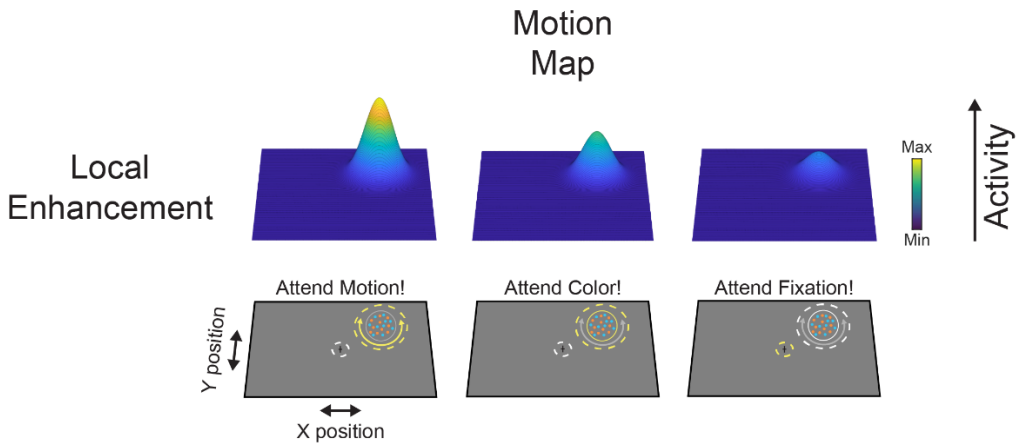

Figure 8: Feature dimension maps prioritize locations based on goal-relevant feature dimension. Priority map theory states that there are various "feature dimension maps" that index the most important locations in the visual field. One factor that determines whether a location is important are the current goals of an individual. If motion is needed for an ongoing task (e.g., finding a darting hummingbird), then activation within the corresponding "motion map" increases the importance associated with the hummingbird's location. Local enhancement could occur, such that only the stimulus location is prioritized. The presence of any salient motion in the visual field, defined by local feature contrast, would lead to increased priority within the motion map across all condition types. Activation within the motion map would not be strongly modulated when color is goal relevant, but if motion is still present, there still may be modest prioritization at that location due to image salience (Thayer & Sprague, 2023). A motion dimension map is depicted here, but modulations would similarly apply for other feature dimensions, such as color.

Materials & Methods

Participants.

We recruited 10 subjects from the University of California, Santa Barbara (UCSB) community to participate in this study (9 female, 18-29 years old). We identified an appropriate sample size for our main effect of interest by conducting a power analysis using pilot data (n = 3). In the pilot study, we had a strong effect ($d_z = 1.24$) which required a sample of 6 subjects to obtain 80% power with an alpha criterion of $\alpha = .05$ as indicated by a power analysis conducted in G*Power (Faul et al., 2007). We collected a large number of measurements from each subject to minimize within-subject variance, which can benefit statistical power more than increased sample sizes (Baker et al., 2021). Subjects reported normal or corrected-to-normal vision and did not report neurological conditions. All procedures were approved by the UCSB Institutional Review Board (IRB# 5-24-0030), and the study was registered on ClinicalTrials.gov (NCT06281457). Subjects gave written informed consent before participating and were compensated for their time ($20/h for scanning sessions, $10-20/h for behavioral familiarization/training).

Stimuli and procedure.

Participants performed a 30-minute training session before scanning so that they were familiarized with the instructions. We used this session to establish the initial behavioral performance thresholds used in the first run of the scanning session. In the main task session, participants were scanned for a single two-hour period consisting of at least 4 runs of a spatial mapping task, which were used to independently estimate encoding models for each voxel, and 8 runs of the experimental feature-selective attention task. All participants also

underwent additional anatomical and retinotopic mapping scanning sessions (1-2x 1.5-2 hr sessions) to identify regions of interest (ROI; see Region of interest definition).

Stimuli were presented using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) for MATLAB (The MathWorks, Natick, MA). Visual stimuli were rear-projected into a screen placed ~110 cm from the participant's eyes at the head of the scanner bore using a contrast-linearized LCD projector (1,920×1,080, 60 Hz) during the scanning session. For two of the participants, a CRS BOLD screen LCD monitor was used to present stimuli (1,920x1,080, 60 Hz) that was placed ~140 cm from the participant's eyes. All results are qualitatively identical for the subjects who used either the BOLD monitor or the projector. In the behavioral familiarization session, we presented stimuli on a contrast-linearized LCD monitor (2,560×1,440, 60 Hz) 62 cm from participants, who were seated in a dimmed room and positioned using a chin rest. For all sessions and tasks (main tasks and mapping task), we presented stimuli on a neutral gray circular aperture (9.15° radius), surrounded by black (only aperture shown in Fig. 9A).

Feature-selective attention task.

For the main task (Fig. 9A), participants were shown a colorful moving dot stimulus in the periphery along with a flickering central fixation cross (0.1° size) on each trial. The colorful moving dot stimulus was a 1.5°-radius circular disc comprised of several individual dots centered 5° from fixation at a random location along an invisible ring from 0°-359°. Individual dots were colored either cyan (RGB: [0; 200; 200]) or orange (RGB: [255; 165; 0]), and moved clockwise, counterclockwise, or in a random direction. Individual dots occupied 0.04° of visual angle, with a dot density was 40 dots/deg^2, and moved at a speed of

4.5°/s. Dots were randomly replotted every 50 ms or when they exceeded the stimulus bounds. Both cyan and orange dots were always present. Dots were always moving, but only some dots on each trial would move clockwise or counterclockwise, while the rest of the dots moved in a random direction (uniform distribution of planar motion directions).

At the start of each trial, participants were shown a letter cue at fixation (0.4° height) that would indicate what aspect of the display to attend. The letter cue was presented for 1000 ms in white (RGB: [180; 180; 180]) and Arial font. The cue could be an 'F' (attend fixation), 'C' (attend color), or 'M' (attend motion). After a 500 ms blank period, with only the aperture present, the colorful moving dot stimulus and fixation cross was shown for 3000 ms. When cued to attend the color of the stimulus ('C'), participants reported which individual feature value (cyan or orange) was most prevalent in the stimulus. When cued to attend the motion of the stimulus ('M'), they reported the coherent motion direction of the stimulus (clockwise or counterclockwise). Since motion and color was present on each trial, visual input was constant, and we were able to isolate changes in the neural stimulus representations across brain regions due to instructed task goals. We adjusted the difficulty of the attend color/motion tasks between runs by changing the coherence of the dots to try and equate task performance between conditions (e.g., 80% dot coherence indicates that 80% of the dots were presented in cyan or were moving clockwise). The first run of the task in the scanner used coherence values acquired from the behavioral training session.

As a neutral baseline condition, we had participants attend the central fixation cross on some trials. The vertical and horizontal lines of the fixation cross were 0.1° of visual angle long and flickered at 3 Hz (10 frames on, 10 frames off at 60 Hz). On each trial, one line of the

fixation cross would slightly change in size. When cued to attend fixation ('F'), participants reported which line increased in length (horizontal or vertical). Since the dot stimulus was ignored on these trials due to fixation task difficulty (Thayer & Sprague, 2023), the neural response associated with the stimulus on these trials was purely driven by visual input. We adjusted the difficulty of the fixation task between runs by altering the degree of size change for vertical/horizontal lines based on behavioral accuracy (range: 0.05° to 0.2°).

Each run had 24 trials. There were 8 trials of each condition (attend fixation, attend stimulus color, attend stimulus motion). All trials were separated by a randomly selected ITI ranging from 6-9 s with an average ITI of 7.5 s. Trial order was shuffled within run. Each run started with a 3 s blank period and ended with a 10.5 s blank period, for a total run duration of 301.5 s. Eye position was monitored throughout the experiment using an Eyelink 1000 eyetracker (SR Research). The right eye gaze position was monitored with the eyetracker at a 500 Hz sampling rate. Participants performed a 5-point calibration at the start of the experiment.

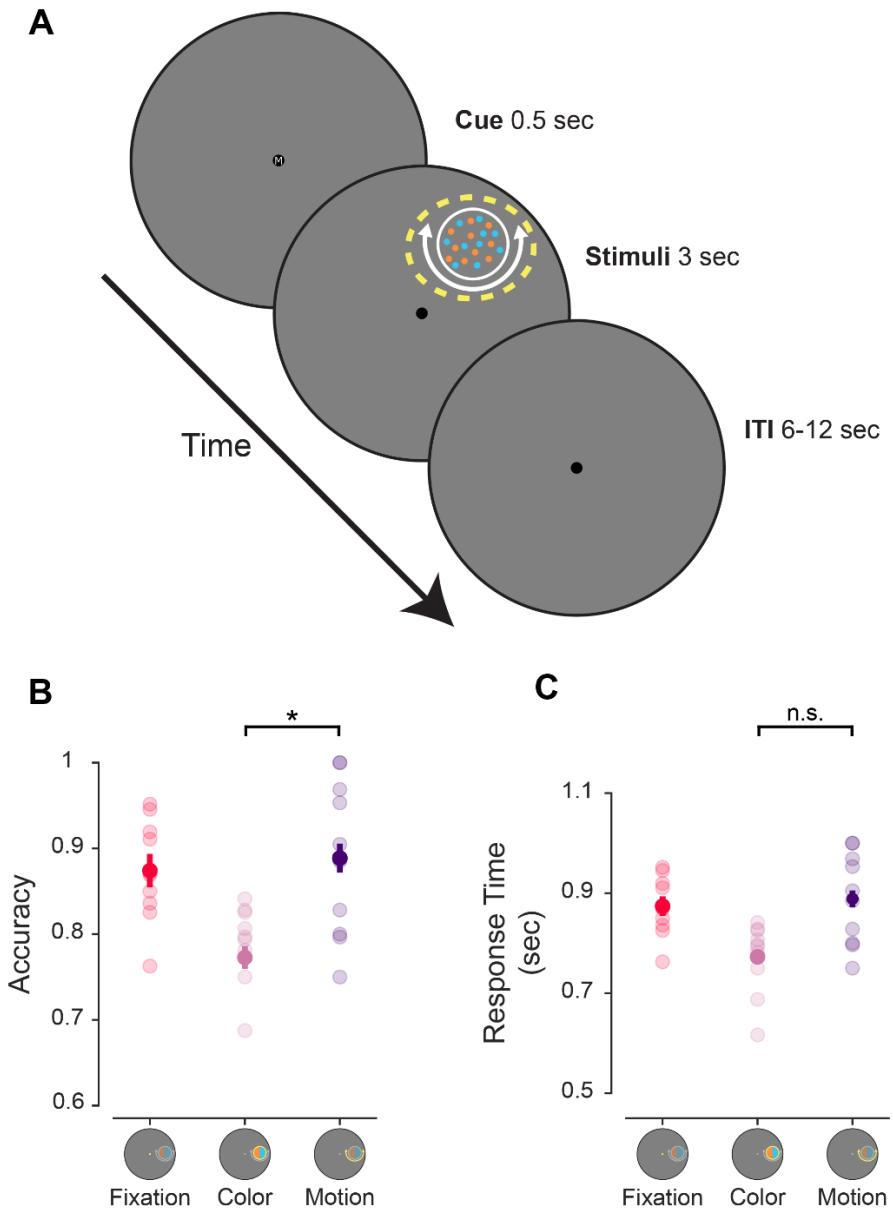

Figure 9: Feature-selective attention task. A. During the main experimental session, participants performed one of three possible attention tasks. On each trial, a dot array comprised of colorful moving dots was shown. Dots were randomly either cyan or orange. Dots were always moving, but only some would move clockwise or counterclockwise, while the rest moved in random direction. In addition, a flickering fixation cross was always present, and the vertical or horizontal arm of the cross would increase in size once during the stimulus presentation period. At the start of each trial, participants were cued to perform one of three attention tasks related to the upcoming stimulus display. They could be cued to attend either the color or motion of the dot array. If this was the case, they had to report which individual feature value was most prevalent in the cued feature dimension (e.g., cued to attend color, are there more cyan or orange dots present). As a baseline, participants were occasionally cued to attend the fixation cross and report which arm of the cross increased in size. **B.** Accuracy for each attention condition. Performance was slightly worse for the attend color than the attend motion or fixation conditions. **C.** Behavioral response times for each attention condition. While performance was faster on the attend fixation trials than either the attend color or motion

conditions, there was no difference between the attend color and attend motion conditions. Error bars indicate SEM across participants. * Significant difference based on permuted paired-samples *t* test ($p < 0.05$).

Spatial mapping task

We also acquired several runs of a spatial mapping task used to independently estimate a spatial encoding model for each voxel, following previous studies (Sprague et al., 2016; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013; Thayer & Sprague, 2023). On each trial of the mapping task, we presented a flickering checkerboard at different positions selected from a hexagonal grid spanning the screen. Participants viewed these stimuli and responded whenever a rare contrast change occurred (6 out 43 trials, 13.9%), evenly split between contrast increments and decrements. The checkerboard stimulus was the same size as the stimulus in the feature-selective attention task (1.5° radius) and was presented at 70% contrast and 6-Hz full-field flicker. All stimuli appeared within a gray circular aperture with a 9.15° radius, as in the feature-selective attention task. For each trial, the location of the stimulus was selected from a triangular grid of 37 possible locations with an added random uniform circular jitter (0.5° radius). The base position of the triangular grid was rotated by 30° on every other scanner run to increase spatial sampling density. As a result, every mapping trial was unique, which enabled robust spatial encoding model estimation.

Each trial started with a 3000 ms stimulus presentation period. If a target was present, then the stimulus would be dimmed/brightened for 500 ms with the stipulation that the contrast change would not occur in either the first or last 500 ms of the stimulus presentation period. Finally, it is were randomly selected to be either 6 or 8.25 s. All target-present trials were discarded when estimating the spatial encoding model. Each run consisted of 43 trials (6 of

which included targets). We also included a 3 s blank period at the beginning of the run and a 10.5-s blank period at the end of the run. Each run totaled 432 s.

Retinotopic mapping task. We used a previously reported task (Mackey et al., 2017; D. D. Thayer & Sprague, 2023) to identify retinotopic regions of interest (ROIs) via the voxel receptive field (vRF) method (Dumoulin & Wandell, 2008). Each run of the retinotopy task required participants attend several random dot kinematograms (RDK) within bars that would sweep across the visual field in 2.25 s (or, for one participant, 2.6 s) steps. Three equally sized bars were presented on each step and the participants had to determine which of the two peripheral bars the motion in the central bar matched with a button press. Participants received feedback via a red or green color change at fixation. We used a three-down/one-up staircase to maintain ~80% accuracy throughout each run so that participants would continue to attend the RDK bars. RDK bars swept 17.5° of the visual field. Bar width and sweep direction was pseudo-randomly from several different widths (ranging from 2.0° to 7.5°) and four directions (left-to-right, right-to-left, bottom-to-top, and top-to-bottom).

fMRI acquisition and MRI preprocessing.

All scanning was conducted at the UCSB Brain Imaging Center with a 3T Siemens Prisma scanner. fMRI data acquisition and preprocessing pipelines in the current study exactly follow our previous report (D. D. Thayer & Sprague, 2023). Briefly, we collected 3 T1-weighted and 1 T2-weighted which were used to coregister all functional images to each participant's native anatomic space. We used freesurfer's recon-all script (version 6.0) to process all anatomical scans and AFNI's afni_proc.py command to preprocess all functional scans.

Region of interest definition.

We identified 15 ROIs using independent retinotopic mapping data. We fit a vRF model for each voxel in the cortical surface (in volume space) using averaged and spatially smoothed (on the cortical surface; 5 mm FWHM) time series data across all retinotopy runs (8-12 per participant). We used a compressive spatial summation isotropic Gaussian model (Kay, Winawer, Mezer, et al., 2013; Mackey et al., 2017) as implemented in a customized, GPU-optimized version of mrVista (see Mackey et al., 2017) for detailed description of the model). High-resolution stimulus masks were created (270 x 270 pixels) to ensure similar predicted responses within each bar size across all visual field positions. Model fitting began with an initial high-density grid search, followed by subsequent nonlinear optimization. Retinotopic ROIs (V1, V2, V3, V3AB, hV4, LO1, LO2, VO1, VO2, TO1, TO2, IPS0-3) were then delineated by projecting vRF best-fit polar angle and eccentricity parameters with variance explained ≥10% onto each participant's inflated cortical surfaces via AFNI and SUMA (Fig. 2). We drew ROIs on each hemisphere's cortical surface based on previously-established polar angle reversal and foveal representation criteria (Amano et al., 2009; Mackey et al., 2017; Swisher et al., 2007; Wandell et al., 2007; Winawer & Witthoft, 2015). Finally, ROIs were projected back into volume space to select voxels for analysis.

Inverted encoding model.

We used a spatial inverted encoding model (IEM) to reconstruct images of stimulus-related activation patterns measured across entire ROIs (Sprague & Serences, 2013; Fig. 10). To do this, we first estimated an encoding model, which describes the sensitivity profile over the relevant feature dimension for each voxel in a region. This requires using data set aside for

this purpose, referred to as the "training set". Here, we used data from the spatial mapping task as the independent training set. The encoding model across all voxels within a given region is then inverted to estimate a mapping used to transform novel activation patterns from a "test set" (runs from the feature-selective attention task) and reconstruct the spatial representation of the stimulus at each timepoint.

We built an encoding model for spatial position based on a linear combination of 37 spatial filters (Sprague et al., 2014; Sprague, Itthipuripat, et al., 2018; Sprague & Serences, 2013; D. D. Thayer & Sprague, 2023). Each voxel's response was modeled as a weighted sum of each identically shaped spatial filter arrayed in a triangular gird (Fig. 10). The centers of each filter were spaced by 2.83° and were Cosine functions raised to the 7th power:

$$f(r) = \left(0.5 + 0.5\cos\frac{\pi r}{s}\right)^7 \quad (1)$$

for $r < s$; 0 otherwise

where r is the distance from the filter center and s is a size constant. The size constant reflects the distance from the center of each spatial filter at which the filter returns to 0. This triangular grid of filters forms the basis set, or information channels for our analysis. For each stimulus used in our mapping task, we converted from a contrast mask to a set of filter activation levels by taking the dot product of the vectorized stimulus mask (n trials × p pixels) and the sensitivity profile of each filter (p pixels × k channels). We then normalized the estimated filter activation levels such that the maximum activation was 1 and used this output as $\boldsymbol{C_1}$ in the following equation, which acts as the forward model of our measured fMRI signals:

$$B_1 = C_1 W \quad (2)$$

$\boldsymbol{B_1}$ (n trials × m voxels) in this equation is the measured fMRI activity of each voxel during the visuospatial mapping task and $\boldsymbol{W}$ is a weight matrix (k channels × m voxels) which quantifies the contribution of each information channel to each voxel. $\widehat{\boldsymbol{W}}$ can be estimated using ordinary least-squares linear regression to find the weights that minimize the differences between predicted values of $\boldsymbol{B}$ and the observed $\boldsymbol{B_1}$:

$$\widehat{W} = (C_1^T C_1)^{-1} C_1^T B_1 \quad (3)$$

This is computed for each voxel within a region independently, making this step univariate. The resulting $\widehat{W}$ represents all estimated voxel sensitivity profiles. We then used $\widehat{W}$ and the measured fMRI activity of each voxel (i.e., BOLD response) during each trial (using each TR from each trial, in turn) of the feature salience task using the following equation:

$$\hat{C}_2 = B_2 \widehat{W}^T (\widehat{W} \widehat{W}^T)^{-1} \quad (4)$$

Here, $\hat{C}_2$ represents the estimated activation of each information channel (n trials × k channels) which gave rise to that observed activation pattern across all voxels within a given ROI ($\boldsymbol{B_2}$; n trials × m voxels). To aid with visualization, quantification, and coregistration of trials across stimulus positions, we computed spatial reconstructions using the output of Equation 4. To do this, we weighted each filter's spatial profile by the corresponding channel's reconstructed activation level and then sum all weighted filters together (Fig. 3).

Since stimuli in the feature-selective attention task were randomly positioned on every trial, we rotated the center position of spatial filters such that the resulting 2D reconstructions of the stimuli were aligned across trials and participants (Fig. 10). We then sorted trials based on condition (attend fixation, attend color, attend motion). Finally, we averaged the 2D

reconstructions across trials within the same condition for individual participants, then across all participants for our grand-average spatial reconstructions (Fig. 12A; Fig. 13B). Individual values within the 2D reconstructed spatial maps correspond to visual field coordinates.

Critically, because we reconstructed all trials from all conditions of the feature-selective attention task using an identical spatial encoding model estimated with an independent spatial mapping task, we can compare reconstructions across conditions on the same footing (Sprague, Adam, et al., 2018; Sprague et al., 2019; Sprague, Itthipuripat, et al., 2018).

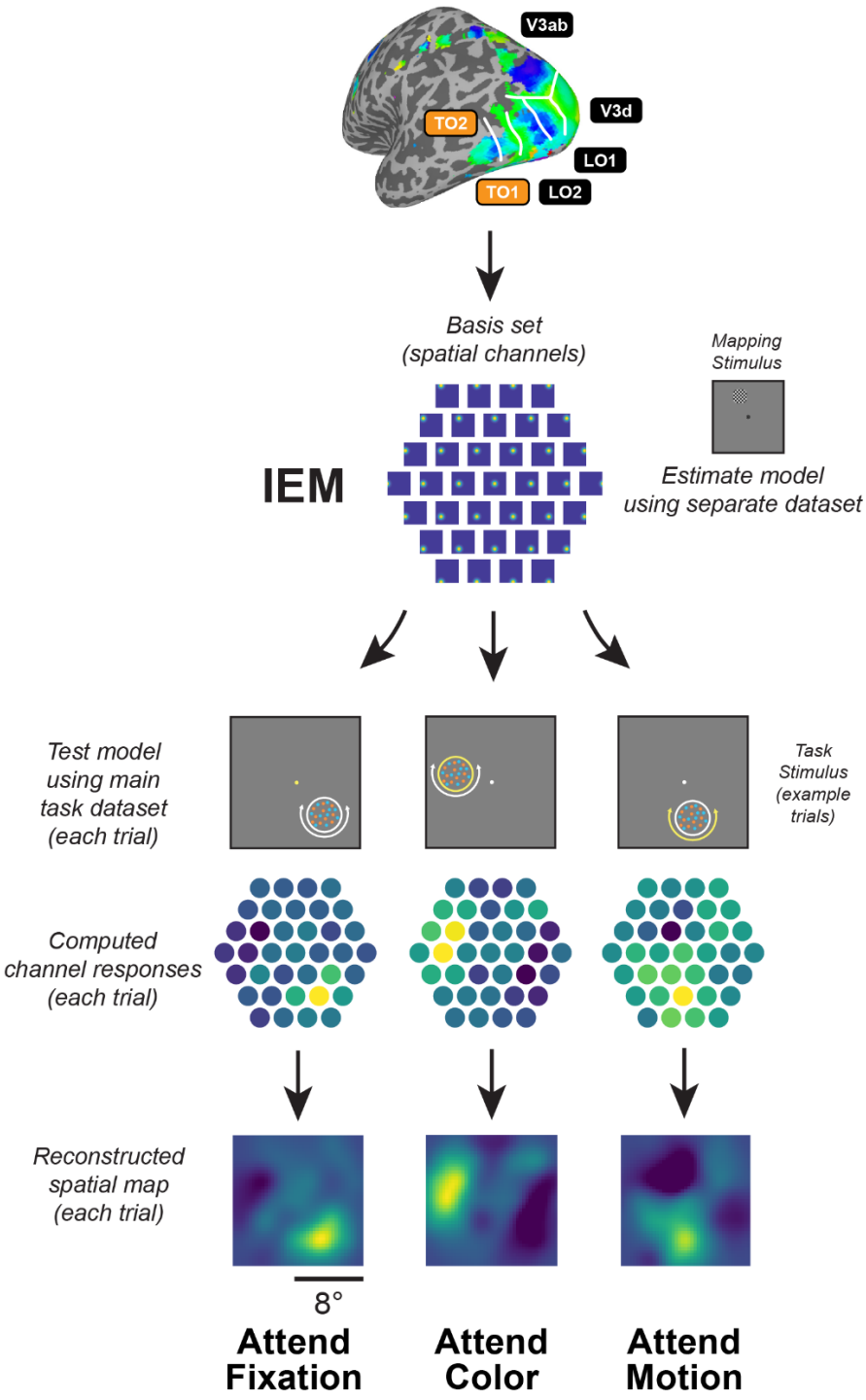

Figure 10: Inverted encoding model used to reconstruct maps of the visual field. Data from an independent spatial mapping task was used to estimate a spatial IEM for each ROI (for details, see Materials and Methods). Through this approach, activation patterns across all voxels in a region can be mapped onto the activation of spatial channels that are then summed to produce reconstructed spatial maps of the visual field. We reconstructed maps for each timepoint on each trial and directly compared the map activation across conditions. Since maps are in visual field coordinates, we were able to rotate and align reconstructed

spatial maps to the known position of the moving dot array. For each trial, we averaged maps from 4-7 s after stimulus onset and then averaged across trials for each condition.

Quantifying neural responses.

For all univariate analyses, we averaged the response across all voxels that had voxel receptive field (vRF) centers (as identified through the retinotopic mapping task) overlapping with the dot stimulus location on a given trial. Voxels overlapped with the stimulus if the vRF center was within the stimulus location, which was a 1.5° radius disk centered at the known position of each stimulus. We compared the response of stimulus-aligned voxels with voxels that had vRF centers overlapping with the exact opposite location of the stimulus within a 1.5° radius disk.

For the multivariate analyses, we quantified the strength of stimulus representations within each reconstructed spatial map. To do this, we computed the mean map activation of pixels located within a 1.5° radius disk centered at the known position of each stimulus (matching the stimulus radius of 1.5°; see Sprague, Itthipuripat, et al., 2018). This provides a single value corresponding to the activation of the stimulus location for a given condition, within each retinotopic ROI. To assess the spatial selectivity of reconstructed spatial maps, we compared the mean map activation at the location of stimulus to map activation at the location opposite fixation using a 1.5° radius disk (Fig. 12).

Lastly, we evaluated how task repetition modulated neural responses (Fig. 13). We did this by computing the neural response corresponding to the stimulus location as a function of whether participants were cued to attend to the same or different feature dimension as the previous trial (repeat and switch trials, respectively). For this analysis, we only considered

trials where the stimulus location was attended (trials preceded by attend-fixation trials were not included).

For both univariate and multivariate data, we computed an attention modulation index (AMI) at the stimulus location (Fig. 11E, 12D). This was done to assess feature preferences of each ROI while minimizing the influence of spatial attention. AMI was computed as the difference in activation at the stimulus location between the attend motion and attend color conditions, where positive values indicate a stronger response to the stimulus when attending motion.

Statistical analysis.

Parametric statistical tests were conducted for all comparisons (repeated-measures ANOVAs and *t* tests). To account for possible non-normalities in our data, we generated null distributions for each test using a permutation procedure (see below) to derive p-values.

First, we used a one-way repeated-measures ANOVA (factor: stimulus condition; 3 levels: attend fixation, attend color, attend motion) to determine whether behavioral performance (accuracy and RT) on the fixation task depended on what was attended on each trial (Fig. 2B/C). We then conducted follow-up paired-samples *t* tests between attention conditions to see which conditions differed.

When analyzing fMRI data, we focused on retinotopic regions that have previously been shown to be color- or motion-selective (Brewer et al., 2005; Huk et al., 2002; Thayer & Sprague, 2023). For both univariate and multivariate neural responses, we first computed a three-way ANOVA with location (2 levels: stimulus and opposite location), attention condition (3 levels: attend fixation, color, and motion), and ROI (2 levels: color-selective

hV4/VO1/VO2 and motion-selective TO1/TO2) as factors (Fig. 11/12). We then computed separate two-way ANOVAs for each relevant location (stimulus and opposite location), with attention condition (3 levels: attend fixation, color, motion) and ROI (2 levels: color-selective hV4/VO1/VO2 and motion-selective TO1/TO2) as factors. Next, we computed the difference between stimulus and opposite locations and conducted a two-way ANOVA to determine the spatial selectivity of neural modulations with attention condition (3 levels: attend fixation, color, motion) and ROI (2 levels: color-selective hV4/VO1/VO2 and motion-selective TO1/TO2) as factors. To directly test whether feature-selective ROIs represent stimulus locations more strongly when goals direct attention to their preferred feature dimension, we computed a paired-samples *t* test on the difference between map activation on attend color and attend motion trials between color-selective and motion-selective ROIs.

Lastly, we used the behavioral and multivariate neural data to assess how responses changed as a function of goal repetition (Fig. 13). To do this, we binned trials based on whether the color/motion attention task instructed on a given trial (N) was the same (repeat) or different (switch) from the previous trial (N-1). Similar to other reports (Maljkovic & Nakayama, 1994; Thayer et al., 2022), we excluded trials with inaccurate behavioral responses. We also excluded trials where the previous trial (N-1) was an attend fixation trial as repetition effects are diminished when attention is directed away from a stimulus (Larsson & Smith, 2012). We then conducted a two-way ANOVA on the behavioral data with trial repetition (2 levels: switch, repeat) and attention condition (2 levels: attend color, attend motion) as factors. We also conducted separate two-way ANOVA using multivariate neural responses for each feature-selective ROI (color-selective hV4/VO1/VO2 and motion-selective TO1/TO2) with

trial repetition (2 levels: switch, repeat) and attention condition (2 levels: attend color, attend motion) as factors.

For our shuffling procedure, we used a random number generator that was seeded with a single value for all analyses. The seed number was randomly selected using an online random number generator (https://numbergenerator.org/random-8-digit-number-generator). Within each participant, averaged data within each condition were shuffled across conditions for each participant individually, and once shuffled, the statistical test of interest was recomputed over 1000 iterations. *P* values were derived by computing the percentage of shuffled test statistics that were greater than or equal to the measured test statistic. We controlled for multiple comparisons using the false discovery rate (Benjamini & Yekutieli, 2001) across all comparisons within an analysis when necessary. Error bars are standard error, unless noted otherwise.

Results

Behavior.

Participants performed a feature-selective attention task while we used fMRI to measure BOLD activation patterns across several independently defined retinotopic regions. Participants were cued to covertly attend to either the color or motion of a colored-moving dot stimulus on every trial, to determine which individual feature value was most prevalent. Accuracy differed across attention conditions (Fig. 9B) as indicated by a permuted one-way repeated measures ANOVA ($p = 0.002$), where accuracy was worse on attend color than attend motion trials ($p < 0.001$). Response times (RT; Fig. 9C) also differed across conditions

($p < 0.001$; permuted one-way repeated measures ANOVA). However, follow-up comparisons indicated that this was driven by faster responses to the attend fixation condition relative to the attend color ($p < 0.001$) and attend motion ($p < 0.001$) conditions. Critically, there a was no significant difference between the attend color and motion conditions ($p = 0.270$). This indicates that there was a similar amount of difficulty in processing the stimulus between these two conditions. Since this is the case, it is unlikely that differences in neural responses are due to behavioral differences.

Univariate neural responses.

For our univariate neural results, we first conducted an analysis to replicate findings from a previous report investigating how goals modulate responses in feature-selective visual cortex (McMains et al., 2007). To do this, we computed the mean response of all voxels that had vRF centers within the known stimulus location as well as the opposite location on each trial. The resulting hemodynamic response functions (HRFs) are plotted in Fig. 11A/B. We averaged responses 4-7 seconds after the letter cue was presented (Fig. 11C-E). A 3-way permuted ANOVA with location (stimulus and opposite), attention condition (attend fixation, color, and motion), and ROI (hV4/VO1/VO2 and TO1/TO2) revealed a three-way interaction ($p < 0.001$). To better identify what was driving these effects, we conducted two-way permuted ANOVAs for each feature-selective ROI with location (stimulus and opposite) and attention condition (attend fixation, color, and motion). The point of this analysis was to determine whether motion- and color-selective regions were modulated by task demands (Fig. 111C). Consistent with this, in color-selective regions there was a main effect of attention condition ($p < 0.001$) and location ($p < 0.001$) as well as an interaction ($p = 0.048$).

Follow-up permuted paired sample *t* tests indicate that there was a stronger neural response from voxels with vRF centers at the stimulus location when attending to stimulus color than when attending to fixation ($p = 0.001$). There was also a stronger response when attending to stimulus motion than fixation ($p = 0.001$). This finding replicates previous work, indicating that color regions are modulated by goals.

Motion-selective regions similarly showed a main effect of attention condition ($p < 0.043$) and location ($p < 0.001$) as well as an interaction ($p = 0.003$). Unlike color regions, motion-selective areas showed no modulation of task demands at the stimulus location as there was similar activation between the attend fixation and each of the attend stimulus feature conditions (color: $p = 0.426$; motion: $p = 0.156$). The interaction effect was likely driven by the relatively strong response at the opposite location when attending to fixation. Since vRFs are much larger in motion-selective areas than in color-selective regions (Wandell & Winawer, 2015), attending to fixation could have increased neural response at the opposite location because many vRFs overlap with both positions. However, when attending to the stimulus, there would not be as much overlap with the opposite location because it is farther away.

Next, we assessed the spatial selectivity of task demand modulations by computing the difference in BOLD response between voxels with vRF centers at the stimulus and opposite locations, where positive values indicate a stronger response corresponding to the stimulus (Fig. 11D). It is crucial that a candidate neural feature dimension map is spatially selective, as this is necessary to index importance throughout the visual field. Previous reports are limited in that they were unable to differentiate between local or global modulations of

attention (McMains et al., 2007). A two-way permuted ANOVA with attention condition (attend fixation, color, and motion) and ROI (hV4/VO1/VO2 and TO1/TO2) indicated a main effect of condition ($p < 0.001$), main effect of ROI ($p = 0.001$), and an interaction between attention condition and ROI ($p < 0.001$). Follow-up comparisons reveal that color-selective regions had a stronger response to the stimulus since both the attend color and motion conditions had a greater response than the attend fixation condition (color: $p < 0.001$; motion: $p < 0.001$). Furthermore, the response was greater when attending to color versus motion ($p = 0.011$), indicating that there was a stronger response at the stimulus location when attending to the preferred feature of the color-selective regions.

Through this analysis, motion-selective regions now had a stronger response to the stimulus location when attending to either the color or motion of the stimulus relative to attending to fixation (color: $p = 0.007$; motion: $p = 0.004$). This is likely due to the subtraction minimizing the possible attention effect from large vRFs in TO1/TO2 during attend fixation trials (Wandell & Winawer, 2015). Even though responses in both attend stimulus conditions were now greater than attend fixation, there was still no difference between the attend color and motion conditions ($p = 0.457$), indicating that task demands modulated responses at the stimulus locations similarly between the attend color and motion conditions.

Critically, to determine if these regions were feature-selective, we computed an AMI for each feature-selective ROI (Fig. 11E). This was the difference between mean responses at the stimulus location between the attend motion and attend color conditions. While color-selective regions were negative ($p < 0.001$), indicating that they had a stronger response to the stimulus when attending to color, motion-selective regions were no different from zero (p

= 0.108). Thus, based on the univariate response, only color-selective regions could be considered a neural dimension map.

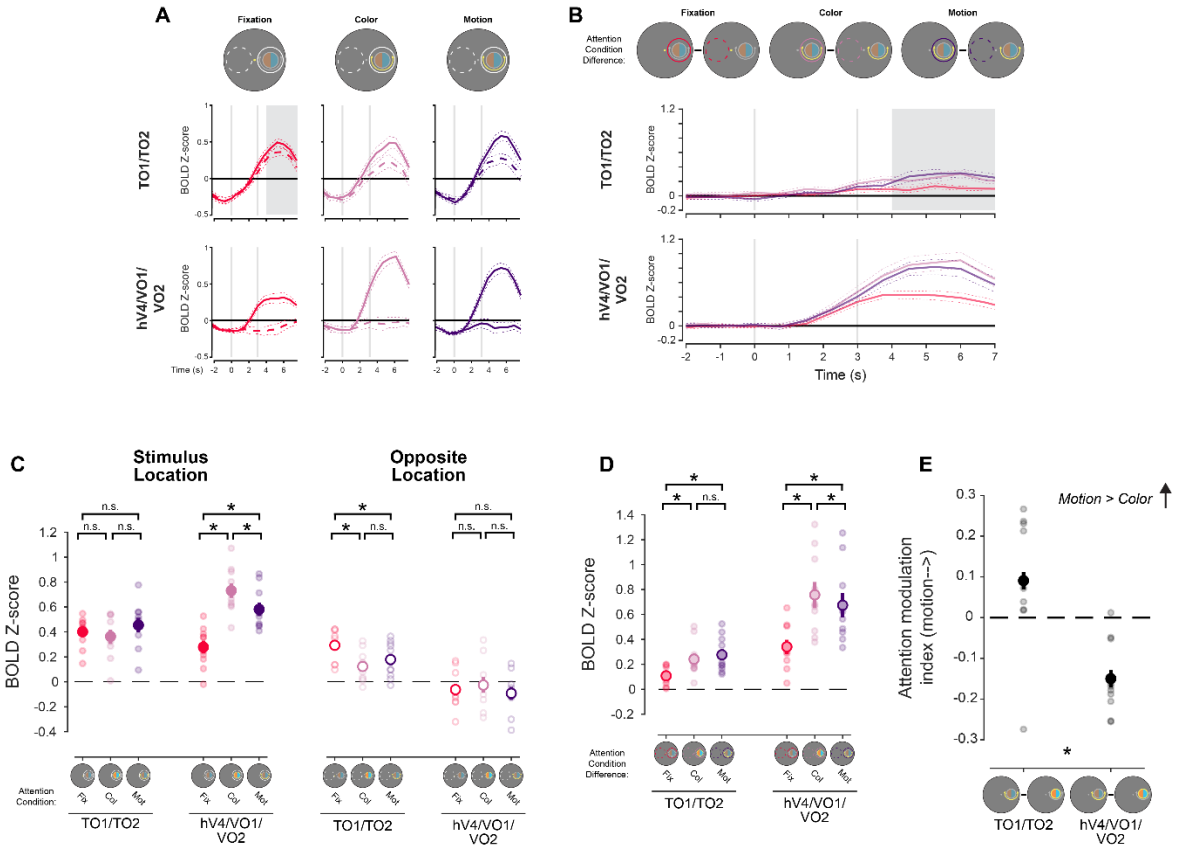

Figure 11: Voxel RF-sorted BOLD responses support neural color dimension map but not motion dimension map. A. HRFs from each retinotopic feature-selective ROI. Univariate BOLD responses were computed using voxels that had vRF centers within the stimulus (solid line) or opposite (dashed line) location. Shaded gray indicates the data that were averaged for plots C-E. **B.** The difference in mean BOLD response between the stimulus and opposite location for each feature-selective retinotopic ROI. Dashed lines indicate within-subject SEM. Qualitatively, color-selective regions had a very strong neural response at the stimulus location. **C.** BOLD responses averaged 4-7 s after instruction cue onset for feature-selective ROIs. Color-selective regions were modulated by task demands at the stimulus location, as BOLD responses were greater when attending to the stimulus than when attending to fixation (color vs fixation: $p < 0.001$; motion vs fixation: $p < 0.001$). However, motion-selective regions were not modulated by task demands, as BOLD responses were the same at the stimulus location across attention conditions (color vs fixation: $p = 0.426$; motion vs fixation: $p = 0.156$. **D.** Difference in BOLD responses for each feature-selective ROI between the stimulus and opposite locations. Color-selective regions again show a modulation of task demands, where activation was greater when attending to the stimulus than when attending to fixation (color vs fixation: $p < 0.001$; motion vs fixation: $p < 0.001$). Similarly, motion-selective regions were modulated by task demands (color vs fixation: $p = 0.007$; motion vs fixation: $p = 0.004$). **E.** Difference in BOLD response at the stimulus location for each ROI between attend motion and attend color conditions. Positive values indicate a stronger response when attending to stimulus motion. Color-selective regions had a negative value ($p < 0.001$), but motion-selective regions did not differ from 0 ($p = 0.108$). * Significant difference based on permuted paired samples *t* test ($p < 0.05$). Error bars indicate SEM within participants.

Multivariate spatial representations.

Based on the univariate findings only color regions reflect the required properties of a feature dimension map. To more finely evaluate whether TO1/TO2 and hV4/VO1/VO2 can be considered neural feature dimension maps, we computed an IEM for each region (see Materials and Methods; Sprague et al., 2018). To do this, we used an independent "mapping" task to estimate a spatial encoding model for each voxel. The model was parameterized as a series of weights across several smooth, overlapping spatial channels. The weights from the spatial encoding model across all voxels that comprised a feature-selective ROI were then inverted, which allowed us to reconstruct spatial maps based on activation profiles from the feature-selective attention task (Fig. 10). Reconstructions are generated for each time point. We averaged reconstructions across time points corresponding to 4-7 s after cue onset.

The reconstructed spatial maps (Fig. 12A) qualitatively have a stronger response at the aligned stimulus location than the rest of the modelled visual field. We quantified whether there was a stronger response to the stimulus by averaging map activation profiles within a 1.5° radius disk (stimulus size) at the aligned stimulus location. We could then compare stimulus-related activation to the opposite location, which participants were not attending to and there was no stimulus present (Fig. 12B). A 3-way permuted ANOVA with location (stimulus and opposite), attention condition (attend fixation, color, and motion), and ROI (hV4/VO1/VO2 and TO1/TO2) revealed a three-way interaction ($p < 0.001$). We conducted two-way permuted ANOVAs for each feature-selective ROI with location (stimulus and opposite) and attention condition (attend fixation, color, and motion) to better determine what

was driving the three-way interaction. Similar to the univariate approach, the aim of this analysis was to determine whether motion- and color-selective regions were modulated by task demands. Consistent with this, color-selective regions showed a main effect of attention condition ($p < 0.001$) and location ($p < 0.001$) as well as an interaction ($p < 0.001$). Follow-up permuted paired sample *t* tests indicate that there was a stronger neural response at the stimulus location when attending to stimulus color than when attending to fixation ($p < 0.001$). There was also a stronger response when attending to stimulus motion than fixation ($p < 0.001$). Motion-selective regions had no effect of attention condition ($p = 0.131$) but did have a main effect of location ($p < 0.001$) and an interaction between location and attention condition ($p < 0.001$). For the univariate analysis, the results from motion-selective regions did not support neural modulations based on task demands. When using IEM, there was a greater response to the stimulus in motion-selective regions as compared to when attending to fixation (color: $p = 0.001$; motion: $p = 0.001$). This indicates that motion-selective regions are responsive to task demands and that this may only be identifiable with more sensitive multivariate approaches, such as IEM.

Next, to evaluate the spatial selectivity of these modulations, we computed the difference in map activation between the stimulus and opposite locations for each attention condition (Fig. 12C). A 2-way repeated measures ANOVA with attention condition (attend fixation, color, and motion) and ROI (TO1/TO2 and hV4/VO1/VO2) as factors reveal a main effect of ROI ($p < 0.001$) and an interaction between ROI and attention condition ($p = 0.006$). Follow up comparisons indicate that in color-selective regions, the stimulus location was more strongly represented than the opposite location when attending to a feature of the stimulus (attend

fixation vs attend color: $p < 0.001$; attend fixation vs attend motion: $p < 0.001$). Furthermore, the stimulus location was more strongly represented when attending to color than when attending to motion ($p = 0.019$), indicating that neural modulation via goals was localized to the stimulus. Motion regions also exhibited a stronger response to the stimulus than when attending fixation as indicated by a stronger response when attending color than attending to fixation ($p < 0.001$) and when attending to motion versus attending fixation ($p < 0.001$). There was no difference between the attend color and motion conditions ($p = 0.745$). It is possible that TO1/TO2 are partly influenced by global modulations of attention, which would lead to increased responses at the stimulus and opposite location primarily when attending to motion (Martínez-Trujillo & Treue, 2002; Treue & Trujillo, 1999b). If so, the difference between locations would have a greater impact on attend motion trials relative to attend color trials, making responses more comparable between conditions even though the attend motion stimulus location had greater activation.

The previous analyses provide evidence that motion- and color-selective regions are spatially selective, as they preferentially tracked the location of the stimulus throughout the visual field; however, a key aspect of neural feature dimension maps is that they respond more strongly to stimuli when observer goals direct attention towards their preferred feature dimension. To quantify this, we computed an AMI for each ROI, where positive values indicate a stronger response to the stimulus when attending to motion (Fig. 12D). Color-selective regions showed a significant negative value ($p = 0.001$). Consistent with the univariate analysis, this demonstrates that color regions preferentially responded to the stimulus when attending to color. Unlike the univariate results, motion-selective regions

exhibited a significant positive value (p = 0.009), meaning that they had a preferential response to the stimulus when attending to motion. Thus, when using a multivariate approach, there is critical evidence that both color- and motion-selective regions represent neural feature dimension maps, as they selectively respond to the locations of important feature-specific stimuli in the visual field.

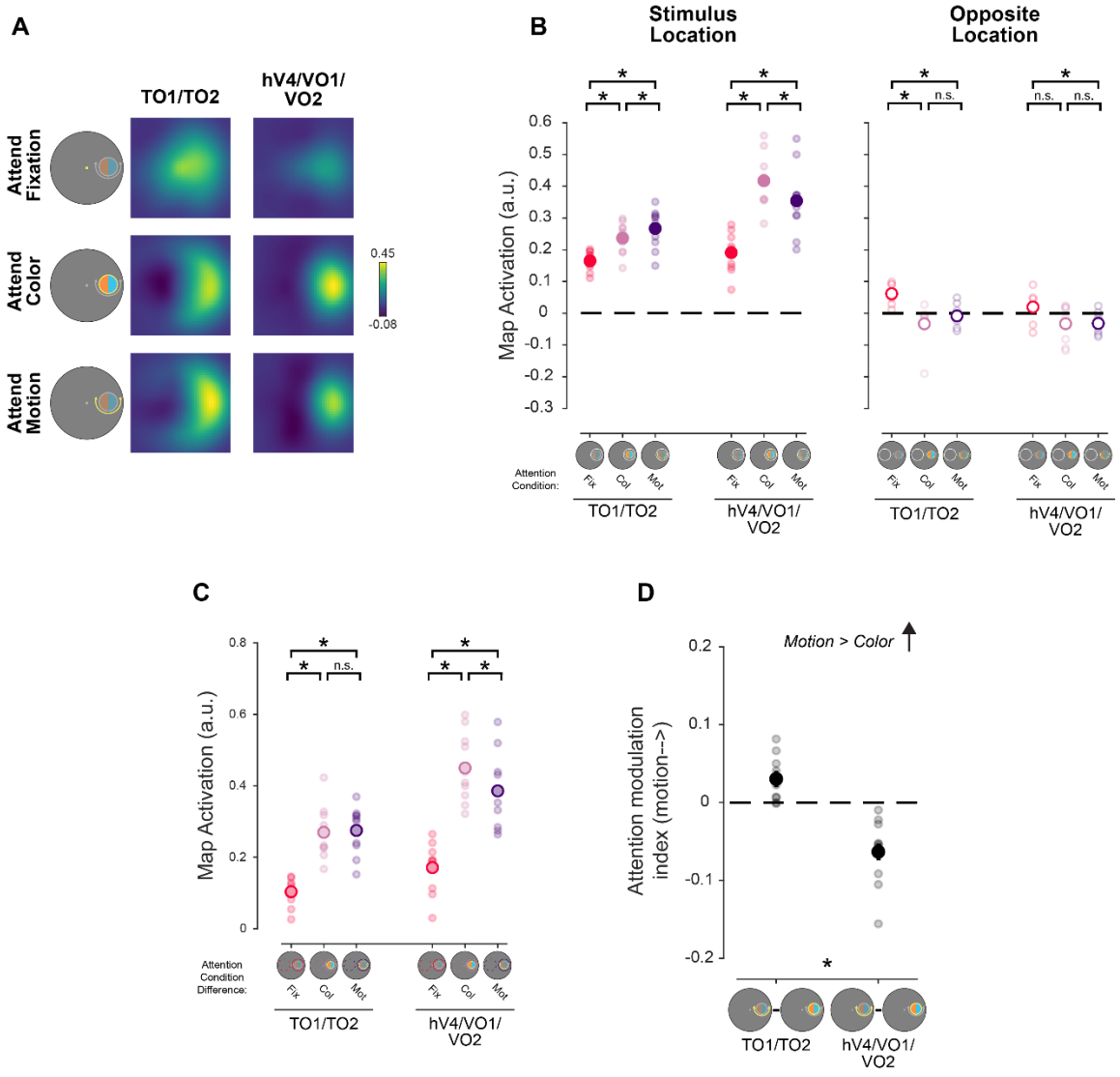

Figure 12: Reconstructed spatial maps support both neural color and motion dimension maps. A. Reconstructions from motion- and color-selective regions from the attend fixation, color, and motion conditions. Qualitatively, there is a strong response at the aligned location, particularly when attending to either the motion or color of the stimulus. **B.** We determined whether each region was modulated by task

demands by assessing the strength of the stimulus representation for each stimulus condition as well as the opposite location. For both regions, the neural response at the stimulus location was stronger when attending to either feature of the stimulus as compared to attend fixation trials, indicating that these regions track general task demands. At the opposite location, responses were stronger when attending to fixation. This is consistent with task demands modulating neural responses only locally, rather than in conjunction with a global modulation. **C.** Difference in map activation between the stimulus and opposite locations for each stimulus condition. Responses were always greater when attending to either the color or motion of the stimulus for both regions. **D.** Difference in map activation between the attend motion and attend color conditions. Responses reliably differed from zero for both ROIs, where motion-selective areas had stronger response when attending to motion and color regions had a stronger response when attending to color. Responses differed between regions as well. * Significant difference based on permuted paired-samples *t* test ($p < 0.05$). Error bars indicate SEM within participants.

Neural and behavioral modulations of task repetition.

Observer goals are crucial for indexing priority in the visual field (Wolfe, 1994), and the multivariate results we present provide the key evidence describing TO1/TO2 and hV4/VO1/VO2 as neural motion and color dimension maps. However, another way which priority can be modulated is through repeated encounters with a stimulus or task instruction (Awh et al., 2012; Luck et al., 2020). Specifically, if a task has been performed in the past, then that can adjust how priority is computed when that task is repeated (Failing et al., 2019; Stilwell & Gaspelin, 2021; Thayer et al., 2022), even when the knowledge of task repetition is implicit (Gao & Theeuwes, 2022). First, we determined whether task repetition influenced behavioral performance during the feature-selective attention task (Fig. 13A). For each attention condition, we sorted trials based on whether participants performed a 'repeat' trial, meaning that the attended stimulus property was the same at least twice in a row (e.g., previous trial was attend motion and current trial is attend motion). We then compared those to 'switch' trials, which were trials where the currently attended stimulus property was different from the previous trial (e.g., previous trial was attend color and current trial is attend motion). By comparing behavioral response times between switch and repeat trials, we

could determine if repeated task demands improved performance (Maljkovic & Nakayama, 1994). We conducted a permuted 2-way repeated measures ANOVA with attention condition (attend color and attend motion) and task repetition (switch and repeat) as factors. There was a main effect of task repetition ($p = 0.039$). Follow-up comparisons show that performance was improved on attend motion repeat trials relative to attend motion switch trials ($p = 0.012$). There was no benefit observed on attend color repeat trials compared to attend color switch trials ($p = 0.148$).

Since there was only a behavioral benefit for the attend motion trials, we only expected a modulation in neural response to the stimulus in motion-selective regions, as the neural motion dimension map should be reflecting this change in prioritization. We evaluated the neural response at the stimulus location within reconstructed spatial maps (Fig. 13B) on switch and repeat trials for each stimulus attention condition (attend motion or attend color). To be complete, we evaluated the neural response for motion- and color-selective regions (Fig. 13C). We conducted a permuted 3-way repeated measures ANOVA, with factors of attention condition (attend motion and attend color), ROI (TO1/TO2 and hV4/VO1/VO2), and task repetition (repeat and switch trials). There was a 2-way interaction between attention condition and ROI ($p = 0.005$).

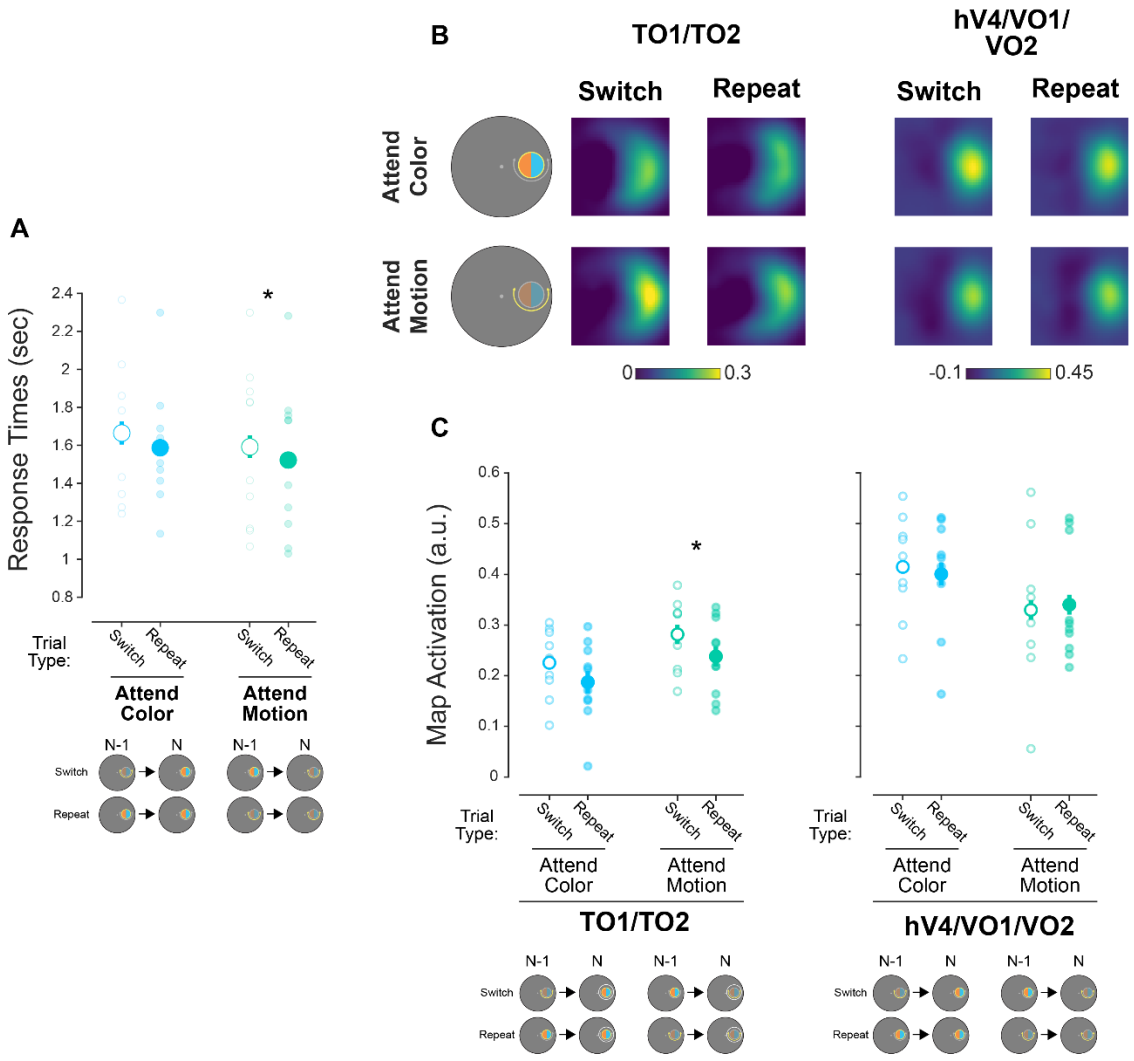

Figure 13: Task repetition modulates behavioral and neural responses for motion. **A.** Behavioral response times for the attend color and attend motion conditions binned by trials where a given trial was the same task at least twice in a row (repeat) or different from the previous trial (switch). Responses only benefited from task repetition on attend motion repeat trials when compared to motion switch trials. **B.** Reconstructed spatial maps from TO1/TO2 and hV4/VO1/VO2 for attention motion/color repeat and switch trials. Even with a reduced number of trials contributing to reconstructions, there was a qualitatively strong response at the stimulus position. **C.** Map activation at the stimulus location for attend motion and attend color switch/repeat trials. In only TO1/TO2, there was a reduced response to the stimulus on repeat trials than on switch trials. This mirrors the behavioral results, as there was only a difference in RT when attending to motion. * Significant difference based on permuted paired-samples *t* test ($p < 0.05$). Error bars indicate within subject SEM.

We conducted permuted 2-way ANOVAs for each feature-selective region with attention condition (attend motion and attend color) and task repetition (switch and repeat trials) as

factors. For color-selective regions, we conducted a permuted 2-way ANOVA. There was a main effect of attention condition ($p = 0.021$), which was due to a stronger neural response when attending to the color of the stimulus. For motion-selective regions, the permuted 2-way repeated measures ANOVA revealed a main effect of attention condition ($p = 0.018$) and task repetition ($p = 0.022$) indicating that there was an overall stronger response to the motion stimulus—consistent with previous analyses—and, surprisingly, a weaker neural response on repeat trials than switch trials. Follow-up two-sample *t* tests indicate that in motion-selective regions, the stimulus representation on attend motion trials was weaker on repeat than switch trials ($p = 0.031$) but this difference of stimulus representation strength was not observed on attend color trials ($p = 0.133$). Since the difference in neural response was *only* observed in motion-selective regions and the behavioral benefit of task repetition was *only* present on attend motion trials, this is suggestive that task repetition modulates priority within neural motion dimension maps. Additionally, since the stimulus was effectively the same on each trial, decreased neural response cannot be explained by perceptual neural adaptation (see Discussion). Overall, the multivariate analyses indicate that responses in color- and motion-selective regions allow us to conclude these regions represent neural feature dimension maps within the priority map framework.

Discussion

The goal of the current study was to establish whether visual cortex contains distinct neural spatial maps that preferentially track the location of stimuli when specific feature dimensions of those stimuli are goal relevant (Fig. 8). We presented a colorful moving dot stimulus on each trial, but cued participants to attend to either the color or motion of that stimulus (Fig.

9). Our design isolated the influence of observer goals because the goal-relevant feature dimension could change on each trial, but the visual input was consistent across the entire experiment. Univariate measures support color-selective regions hV4/VO1/VO2 as a representative neural color map but fall short of supporting motion-selective areas TO1/TO2 as a motion dimension map (Fig. 11). However, by using an inverted encoding model (Fig. 10) to reconstruct spatial maps from both motion- and color-selective regions, we were able to show that stimulus representation in both regions were modulated by task demands and was strongest when attending to each region's preferred feature (Fig. 12). Additionally, motion-selective regions were modulated by task repetition, which mirrored behavioral response times benefits that were only observed on attend motion trials (Fig. 13). These findings provide the necessary evidence to identify feature-selective retinotopic regions as 'neural feature dimension maps'—a critical component of priority map theory.

There are several properties that need to be reflected in a neural feature dimension map: spatial selectivity, feature selectivity, and computation of important information at specific locations. The findings in this study supports color- and motion-selective regions as neural feature dimension maps, but it additionally provides a complement to previous findings from our lab (Thayer & Sprague, 2023). In this report, we presented dot displays which contained either salient color or motion stimuli that were ignored. In the same feature-selective regions, reconstructed spatial maps showed a stronger response at the location of a stimulus when the salient feature dimension matched the feature selectivity of each region. This was considerable evidence supporting color- and motion-selective regions as neural feature dimension maps, but nearly all priority map models emphasize how observer goals sculpt

priority. Thus, in conjunction with these previous findings, there is now mounting support for the role of feature-selective regions in computing attentional priority.

Previous reports (McMains et al., 2007; Runeson et al., 2013) have similarly investigated how color- and motion-selective regions are modulated by task demands. In both studies, participants viewed colorful moving dot stimuli and reported that color- and motion-selective regions were preferentially modulated when attending to the preferred feature of each region. However, neither study indexed the spatial selectivity of feature response modulations, so they may have occurred throughout the entire visual field. These findings fall short of establishing an area as a neural feature dimension map, as a critical aspect of these maps is that the stimulus location should be prioritized above all other locations if there is nothing else salient or relevant in the visual field. This is a limitation of univariate averaging approaches. Because univariate methods usually average across voxels with variable spatial selectivity (Kriegeskorte & Bandettini, 2007; Serences & Saproo, 2012), it is difficult to ascertain any map-like structure in these regions. We attempted to circumvent this limitation by leveraging a vRF encoding model (Dumoulin & Wandell, 2008) to identify voxels that preferentially responded to the stimulus location. Even though this method is still univariate, as models are fit to individual voxels, it is a powerful tool for assessing spatially selective responses (Wandell & Winawer, 2015). While there was now satisfactory evidence to support hV4/VO1/VO2 as a color dimension map, TO1/TO2 showed no feature selective response to the stimulus. In fact, voxels corresponding to the stimulus location in TO1/TO2 were not modulated by task demands at all. However, using inverted encoding models, we were able to provide the necessary evidence for both color and motion regions to identify

them as neural feature dimension maps. IEM leverages weighted responses across voxel activation patterns, which what led to the detectability of task demand modulations in motion regions (Serences & Saproo, 2012).

The primary goal of this study was to provide evidence that neural feature dimension maps track the location of goal-relevant feature-selective stimuli throughout the visual field, but information can also be prioritized through repeated exposure to a stimulus (Awh et al., 2012; Luck et al., 2020). History with a stimulus can enhance or suppress specific locations or features (Adam & Serences, 2020; Addleman & Störmer, 2023; Maljkovic & Nakayama, 1994; Stilwell et al., 2019; Thayer et al., 2022), and it was possible to evaluate the influence of task repetition in the current study. Due to the trial-by-trial variability of our design, we could assess how task repetition adjusted the stimulus representation in neural feature dimension maps. Behaviorally, there was a response time benefit on attend motion trials when that task repeated (Maljkovic & Nakayama, 1994; Wolfe et al., 2003). Interestingly, only the neural response in motion-selective regions changed as a function of repetition, as indicated by decreased stimulus strength. Initially, this may appear to contradict the behavioral finding, because if activation within the motion dimension map corresponds to prioritization, then perhaps the stimulus representation should be stronger on repeat trials. However, it may be that task repetition makes it easier to read out information from a region. This could occur through greater inhibitory signals in neural populations that code task-irrelevant information, which would help minimize the variability of neural responses by sharpening task-relevant information (Ringo, 1996; Schacter & Buckner, 1998; Wiggs & Martin, 1998). It may be that neural populations in TO1/TO2 are suppressed at the stimulus

location that code task-irrelevant feature information. For example, some individual neurons may have receptive fields that overlap with the stimulus but are not tuned to the specific motion directions used in our task. Since these neurons do not contribute to motion discrimination, and since spatial information is not needed to complete the task, those neurons may be inhibited. This would lead to a weaker spatial representation of the stimulus but one that contains the same amount of task-relevant information, and presumably an improved representation of the goal-relevant motion direction. Future studies should test this interpretation by applying variants of the IEM approach that model feature information, like motion or color (Brouwer & Heeger, 2009; Sahan et al., 2020).

Neural adaptation studies may also support this interpretation. Neural adaption refers to the decreased neural response that occurs when stimuli with the same properties are repeatedly presented and is only observed in neurons that are tuned for the repeated stimulus feature(s) (Blakemore & Campbell, 1969). This is used as evidence in fMRI studies to identify feature-selective regions (Grill-Spector & Malach, 2001) and can occur through perceptual or conceptual repetitions (Schacter & Buckner, 1998). It is unlikely that the decreased neural response in the current study is due to perceptual neural adaptation because the stimulus was effectively identical throughout the study. Even though most adaptation studies observe decreased neural responses when the physical properties of a stimulus are the same across presentations, adaption can occur through iterative task demands (Larsson & Smith, 2012). Since stimuli in the current study were always identical throughout the experiment, then task repetition must be driving the difference between switch and repeat trials in TO1/TO2. This bolsters the claim that motion-selective regions represent a motion dimension map, as task

repetition is a form of selection history, which is another modulatory factor described by priority map theory. Even though repetition effects were not observed in color regions, there was no response time benefit due to task repeats. Because no behavioral effect was present, there should not be any observed neural effect. In future studies, the task could be optimized for repetition effects when attending to color, and perhaps then there would be an observable neural effect.

In sum, we found that feature-selective retinotopic regions in visual cortex track the location of stimuli throughout the visual field and had a stronger response to those stimuli when each region's preferred feature was goal-relevant. This is necessary evidence for priority map theory, which identifies these regions as neural feature dimension maps. These findings solidify TO1/TO2 and hV4/VO1/VO2 as necessary in processing important color and motion information in the visual field, which can be exploited to fully understand how we successfully navigate our visual environment.

Chapter IV: Behavioral Prioritization of Learned Features

Introduction

The visual system is constantly bombarded with information, of which only a small portion can be attended. When searching the kitchen for ingredients to make pizza, features and locations in the kitchen that are aligned with the goal of making pizza get prioritized. For example, one may prioritize search for the red tomato sauce in the cabinet. Search for pizza-related items can be disrupted, such as when there is a salient, unexpected, and abrupt appearance of a roommate in the kitchen. The processing of other salient items in the scene, such as the bright green parsley growing in the window, are not disruptive, but instead might be suppressed due to their regular presence in the kitchen. The ability to prioritize specific information based on one's goals, the automatic capture from abrupt onset salient stimuli, and the learned suppression of regularly presented items all interact to produce the phenomenon of attentional control (Awh et al., 2012; Luck et al., 2020).

The interplay among these signals has been characterized within the priority map framework (Itti & Koch, 2001; Koch & Ullman, 1985; A. M. Treisman & Gelade, 1980; Wolfe, 1994; Zelinsky & Bisley, 2015), where a priority map reflects the importance of specific locations within the visual field. To compute a feature-agnostic priority map, individual maps of specific feature dimensions (e.g., color or orientation), which contain information corresponding to locations that are important based on being physically salient as well as based on their relevance for ongoing goals, are summed. Bottom-up and top-down inputs have been well established to drive attentional selection through behavioral (Bundesen, 1990; Duncan & Humphreys, 1989; Olivers et al., 2006) and neural (Fecteau & Munoz, 2006;

Gottlieb et al., 1998; Serences & Yantis, 2006) studies and can modulate priority at the level of individual feature maps (McMains et al., 2007; Runeson et al., 2013; Saenz et al., 2002; Serences & Boynton, 2007) or an integrated priority map (Bisley & Goldberg, 2003, 2006; Bogler et al., 2011, 2013).

The contribution of a third category, selection history, has been proposed due to results that do not adhere to the canonical top-down/bottom-up dichotomy (Awh et al., 2012; Shomstein et al., 2022). Selection history is distinct from top-down attention, as the influence of previous deployments of attention modulate priority without the explicit awareness of an individual and can even interfere with ongoing goals (Hickey et al., 2010). Additionally, selection history is distinct from bottom-up salience because selection history clearly cannot influence the physical properties of stimuli which render them salient.

One primary means by which selection history influences the allocation of attention is by deprioritizing regularly presented distractors (Gaspelin et al., 2019; Stilwell et al., 2019; Wang & Theeuwes, 2018). Such distractor suppression is often studied using the additional singleton paradigm (Theeuwes, 1991, 1992). Briefly, this task commonly involves searching for a target shape among various distractor shapes, such as a target diamond among circle distractors. On some trials, a critical distractor appears that is presented in a distinct color from the rest of the display (e.g., red distractor among green items). When present, this distractor tends to slow response times (RTs), which is due to attention being directed to the location of the distractor based on its salience (Jonides & Yantis, 1988; Theeuwes et al., 2003). However, when the critical distractor is regularly presented at a specific position within the search array, capture effects are diminished (Stilwell et al., 2019), or even

completely abolished such that performance is the same as distractor-absent trials (Wang & Theeuwes, 2018). This modulation occurs without explicit knowledge of the location regularities (Gao & Theeuwes, 2022), indicating a process distinct from top-down influences.

Suppression is thought to occur via two mechanisms: proactive inhibition and reactive rejection (Geng, 2014). Proactive inhibition deprioritizes information prior to the onset of a visual display. For instance, fewer saccades are directed towards the location where a singleton was usually presented than any other location (Gaspelin et al., 2019; Stilwell & Vecera, 2022), consistent with the possibility that the learned location was suppressed prior to display onset. Reactive mechanisms involve the rapid disengagement from distracting information after covert or overt attention has already been captured (Theeuwes, 2010). They are thought to act primarily within a spatial context, as evidence shows suppression restricted to a specific location (Theeuwes et al., 2003). Thus, mechanisms of suppression likely act on a feature-agnostic priority map, and not necessarily at the level of individual feature dimension maps (Luck et al., 2020). This raises the question: to what extent do nonspatial stimulus features (e.g., color hue, shape) contribute to distractor suppression?

The additional singleton paradigm lends itself to investigating the learned suppression of features such as color (Failing et al., 2019; Stilwell & Gaspelin, 2021; Vatterott & Vecera, 2012). For example, Stilwell, Bahle, and Vecera (2019) reported that when the location of a singleton is completely randomized, but presented in one high-probability color, RTs were faster than when the singleton was a low-probability color. This is consistent with participants suppressing specific color values when beneficial for task performance. However, an important aspect of the visual search tasks used in previous research is that each

item in the display has a distinct spatial position, which inserts ambiguities on whether feature control mechanisms were implemented independent of space; it could be the case that only after a feature singleton captures attention, then reactive mechanisms suppress the *location* corresponding to the salient singleton (Luck et al., 2020; Moher & Egeth, 2012; Theeuwes et al., 2003). One way to disentangle the influences of features and space is to demonstrate feature-specific deprioritization *independent* of location.

A common procedure to minimize the impact of space is to use overlapping stimuli (Duncan, 1984; Giesbrecht et al., 2003; Liu et al., 2003; O'Craven et al., 1999; Saenz et al., 2002; Yantis & Serences, 2003). This way, spatial location is shared among stimuli, which isolates feature-specific mechanisms and minimizes the ability of a spatially driven mechanism to selectively suppress one, but not another, stimulus. We adopted this strategy in the current study by having participants perform an orientation discrimination task on two spatially overlapping colored line arrays. In this task (Fig. 1A), participants identified which of two arrays had more lines, then determined the orientation of the higher-density ('target') line array. Critically, the low-density ('distractor') array was typically presented in one color (Fig. 1B). If feature control mechanisms can specifically suppress the representation of a stimulus without necessarily suppressing all stimuli at a given location, then we expected behavioral performance to be faster when the distractor array was presented in the high-probability color. However, if reactive mechanisms are suppressing the location corresponding to the distractor array, then the target array would also be suppressed due to their spatial overlap. If this latter account is true, then we would expect to see no difference

in behavioral performance whenever the distractor array was shown in the high-probability color or any of the low-probability colors.

Additionally, it is imperative to pinpoint the duration of suppression effects, as it is informative about the mechanism of prioritization (Wöstmann et al., 2021). Studies have shown that inter-trial priming, or the influence of the previous trial on current trial performance (Maljkovic & Nakayama, 1994), and statistical learning, or the extraction of long-term display regularities to adjust future performance (Jiang, 2018; Jiang et al., 2013; Vatterott & Vecera, 2012), both influence distractor suppression.

To isolate the duration of feature suppression, we included several blocks in which color regularities were removed from the display. In Experiment 1, during these regularity-absent blocks, the distractor array had an equal chance of being shown in any of the possible colors (Fig. 14C). Whereas, in Experiment 2, both the target and distractor array had an equal chance of being presented in the previously high-probability color (Fig. 14C). These regularity-absent blocks allowed us to determine the specific mode of suppression by evaluating whether effects persisted after learning blocks, consistent with statistical learning, or whether they were primarily driven by inter-trial priming within the learning phase itself.

In both experiments, we found robust suppression of the high-probability distractor color when regularities were present. Furthermore, subjects who showed the suppression effect when regularities were present continued to suppress the high-probability color when regularities were absent. In Experiment 2, we found additional evidence for long-term distractor suppression, as RTs were slower when the target array was presented in the previously learned high-probability distractor color. Overall, our results demonstrate that

learned distractor colors can be suppressed independent of a spatial suppression mechanism, and that this suppression is supported by statistical learning of distractor feature values.

Experiment 1

The goal of Experiment 1 was to determine whether feature-specific suppression occurs when stimuli are spatially overlapping. If so, this would suggest that feature control mechanisms can be independent of spatial control operations. We also sought to test whether suppression was transient, consistent with intertrial priming, or whether suppression persisted over longer periods of time, consistent with statistical learning.

Method

Participants.

The study protocol was approved by the UCSB institutional review board. Twenty-four participants (16 female, mean age = 18.5 years) were recruited from the University of California, Santa Barbara (UCSB) subject pool. All participants reported normal or corrected-to-normal vision and either received course credit or $10/hr upon completing the experimental session. Participants gave written consent prior to participating in the study. Previous work investigating color suppression using a visual search task (Stilwell et al., 2019) reported an effect size of η_p^2 = 0.68 and a power analysis using this effect size indicated that 4 subjects were needed to obtain 80% power. Since our study used a different task, we collected data from 24 participants to ensure enough statistical power to detect effects in our experiments.

Apparatus and stimuli.

Participants viewed stimuli in a darkened room on a 25-in LED-backlit LCD screen with a resolution of 2560 x 1440 pixels. They were seated approximately 60 cm away from the screen. Stimuli were presented using Matlab and Psychtoolbox (Brainard, 1997).

A white (80.1 cd/m^2) dot centered at fixation was presented at the start of each block with a radius of 0.15° visual angle against a gray (49.4 cd/m^2) background (Fig. 1). The fixation stimulus was visible throughout the whole block. On each trial, two oriented line arrays were presented. All lines in one array were oriented 45° clockwise of vertical, while the lines of the other array were oriented 45° counterclockwise. The orientation of the line arrays was randomized on each trial. Jitter was independently added to the orientation of both arrays randomly selected from 0.3-1.2° orientation. Both arrays were presented within an imaginary circle with a radius of 10.5° visual angle. One array always contained 60 ± 20 (randomly selected on each trial) more lines than the other array. The array with more lines was the 'target', while the other array was the 'distractor'. The number of lines in the target array had a range of 150-170 lines, while the distractor array could contain 90-110 lines. Individual lines had a length of 1.5° visual angle and a width of 0.05° visual angle. The color of either array was selected from the following four isoluminant colors in CIE color space: green (40.7 cd/m^2, x = 0.243, y = 0.397), red (40.6 cd/m^2, x = 0.421, y = 0.285), blue (40.3 cd/m^2, x = 0.182, y = 0.175), and yellow (40.3 cd/m^2, x = 0.450, y = 0.481). The target array was always a different color from the distractor array. Feedback text at the end of each block was presented in gray Arial font (RGB: 100, 100, 100). Participants reported whether the target array was oriented clockwise or counterclockwise from vertical with a left or right button press using a USB response pad.

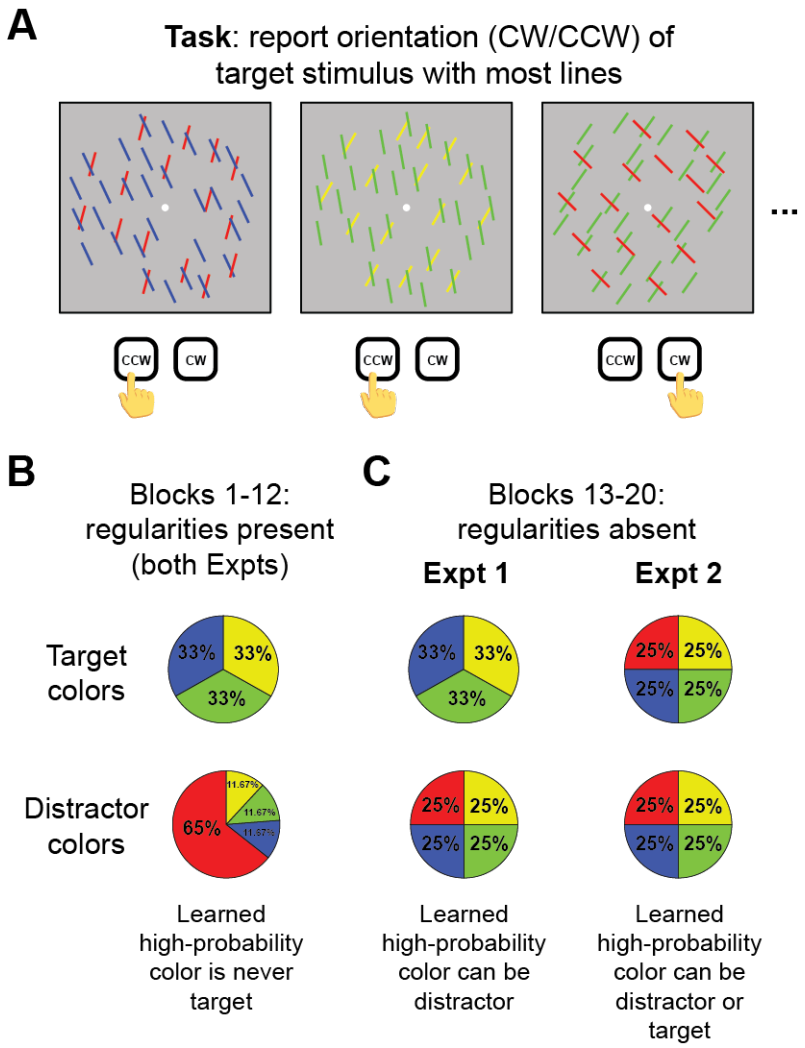

Figure 14. Discrimination task. (A) On each trial, participants were shown two oriented line arrays, each presented in one of four different equiluminant colors. One array was tilted clockwise from vertical and the other was counterclockwise from vertical. Participants determined which array had the most lines, and then reported the orientation of that array with a button press. **(B)** Color regularities were present during the first 12 blocks of both experiments, such that the array with fewer lines was usually presented in the high-probability color (65% of trials). The target array was never presented in the high-probability color when regularities were present. **(C)** Regularities were removed in the last 8 blocks of both experiments, meaning that the array with the fewest lines had an equal chance of being any of the four possible colors (25%). In Experiment 1, the array with the most lines was still never presented in the previously high-probability distractor color. Experiment 2 allowed both the target and distractor line arrays to be shown in any of the four colors with equal probability. The arrays were never presented in the same color on a given trial. Images here are illustrative cartoons; actual colors were equiluminant and line density/orientation are described in detail in Methods.

Design and procedure.

The fixation dot was presented at the start of the experiment and was visible throughout the whole block of 60 trials. At the start of each trial, the fixation dot was presented alone for 1,000 ms. Participants were instructed to attend and fixate the central dot until stimulus array onset. Next, the target and distractor line arrays were presented for up to 3,000 ms or until response. Participants determined whether there were more lines tilted counterclockwise or clockwise of vertical and reported the corresponding orientation with a left/right button press. They were encouraged to respond as fast as possible while still being accurate. At the start of the experiment, a random color was selected to be the prevalent distractor color for each subject (selected from red, green, blue, and yellow). During the first 12 blocks of the experiment, on 65% of trials, the distractor array was presented in the selected high-probability color. For the remaining 35% of trials, the distractor array was equally presented in one of the other three low-probability colors (11.67% of trials for each remaining color). The target array was never presented in the high-probability distractor color. The color of the target array was randomly selected from the remaining three colors with equal probability (33% of trials for each color), with the additional stipulation that the target and distractor were always different colors on a given trial. By comparing response time (RT) and accuracy on these *regularity-present blocks*, we could determine whether participants report the target orientation more quickly and accurately when the distractor appeared in a high-probability color.

After the first 12 blocks, where color regularities were present, participants performed 8 more blocks of the discrimination task. During these last 8 blocks, the target array color was

chosen as before (33% of each non-distractor color). However, now the distractor array had an equal chance of being presented in any color (25% of trials for each color). Other than the change in color probabilities, the last 8 blocks were identical to the first 12 blocks. Participants were not informed about a change in target/distractor color probabilities throughout the experiment. By comparing RT and accuracy in these *regularity-absent blocks*, we were able to determine if participants continue to suppress the distractor color even when this is no longer useful. Overall mean accuracy on the task was shown to the participants at the end of each block of the experiment (regularity-present and -absent blocks).

Before starting the main session, participants completed a practice session of the task, which consisted of 60 trials of the orientation report task without any color regularities. There were 60 trials per block of the main session, and participants completed a total of 20 blocks over ~1 hr. Upon completing the experiment, we interviewed participants to determine whether they were aware of the color regularities. First, they were asked if they noticed any patterns or consistencies with the stimuli during the experiment. Second, they were told that the distractor array was usually one color and were instructed to guess the high-probability color.

Data analysis & statistical procedures.

Trials with an RT 2.5 SDs above or below the individual participant's mean RT, along with trials that were faster than 100 ms or slower than 2500 ms, were removed from RT analyses. An average of 4% (SD = 1.51%) of trials were removed per participant after applying these exclusion criteria. We also excluded trials with an inaccurate orientation report from all RT analyses (13.8% of remaining trials). The task was intentionally made difficult to avoid

ceiling effects, which explains the relatively high percentage of inaccurate trials. None of the experiments were preregistered.

We compared mean RT and accuracy on regularity-present blocks using paired sample *t*-tests to determine whether participants reported the target orientation more quickly and accurately when the distractor appeared in a high-probability color. To see if color suppression persisted when regularities were removed from the display, we computed a two-way repeated measures ANOVA with color condition as the first factor (high-probability color vs low-probability colors) and regularity presence as the second factor (regularity-present blocks vs regularity-absent blocks). This analysis was followed by a *t*-test comparison between mean RTs in the high- and low-probability distractor color conditions during regularity-absent blocks. Finally, we computed the linear correlation between suppression observed in regularity-present and regularity-absent blocks, where suppression was defined as the difference in mean RT between low-probability and high-probability distractor color trials. For all pairwise tests, we reported Bayes Factor (BF) results using the bayesFactor package for MATLAB (Krekelberg, 2018). Evidence in favor of the null (BF_{01}) is reported for non-significant tests, and evidence against the null is reported for significant tests (BF_{10}). We used d_z as a metric of effect size for all *t*-test comparisons to account for shared variance in our repeated measures design (Lakens, 2013).

7 subjects correctly identified the high-probability distractor color during a post-experiment interview, which did not differ from chance (binomial test: $p = 0.393$). All reported results were qualitatively the same when excluding participants who correctly reported the high-probability color. We analyzed the regularity-present suppression effect (low-probability -

high-probability RT) separately for those who correctly identified the high-probability color and found no significant difference compared to those who were unaware of the color regularities (Figure 19A).

Data & code availability.

All data (behavioral performance on each trial) and code (experiment presentation, data analysis) necessary to reproduce results presented in this manuscript are freely available at: https://osf.io/h7dg6/.

Results and discussion

Regularity-present performance.

First, we compared RT for target orientation discrimination across all task blocks throughout the experiment (Fig. 15A). Qualitatively, RTs were faster when the distractor appeared in the high-probability distractor color than when it appeared in another color. Additionally, RTs qualitatively sped up through the experiment. Next, we quantitatively established whether participants could more efficiently report the target orientation when a high-probability distractor color was present in the display during regularity-present blocks (Fig. 15B). We compared RTs (averaged across the initial regularity-present blocks 1-12; Fig. 15A) on trials with the high-probability distractor and trials with another color distractor. Correct orientation reports on trials with high-probability distractor color were significantly faster than on trials with low-probability distractor colors ($t(23) = 4.04$, $p < 0.001$, $d_z = 0.83$, $BF_{10} = 63.39$). There was not a significant difference in orientation report accuracy between these trials (Table 3; $t(23) = 1.55$, $p = 0.134$, $d_z = 0.32$, $BF_{01} = 1.63$), indicating that the RT

advantage is not due to a speed-accuracy tradeoff. These results suggest that the high-probability distractor color was suppressed when stimulus regularities were present.

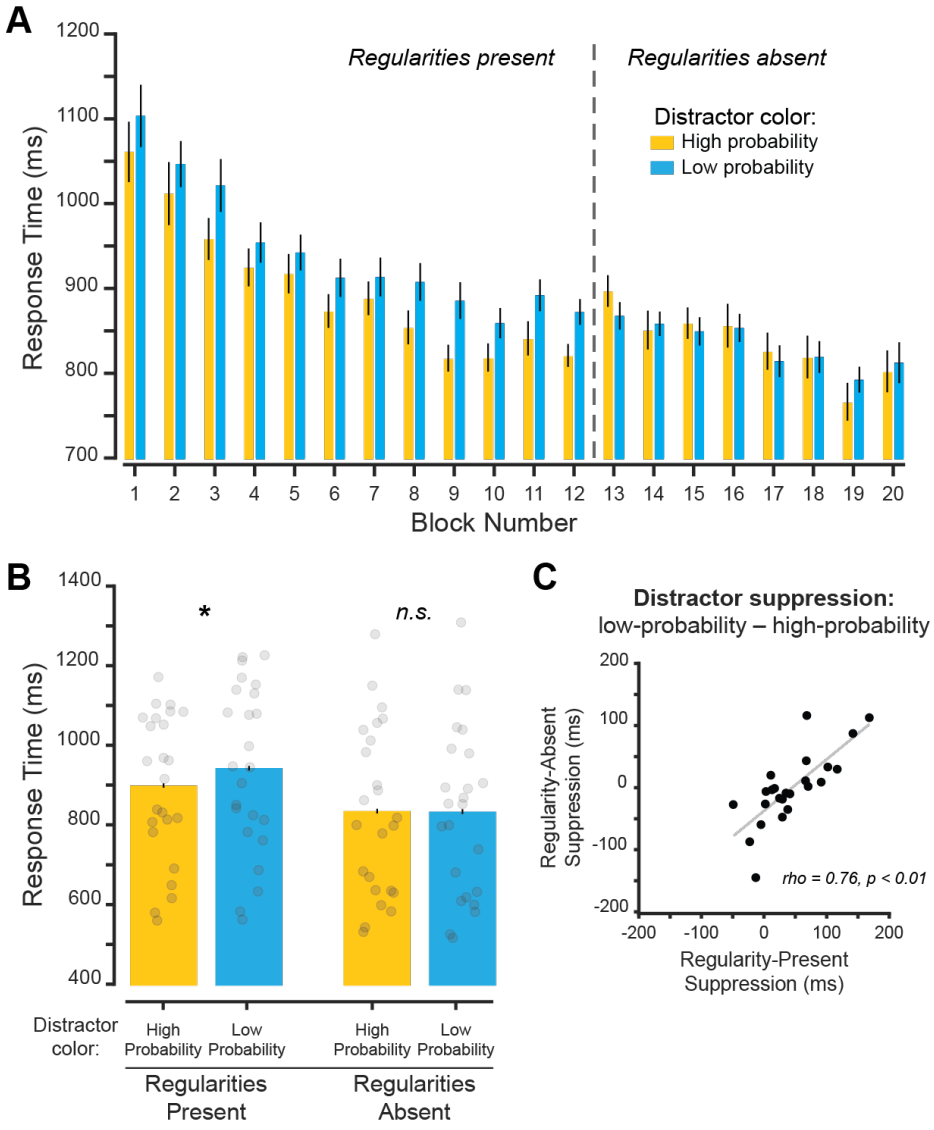

Figure 15. Experiment 1: high-probability distractor color is suppressed during learning and over an extended interval. **(A)** Mean RT for each block on trials with correct orientation reports. Dashed line indicates when distractor color regularities were removed from the display. **(B)** Mean RT across regularity-present and regularity-absent blocks for both high- and low-probability color conditions. Individual subject data points shown. Significant differences between color probability conditions indicated with * for p values < 0.05. **(C)** Correlation between suppression effect during regularity-present blocks and regularity-absent blocks. Suppression effects were computed as the difference in RT between the low- and high-probability color conditions. Error bars are within-subject standard error of the mean.

Table 3. Experiment 1 Accuracy (± SEM)

	Regularities Present	Regularities Absent
High-Probability Color	87.02% (1.08)	85.76% (1.10)
Low-Probability Color	83.65% (1.08)	84.06% (1.10)

Regularity-absent performance.

Next, we identified whether distractor suppression persisted when color regularities were removed (Fig. 15B). To see if the difference between color conditions changed as a function of regularity presence, we performed a two-way repeated measures ANOVA using RT data with distractor array color (high-probability color vs low-probability colors) and regularity phase (regularity-present vs regularity-absent) as factors. There was no effect of color condition, $F(1,23) = 4.01$, $p = 0.057$, $\eta_p^2 = 0.56$, but there was a significant effect of phase $F(1,23) = 18.14$, $p < 0.001$, $\eta_p^2 = 0.96$. The significant main effect of phase reflects the overall faster RTs during the later regularity-absent blocks as participants were getting better at the task. Importantly, there was a significant interaction, $F(1,23) = 32.64$, $p < 0.001$, $\eta_p^2 = 0.59$. This result suggests that, at a group level, once regularities were removed from the display, the previously high-probability color was no longer suppressed. Follow-up comparisons between high- and low-probability distractor colors using data from the regularity-absent blocks are consistent with this conclusion: there was no significant difference between RTs when the distractor appeared in the previously high- vs low-probability distractor color ($t(23) = 0.11$, $p = 0.917$, $d_z = 0.02$, $BF_{01} = 4.64$).

While the above results suggest that, on average, participants no longer suppress the distractor color with learned regularities when the regularities are removed, we next considered the possibility that the magnitude of suppression during the regularity-present blocks within individual participants was carried over to the regularity-absent blocks. That is—do participants who most strongly suppress the learned distractor color when regularities are present also suppress the distractor color more than other participants when regularities are removed?

To test this, we calculated the correlation between the amount of suppression (defined as the difference in mean RT between low-probability and high-probability distractor color trials) during the regularity-present blocks and during the regularity-absent blocks (Fig. 15C). There was a strong positive relationship between these variables ($r = 0.76$, $p < 0.001$), indicating that subjects who suppressed the distractor during regularity-present blocks continued to suppress the distractor during regularity-absent blocks, despite no overall mean difference in the regularity-absent blocks across our participant sample (Fig. 15B).

Together, these findings show that a high-probability distractor color can be suppressed when regularities are present within a block. Critically, suppression occurred even when the arrays spatially overlapped, indicating that a specific color can be suppressed independent of spatial location. It is possible that suppression in this experiment was due to inter-trial priming (Maljkovic & Nakayama, 1994), as suppression did not, on average, persist across our participant sample once regularities were removed. But, an analysis of individual participants showed that those with stronger suppression effects when regularities were present continued to suppress the distractor color once regularities were removed (Fig. 15C). In addition to

evaluating continued distractor suppression, previous studies have observed suppression effects when a target stimulus is presented at a learned distractor location (Britton & Anderson, 2020; B. Wang & Theeuwes, 2018). Since the target array was never presented in the high-probability color, it was not possible to conduct a similar analysis in Experiment 1. Experiment 2 was designed to better understand whether non-spatial color suppression mechanisms are transient or whether they result in suppression that persists over longer periods of time by including trials to directly measure suppression of the target array when regularities are removed (Fig. 14).

Experiment 2

In Experiment 1, we showed that a regularly presented distractor color can be suppressed even when suppression cannot operate via a spatial location. Suppression occurred if regularities were present, but not once regularities were removed, consistent with a transient suppression effect such as inter-trial priming (Maljkovic & Nakayama, 1994). However, spatial and color regularities have been shown to persist beyond the effects of priming in previous studies using visual search paradigms (Stilwell et al., 2019; Vatterott & Vecera, 2012). There was a hint of this effect at the level of individual subjects, where those who suppressed the distractor during regularity-present blocks continued to suppress the distractor during regularity-absent blocks. It may have been difficult to observe continued suppression during regularity-absent blocks due to relatively fast performance. In Experiment 2, we aimed to better probe the persistence of these suppression effects by evaluating performance in the regularity-absent blocks when the previously high-probability color appears as the target array (Britton & Anderson, 2020; B. Wang & Theeuwes, 2018). If the learned color is

being suppressed, and the persistence of this effect was masked due to fast performance in the regularity-absent blocks, then suppressing the target array may allow for long-term suppression to be more readily observed via a *slowing* in discrimination performance when the target appears in the previously high-probability distractor color.

Method

Participants.

We recruited 24 new participants (18 female, mean age = 20 years) from the UCSB subject pool. Subjects were compensated with either course credit or $10/hr upon completing the task. None of the participants recruited for Experiment 2 participated in Experiment 1.

Design and procedure.

Experiment 2 was identical to Experiment 1 during the regularity-present blocks (blocks 1-12). The one critical change occurred in the regularity-absent blocks (blocks 13-20; Fig. 14B). Similar to Experiment 1, the previously high-probability color had an equal chance of being the distractor array color (25% for each color). However, now the target array was presented in any of the four colors with equal probability (25% for each color), with the stipulation that the target and distractor colors were nonidentical. If suppression is due to long-term learning, then we would expect to see slower RTs when the target array was presented in the previously high-probability color. Furthermore, as in Experiment 1, it is possible that continued suppression effects persist in the regularity-absent blocks when the distractor array is presented in the high-probability color. However, due to the results of

Experiment 1 (Fig. 15), we primarily expected this suppression effect to occur on the individual-subject level.

Analysis & statistical procedures.

We removed trials that were faster than 100 msec and slower than 2,500 msec as well as trials 2.5 SDs above or below individual subject means. An average of 4.28% (SD = 1.59%) of trials were removed per participant. Trials with inaccurate responses were also removed from RT analyses (14.72% of trials). 6 participants correctly reported their high-probability color, which did not differ from chance (binomial test: $p = 0.578$). Results are qualitatively the same when we exclude participants who correctly identified the high-probability color. Specifically, there was no difference in regularity-present distractor suppression between those who were aware and unaware of the high-probability color (Figure 19B).

The same statistical tests computed for Experiment 1 were conducted in Experiment 2 when evaluating the influence of distractor array probabilities on performance. Additionally, to see if the high-probability color was suppressed during regularity-absent blocks, we computed a two-way repeated measures ANOVA with *target* array color as the first factor (previously high-probability distractor color vs previously low-probability distractor color) and block as the second factor (regularity-absent blocks 1 through 6). This was followed by a paired samples *t*-test comparing the mean RT across regularity-absent blocks of high- and low-probability target color conditions. Finally, we computed the linear correlation between *distractor* suppression in the regularity-present blocks and *target* suppression in the regularity-absent blocks to evaluate individual subject long-term suppression.

Results and discussion

Regularity-present performance.

First, we verified that we could replicate the suppression effect observed in Experiment 1 during the regularity-present blocks (Fig. 15). Matching the results from Experiment 1, we saw that RTs were qualitatively faster when the high-probability color was shown as compared to the low-probability colors and that RTs increased throughout the experiment (Fig. 16A). We then compared the mean RTs from the regularity-present blocks between trials with a high-probability and low-probability distractor color (Fig. 16B). RT was faster in the high-probability color condition than the low-probability color condition ($t(23) = 4.17$, $p < 0.001$, $d_z = 0.85$, $BF_{10} = 84.87$). This replicates the main findings in Experiment 1, where the high-probability distractor color was suppressed when regularities were present, resulting in faster target discrimination performance. In addition, accuracy was greater for the high-probability distractor color condition (Table 4; $t(23) = 2.77$, $p = 0.011$, $d_z = 0.57$, $BF_{10} = 4.53$). Our accuracy results indicate that there was no speed-accuracy tradeoff and that target identification accuracy was improved when the prevalent distractor color was present in the array.

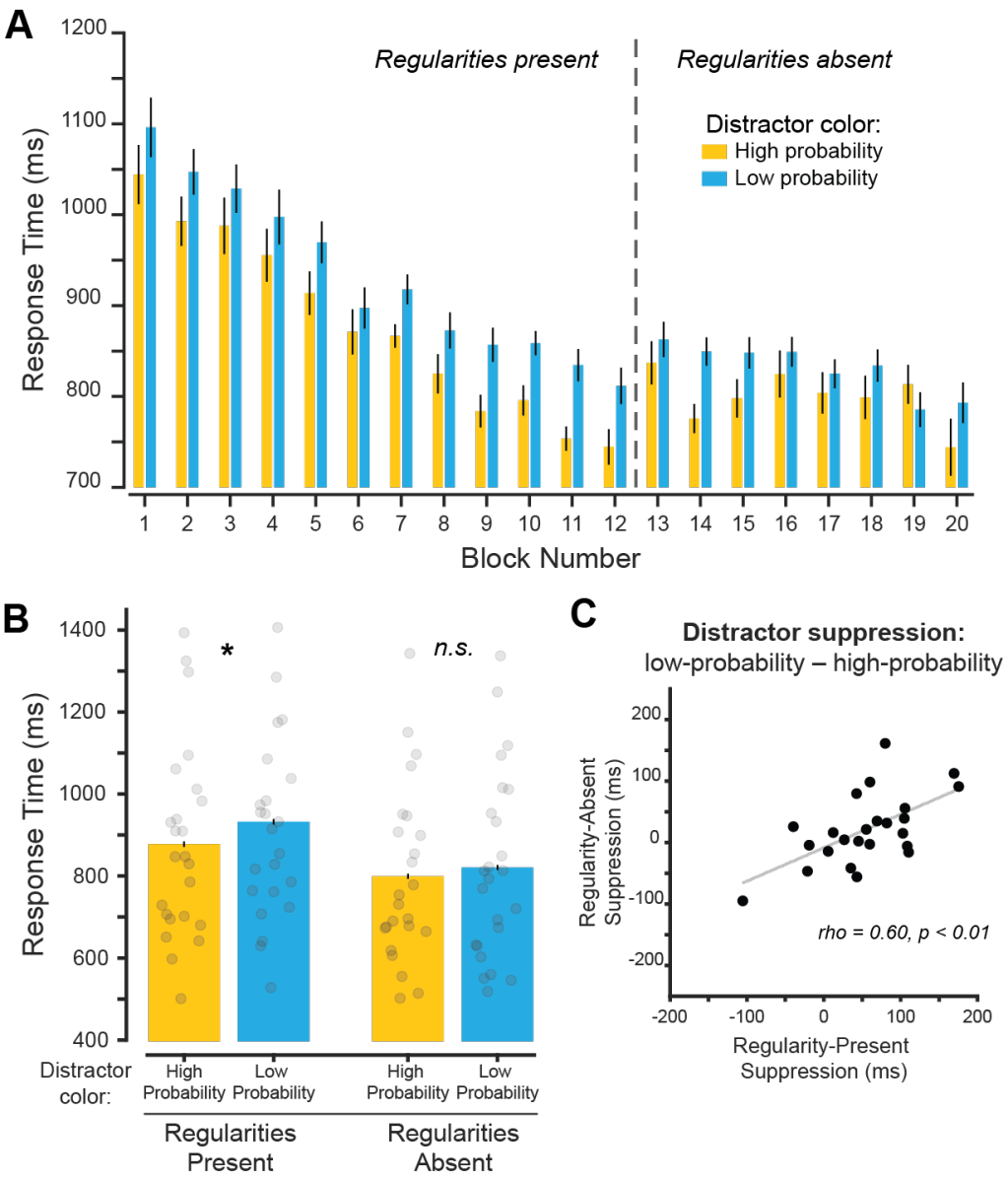

Figure 16. Experiment 2: high-probability distractor color is suppressed during learning and over an extended interval. **(A)** Mean RT for each block on trials with correct orientation reports. Dashed line indicates when distractor color regularities were removed from the display. **(B)** Mean RT across regularity-present and regularity-absent blocks for both high- and low-probability color conditions. Individual subject data points shown. Significant differences between color probability conditions indicated with * for *p* values < 0.05. **(C)** Correlation between suppression effect during regularity-present blocks and regularity-absent blocks. Suppression effects were computed as the difference in RT between the low- and high-probability distractor color conditions. Error bars are within-subject standard error of the mean.

Table 4. Experiment 2 Accuracy (± SEM)

	Regularities Present	Regularities Absent
High-Probability Color	86.94% (0.86)	85.45% (1.23)
Low-Probability Color	82.19% (0.86)	81.66% (1.23)

Regularity-absent performance.

During the regularity-absent blocks, the previously high-probability color could be present in either the target or distractor array, but was presented with the same probability as all other colors. Similar to Experiment 1, we conducted a two-way repeated measures ANOVA with *distractor color* as the first factor (high probability color distractor vs low-probability colors) and phase as the second factor (regularity-present vs regularity-absent; Fig. 3B). There was a main effect of distractor color condition, $F(1,23) = 11.92$, $p = 0.002$, $\eta_p^2 = 0.69$, and phase $F(1,23) = 21.08$, $p < 0.001$, $\eta_p^2 = 0.91$. These findings demonstrate that participants were overall faster to respond when the distractor was shown in the high-probability color and that RTs were faster during regularity-absent blocks. Importantly, there was no interaction between these variables, $F(1,23) = 3.74$, $p = 0.066$, $\eta_p^2 = 0.14$, which leaves open the possibility that suppression of the previously high-probability color continued when regularities were removed. However, follow-up comparison showed that distractor suppression across subjects did not persist into regularity-absent blocks ($t(23) = 1.80$, $p = 0.086$, $d_z = 0.37$, $BF_{01} = 1.17$).

To test if individual subjects continued to suppress the learned distractor color, we computed the correlation between the distractor suppression effect during the regularity-present blocks

and the distractor suppression effect during the regularity-absent blocks (Fig. 16C). There was a positive correlation when comparing the regularity-present distractor suppression and regularity-absent distractor suppression ($r = 0.60$, $p = 0.002$). Consistent with Experiment 1, this result shows that participants who suppressed the high-probability color when regularities were present continue to suppress the color when regularities were removed.

Next, we compared RT across blocks when the *target array* was presented using either the previously learned high- or low-probability distractor color(s) to determine if suppression effects persist when regularities were no longer present (Fig. 17A). A two-way repeated measures ANOVA showed that there was a main effect of condition, $F(1,161) = 5.62$, $p = 0.027$, $\eta_p^2 = 0.16$, as well as a main effect of block $F(7,161) = 2.82$, $p = 0.008$, $\eta_p^2 = 0.29$. There was no interaction, $F(7, 161) = 1.52$, $p = 0.165$, $\eta_p^2 = 0.06$. A paired sample *t*-test showed a significant difference between target color conditions ($t(23) = 2.37$, $p = 0.027$, $d_z = 0.48$, $BF_{10} = 2.16$). Thus, when suppression was measured by presenting the *target array* in the learned high-probability distractor color, we observed persistent suppression after regularities were removed.

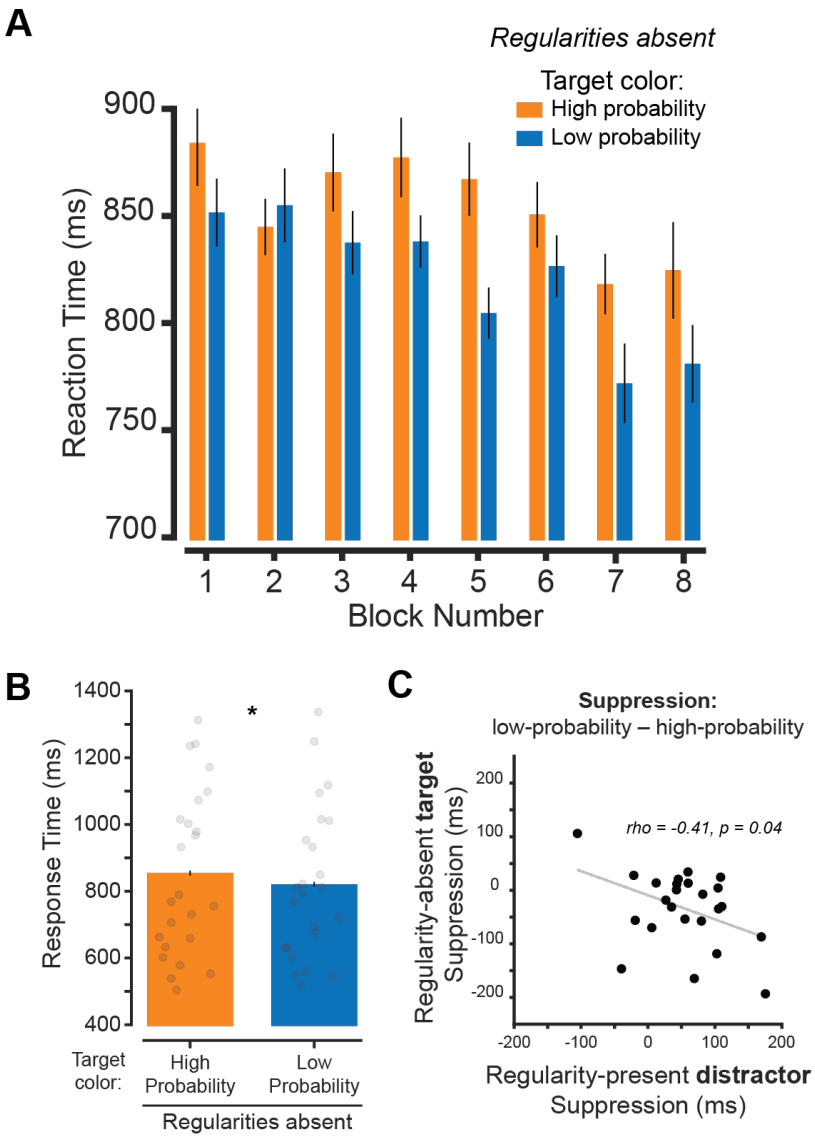

Figure 17. Learned distractor color is suppressed when used as target color after regularities are removed. (A) Mean RT for each regularity-absent block on trials where the target array was presented in the high- and low-probability color(s). **(B)** Mean RT across regularity-absent blocks for both the high- and low-probability color conditions. Individual subject data points shown. Significant differences between color probability conditions indicated with * for p values < 0.05. **(C)** Correlation between suppression effect during regularity-present blocks and regularity-absent blocks. Suppression effect for regularity-present blocks was computed as the difference in RT between the high- and low-probability *distractor* color conditions. Suppression effect for regularity-absent blocks was computed using high- and low-probability *target* color conditions. Error bars are within-subject standard error of the mean.

We then determined whether suppression effects during the regularity-present blocks in individual subjects predicted target suppression in regularity-absent blocks (Fig. 17C). There was a negative correlation between distractor suppression on regularity-present blocks

(measured as the difference in low- vs high-probability distractor color RTs) and target array suppression on regularity-absent blocks (measured as the difference in RT when the *target* was the previously low- vs high-probability distractor color; $r = -0.41$, $p = 0.045$), indicating that participants who suppressed the distractor when regularities were present (resulting in a faster target discrimination response) tended to respond slower to the target array when it was presented in the high-probability color. The negative correlation is expected since continued suppression of previously high-probability color should lead to worse performance when that color was present in the target array, even though the same suppression was helpful during the regularity-present blocks.

Overall, results from Experiment 2 showed that color suppression can occur independent of attenuation of specific spatial locations, replicating our main finding from Experiment 1. Additionally, suppression persisted even when stimulus regularities were no longer present, such that responses were slower when the target array was presented in the suppressed high-probability distractor color than when the target array was presented in a low-probability distractor color.

Inter-trial priming: analysis of aggregate data across experiments.

As a final test of whether the non-spatial color suppression we observed was due to statistical learning (regularities learned throughout the experiment) or to inter-trial priming (transient influence of previous trials), we computed the mean RT of each condition using only 'switch' trials, or trials where the distractor probability was different from the distractor probability on the previous trial. We compared switch trials to 'repeat' trials, where the distractor probability was the same as the distractor probability from the previous trial. This

analysis allowed us to assess the individual contribution of priming, which is expected to result in a stronger effect on repeat than switch trials, and perseverant learning, which should still be present in switch trials. In addition to analyzing data after sorting each trial (n) based on the switch/repeat status of the previous trial (n-1), we also looked at trials farther back in the experiment where the distractor probability matched/mismatched the current trial distractor probability in a serial manner (n-k). We sorted each trial (n) based on the trial label 1-8 trials previous (k = 1:8), because previous research has shown that priming no longer impacts RT after approximately 7 trials (Maljkovic & Nakayama, 1994). Since this removes a large proportion of trials, and because blocks 1 through 12 were identical in both experiments, we collapsed across data from both experiments to ensure adequate power (total n = 48).

Figure 18A shows a significant three-way interaction between distractor probability (high- and low-probability), priming (switch and repeat), and serial position (n-1 through n-8), $F(7,329) = 3.88$, $p < 0.001$, $\eta_p^2 = 0.08$. This demonstrates that distractor suppression was modulated by priming, but that this effect changed as a function of how far back in the trial sequence a repeat occurred. To better visualize the influence of priming at each serial position, we computed a priming distractor suppression value by first finding the difference in RT between the high- and low-probability distractor color conditions independently for switch and repeat trials, then computing the difference between these values. Positive values of this measure indicate greater distractor suppression on repeat trials (Fig. 18B)[1].

[1] Our analysis is unable to completely exclude priming as a contributing factor, as we did not account for repeat trials between or after the analyzed trials (e.g., trials 1 through 7 when analyzing n-8). This is because trials with

Priming indeed had a diminishing effect on RT the farther back a repeat occurred in the trial sequence, with most influence absent after n-5. Importantly, when comparing the high- and low-probability distractor color conditions using only n-1 switch trials (where inter-trial priming had the strongest influence; Fig 18B), RTs were still significantly faster when the high-probability color was shown ($t(47) = 2.25$, $p = 0.029$, $d_z = 0.32$, $BF_{10} = 1.54$; Fig. 18C). This is additional evidence suggesting that feature suppression is due, at least in part, to long-term learning of stimulus regularities.

no repeats between/after trial n were exceedingly rare, making it difficult to interpret results. However, our findings do clearly demonstrate that priming cannot entirely account for learned color suppression.

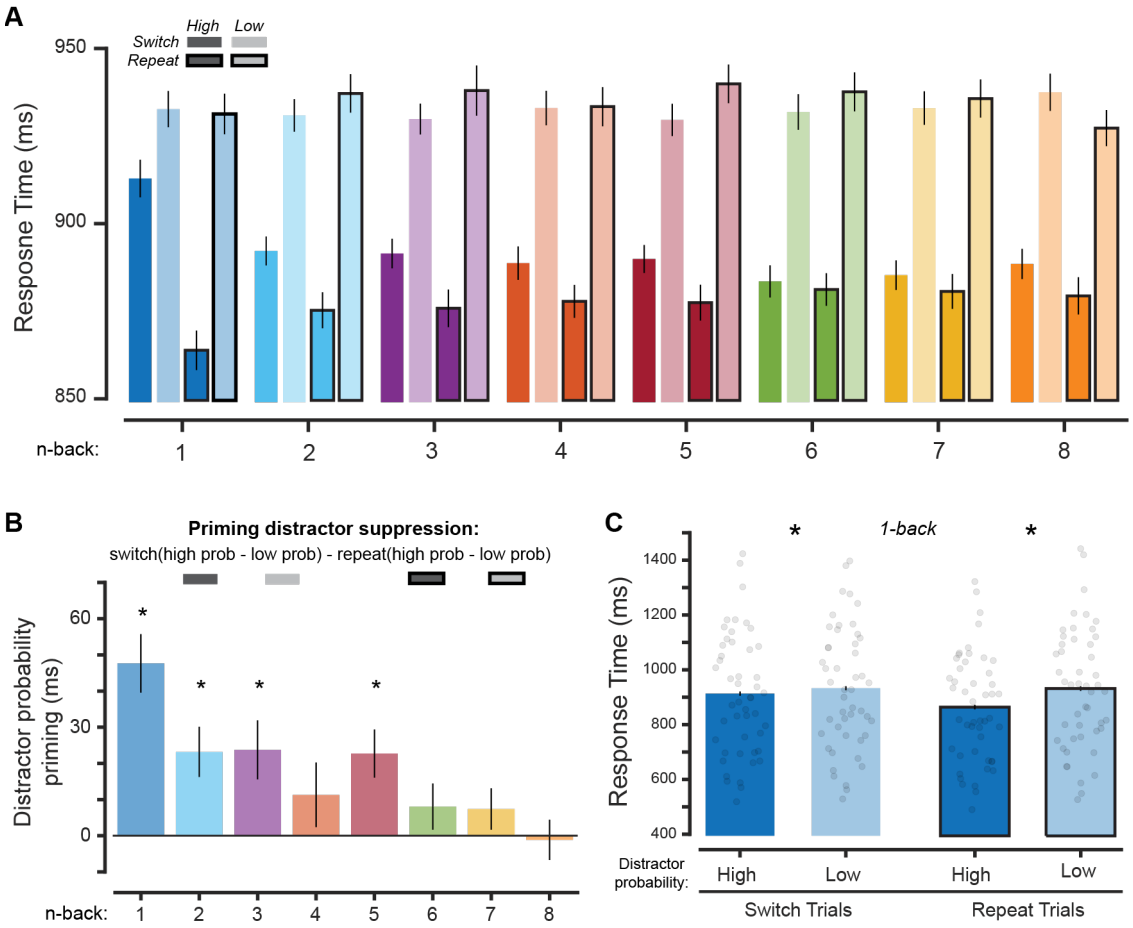

Figure 18. Suppression is not entirely explained by intertrial priming. **(A)** Mean RT was computed for the distractor array high-probability and low-probability conditions on trials where the previous trial used a different, or the same, distractor probability than that presented on the analyzed trial ('switch' and 'repeat' trials respectively). We performed this analysis serially, sorting by switch/repeat based on trials 1-8 prior to the current trial. **(B)** Difference between 'switch' and 'repeat' trial suppression effects. Suppression was computed independently for switch/repeat trials as the difference between high- and low-probability distractor color conditions. Then, the difference between switch and repeat suppression effects were plotted, where positive values indicate greater suppression on repeat trials. * indicates significant difference, $p < 0.05$, one-sample t-test. **(C)** Mean RT for the distractor array high- and low-probability color conditions for 'switch' and 'repeat' trials on n-1 trials. * indicates significant difference, $p < 0.05$, paired t-test. Data from both experiments were used to ensure enough power to detect an effect. Error bars are within-subject standard error of the mean.

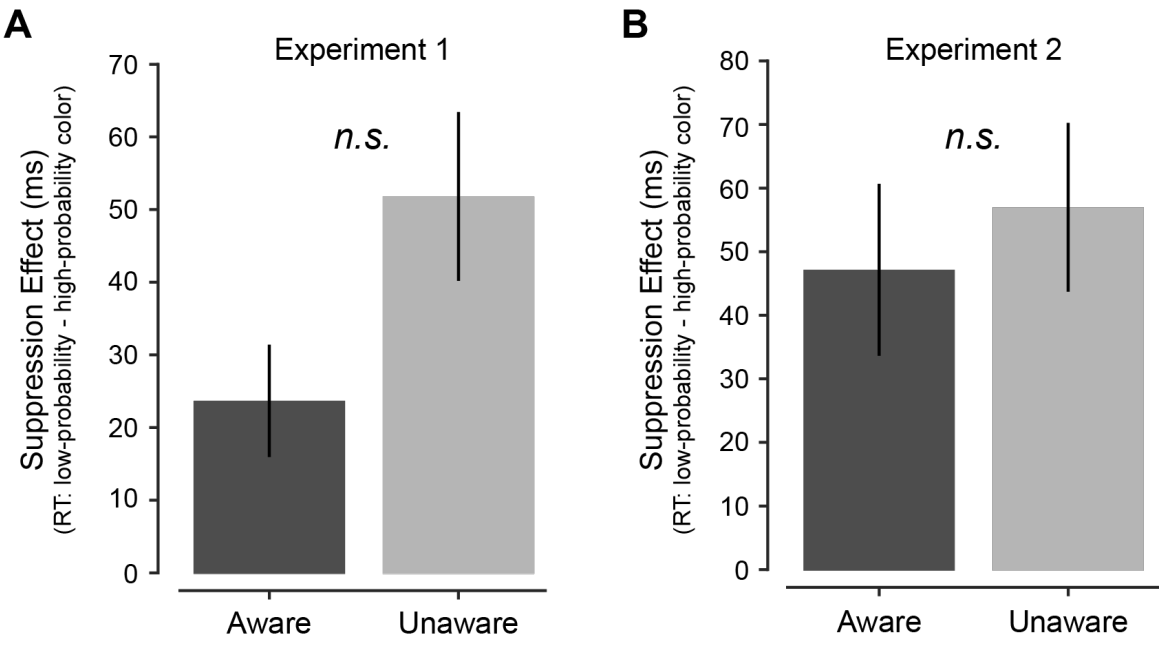

Figure 19: Distractor suppression does not require awareness. Suppression effect (target discrimination RT for low-probability minus high-probability distractor color; regularity-present trials) compared across groups of participants aware or unaware of distractor color regularities. For each Experiment (A: Expt 1; B: Expt 2), we split participants into groups based on whether they correctly reported the regular distractor color ("Aware") or not ("Unaware"). Distractor suppression did not differ between these groups, and qualitatively, distractor suppression was stronger when participants did not correctly detect regularities. Error bars are between-participant SEM.

General Discussion

The current study was designed to understand whether a distracting stimulus defined based on its color could be suppressed independent of spatial location. If true, target discrimination performance should be improved when a high-probability distractor color is present at the same location of a target stimulus as compared to when any low-probability distractor color is presented. We tested this by showing participants two overlapping line arrays, where they had to report the orientation of the array with the most lines (Fig. 14A). During regularity-present blocks, the distractor array was usually presented in one color. Over the course of both experiments, RTs were faster when this high-probability color was present in the

distractor array relative to one of the other low-probability colors, indicating that the distractor color was suppressed (Figs 15B, 16B). Distractor suppression persisted when color regularities were removed from the display for subjects utilizing them during regularity-present blocks, indicating that suppression cannot be fully explained by priming (Figs. 15C, 16C). In Experiment 2, we found stronger evidence in favor of long-term suppression: RTs were slower when the target array was presented in the high-probability color (Fig. 17), and this suppression persists when we only analyzed switch trials during regularity-present blocks (Fig. 18).

Our findings build on the growing literature demonstrating feature suppression through repeated exposure to regularly presented visual search singletons (Failing et al., 2019; Gaspelin & Luck, 2018; Stilwell & Gaspelin, 2021; Stilwell et al., 2019; Vatterott & Vecera, 2012). Importantly, the effects in the present study were identified when spatial suppression mechanisms could not be used to lower the prioritization of distracting items. In all of the aforementioned studies, visual search tasks were employed, which have been useful in identifying when particular display statistics are used to guide search behavior (Stilwell et al., 2019) as well as how regularities may interact within and between feature dimensions to modulate suppression (Failing et al., 2019). However, to further understand how these regularities are deployed, it is important to understand each one in isolation. Our stimulus, in which *only* color regularities could contribute to guiding suppression, could be a useful tool for future studies to isolate feature-specific suppression mechanisms from their spatial counterparts.

Potentially contrasting with findings of feature-specific suppression are studies indicating that only stimulus *locations* can be deprioritized (Theeuwes, 2010; Moher & Egeth, 2012). For example, Moher and Egeth (2012) had participants perform a target detection task where a cue was given at the start of each trial. This cue was informative about the color of distractors in an upcoming multi-item display, where each item occupied a unique location. Target detection was faster when an informative distractor cue was provided as compared to a neutral cue, but this effect was only observed when the distractor location was attended prior to the onset of a target. This result led to their *search-and-destroy* hypothesis, which states that a location needs to be selected first and then a distractor presented in a learned feature can be suppressed. Further evidence suggesting that suppression is location-dependent comes from their Experiment 4, as suppression was not improved when there were several distractors of the same color, inconsistent with accounts of feature-based attention in which a specific feature value can be up/down-regulated across the entire screen simultaneously (Maunsell & Treue, 2006; Treue & Trujillo, 1999a).

There are two noteworthy differences between our study and Moher and Egeth (2012). First, as mentioned previously, the spatially overlapping stimulus used in the current study discouraged the use of any spatial suppression mechanisms, as it would not have benefited target detection. Second, Moher and Egeth (2012) used cues to direct volitional control towards suppressing task-irrelevant information. In our study, subjects were unaware of the display statistics, yet their performance was modulated by the presence of a high-probability color. It appears that top-down control cannot be used to suppress distracting information in a parallel feature-based manner, but implicit mechanisms allow for a more global suppression.

While this may be the case, a potential downside to implicit learning is that suppression persists even when it is no longer useful, as was evident in our Experiment 2 (Fig. 17B/C), whereas top-control can be implemented on a trial-by-trial basis (Cunningham & Egeth, 2016). Overall, our results provide strong evidence that feature-specific suppression obtained through statistical learning can occur *independent* of top-down spatial suppression operations such as reactive control (Theeuwes, 2010) or search-and-destroy processes (Moher & Egeth, 2012).

While we ruled out reactive spatial mechanisms as a possible alternative to learned feature suppression, the current findings are unable to address whether suppression exclusively occurred proactively or reactively. It could be the case that the distractor array is less likely to be selected when presented in the high-probability color (Gaspelin et al., 2019), or the distractor array is still selected but the high-probability color is rapidly suppressed through reactive feature suppression. Without explicit knowledge of the color regularities, both line arrays need to be attended to determine which is the target. Thus, we speculate that reactive mechanisms were deployed when attending our stimuli. However, it is plausible, especially during regularity-present blocks, that a proactive mechanism was also used as participants implicitly learned the high-probability color. Ultimately, both strategies can be implemented (Geng, 2014). For instance, it is often more efficient to proactively ignore distracting stimuli but, since these regularities may not persist—as is the case in our regularity-absent blocks—it can be beneficial to allow for learned distractors to occasionally capture attention to update learned regularities. This is even more effective with a reactive mechanism to quickly disengage from stimuli as long as they are still distracting.

How do our results fit with priority map theory? Within this framework, maps corresponding to individual feature dimensions are integrated into a feature-agnostic priority map (Itti & Koch, 2001, 2001; Wolfe, 1994). Locations with the greatest prioritization are selected for the allocation of attention. Mechanisms for distractor suppression generally fit nicely within this model, as they explain how locations within these maps are deprioritized (Failing et al., 2019; Luck et al., 2020). Whenever feature-specific suppression is engaged, modulations are thought to occur within the corresponding feature map. For example, a regularly presented red singleton will have lower activation in the red feature map, which results in lower activation in the summed priority map. It is difficult to reconcile our results exactly within this structure, as suppressing the distractor location would also deprioritize the target item due to their shared location. Rather than specific locations being the target of prioritization, others have proposed that modulation can occur at the level of individual objects (Shomstein, 2012; Shomstein & Yantis, 2002), even when they are occluded (Moore et al., 1998). According to this account, after directing spatial attention, goal-relevant objects at that location are selected before other less-relevant objects. This mechanism of object-based attention is compatible with the spatially overlapping stimuli used in the current study. When considered within the context of feature maps, in addition to the high-probability color being suppressed, the orientation of the lines associated with the high-probability color would be suppressed allowing for the other object in the display to be selected first. Future work can manipulate the statistics in this paradigm to tease apart when objects, features, and/or locations are suppressed.

Rather than a suppressive reweighting of objects within the priority map framework, it could be the case that distractor statistics are used to shift or enhance the representation of the three possible target colors. Recent evidence shows that when distractors are regularly presented in colors that are linearly separable from the target color in feature space, the representation of a target color shifts away from the color of distractors (Navalpakkam & Itti, 2007; P. Witkowski & Geng, 2019; Yu & Geng, 2019). By shifting the target representation, it makes it harder for distractors with a similar color as the target to capture attention (Duncan & Humphreys, 1989). In the current study, it is possible that the representation of each target color was shifted away from the learned distractor color to improve performance. However, there are a couple of aspects of our design that are difficult to reconcile with this account. First, the high-probability color was only the most likely distractor color—the other three colors were the distractor on some trials. Yu & Geng (2019) showed that when distractor colors were sampled from either side of feature space, the target color representation no longer shifted. Second, these previous studies have primarily investigated how distractor statistics influence target representations, but there has yet to be a study showing how implicitly learned distractor information is modulated in similar visual search paradigms, so it is unclear whether both targets and distractors are influenced. Regardless, this remains an interesting mechanism and further studies should investigate the degree of influence distractor statistics have on both targets and distractors.

In both experiments, there was evidence that learned distractor suppression may be a variable characteristic across individuals since suppression effects persisted in regularity-absent blocks for participants who showed an effect during regularity-present blocks (Figs. 15C,

16C, 17C). This may not come as a surprise, as similar findings are apparent in the working memory literature (Luria et al., 2016). As an example, individuals who perform well on memory tasks tend to be better at ignoring distracting information (Vogel et al., 2005). While speculative, it is possible that the ability to leverage distractor statistics to prioritize target information is related to the ability to prevent distracting information from entering visual working memory. In fact, the distractors used in the primary experiment of Vogel et al. (2005) were always red—a feature regularity that could be used in a manner consistent with learned suppression. This is further supported by the strong relationship between visual working memory and attention (Awh & Jonides, 2001; Bahle et al., 2018; Olivers et al., 2006). However, additional studies are needed to directly test whether distractor suppression observed in studies of selection history is related to the ability to prevent irrelevant information from entering working memory.

Conclusion

It is imperative to suppress distracting information for effective selection of relevant stimuli in service of goal-oriented behavior. Mounting evidence has shown that locations corresponding to a distractor can be suppressed (Gaspelin et al., 2015; Stilwell et al., 2019; Wang & Theeuwes, 2018), but it is important to understand whether non-spatial features can be inhibited when space-based suppression is not beneficial. Our study showed that when overlapping stimuli are presented, a high-probability distracting color is suppressed to improve target discrimination performance. This suppression persisted even when regularities were removed from the display, indicating that learned statistics contributed to

this effect. Overall, we provide strong evidence that features can be suppressed independent of spatial location.

Chapter V: Prioritizing Feature-Specific Information

The studies presented here provide evidence required to support the role of feature dimension maps, both behaviorally and neurally, within the priority map framework (Bisley & Goldberg, 2010; Fecteau & Munoz, 2006; Itti & Koch, 2001; Katsuki & Constantinidis, 2014; Serences & Yantis, 2006; Treisman & Gelade, 1980; Wolfe, 1994). Chapters 2 and 3 identified the neural correlates of feature dimension maps, with hV4/VO1/VO2 representing a neural color dimension map and TO1/TO2 representing a neural motion dimension map. For a region to be considered a feature dimension map, they need to be spatially selective, feature selective, and index important feature-specific locations. This is supported by the findings that each region: 1) tracked the location of stimuli throughout the visual field, demonstrating spatial organization; 2) responded more strongly to stimuli that were presented in the preferred feature of a region or when that feature was goal relevant, demonstrating feature selectivity; and 3) responded strongly to the location of both salient and relevant stimuli, demonstrating that each region represents important stimuli.

As mentioned in Chapter 1, a third factor can prioritize information in the visual field—selection history. Selection history is how previous encounters with a stimulus influence what is prioritized in the future. Two chapters support the idea that selection history is used to compute priority. In Chapter 3, motion-selective regions prioritized stimuli based on stimulus history, as the stimulus representation was modulated when participants performed the attend motion task at least twice in a row. However, some claim that modulations of priority due to selection history can only occur spatially (Theeuwes, 2010; Theeuwes et al., 2003). Since the results in Chapter 3 were based on a spatial encoding model, it may be that

modulations emerged through a spatial mechanism. In Chapter 4, a behavioral study was conducted to assess if certain features, specifically color, could be suppressed independent of spatial location. Whenever a learned color distractor appeared in the display, task performance was improved, which is evidence that selection history prioritized specific features and is not restricted to spatial positions. In this chapter I will discuss how this collection of work impacts models of attention as well as briefly discuss future research directions that are now possible because of these findings.

Implications for models of attention

Identifying hV4/VO1/VO2 and TO1/TO2 as neural feature dimension maps solidifies the functional role of these regions for prioritizing important information. In most priority map models, feature dimension maps are used to compute feature-specific information before summing into the feature-agnostic priority map. Then, attention is directed to locations with the highest levels of activation in the priority map. But what happens when only specific feature information is necessary? As described by original priority map models (Koch & Ullman, 1985), if information from a given feature dimension is not relevant or salient, there would simply be low activation within the corresponding dimension map. Once maps are combined to form the priority map, there would be minimal influence from the irrelevant feature dimension. However, since the brain is a noisy system, then there would still be low levels of input from the neural feature dimension map driving a weak response in the neural priority map. Noise introduced in this way would not result in behavioral detriments most of the time, but there are situations where high performance is required (Theeuwes, 2023a).

Thus, it may be beneficial to actively suppress neural feature dimension maps when that information is not relevant, especially when it would be detrimental to behavior.

One way this could occur is through dimensional weighting (Found & Müller, 1996; Liesefeld & Müller, 2019; Müller et al., 1995). The dimensional weighting account (DWA) is a model grounded in priority maps, stating that observer goals can increase or decrease activity across an entire feature dimension. When a certain feature dimension is relevant (e.g., color), then all feature values in that dimension (e.g., all colors, such as red and blue) are upweighted. This contrasts with early priority map models, as something like Guided Search (Wolfe, 1994) would state that only specific goal-relevant feature values are upweighted (e.g., red rather than all colors). The neural results presented here are consistent with dimensional weighting, as responses were modulated across each feature-selective region. This may be a result of an entire feature dimension being enhanced, but since a spatial encoding model was used to infer feature-specific modulations, a plausible alternative is that neural populations tuned to specific feature values were driving this effect. To answer this question, color or motion inverted encoding models could evaluate neural responses to specific features (Brouwer & Heeger, 2009; Saproo & Serences, 2014).

An additional claim of DWA is that task-irrelevant feature dimensions can be suppressed (Liesefeld et al., 2019; Sauter et al., 2018). For example, when a distracting feature dimension is known in advance (e.g., motion), then all information within the corresponding feature dimension map will be suppressed (e.g., motion map suppressed). This operation prevents specific feature dimensions from integrating into the priority map, which would be a more effective way to prioritize important stimuli as it could fully remove distracting inputs.

Support for this comes from work on interregion suppression via inhibitory gating (Bonnefond et al., 2017; Jensen & Mazaheri, 2010). Areas that are associated with attentional control, such as posterior parietal cortex (Corbetta & Shulman, 2002), can suppress information received from various sensory areas in visual cortex through decreased coupled oscillatory neural activity within frequencies of 8-12 Hz (Jokisch & Jensen, 2007; Mazaheri et al., 2014). Since these areas represent candidate priority and feature dimension maps, then dimensional weighting may occur by suppressing connections between these areas.

While the work presented here focuses on how neural feature dimension maps are recruited for attention, it could be the case that these regions are needed for other cognitive functions, such as working memory (WM; Jerde et al., 2012; Jerde & Curtis, 2013; Zelinsky & Bisley, 2015). There is a large degree of overlap between the functional role of attention and WM (Awh & Jonides, 2001; Kiyonaga & Egner, 2013; Larocque et al., 2014; Oberauer, 2019; Olivers et al., 2011), with one key similarity being that both select important information—selection occurs externally in the case of attention and internally in the case of WM (Chun, 2011). Due to these similarities, processing related to one cognitive function is often used to infer properties of the other (Hollingworth & Luck, 2009; Olivers, 2008; Soto et al., 2005, 2008; Thayer et al., 2021; Thayer & Sprague, 2023). Furthermore, the same neural regions that are modulated by attention in this manuscript are recruited for WM maintenance (Harrison & Tong, 2009; Postle, 2006; Serences et al., 2009; Sreenivasan & D'Esposito, 2019). It seems plausible that neural feature dimension maps are recruited to maintain information in memory when there is no sensory input. There is certainly preliminary evidence indicating that this is the case, as feature-specific memories are decodable from

visual cortex (Ester et al., 2009; Leavitt et al., 2017; Roussy et al., 2021; Sahan et al., 2020). However, to provide a concrete connection to the priority map architecture, it is necessary to demonstrate that neural feature dimension maps are specifically recruited to maintain information in WM.

The primary goal of priority map theory is to understand how we select items of interest in the environment. Identifying how selection occurs in the brain, and which regions are responsible for this operation, will improve our understanding of information processing. However, part of what makes this a difficult endeavor is that there are several regions considered priority maps. One explanation is that all these regions are redundant to ensure a robust neural code for directing attention (Zelinsky & Bisley, 2015). Alternatively, each region could serve unique functions depending on task demands or required motor output, such as FEF being recruited for eye movements (Moore & Armstrong, 2003) and parietal areas recruited for other motor movements (Murata et al., 2000). Since feature maps sum into the priority map, then one way to identify whether a region has a unique role as a priority map is to evaluate when signals are shared with feature dimension maps under various task demands through functional connectivity (Bassett & Gazzaniga, 2011; Gonzalez-Castillo & Bandettini, 2018; van den Heuvel & Hulshoff Pol, 2010). It may be that parietal regions have stronger connectivity with neural dimension maps when performing a grasping task, while connectivity with FEF is stronger when making an eye movement. This approach could also help rule out a region as a priority map. Candidate priority maps that have no functional relationship with neural feature dimension maps may in fact not be priority maps, but could instead be inheriting information from other regions (Mirpour & Bisley, 2012). Thus, neural

feature dimension maps provide a way to clearly categorize the role of other regions in directing attention.

Perhaps neural feature dimension maps themselves are used to direct attention more than originally theorized. Since specific regions may be recruited to direct attention depending on motor output, it could be that neural feature dimension maps are recruited when a task can be completed with only feature-specific information. For example, looking for a banana slug in the forest can be accomplished by searching for the color yellow. In this circumstance, the color dimension map contains the majority of information needed to find the banana slug. If responses are based directly on color map activation, then this task can be efficiently completed. An additional benefit is that this approach would have relatively minimal metabolic cost, as fewer regions need to be involved with the selection process.

Future tests of feature dimension maps

While the work presented here provides the necessary evidence to establish hV4/VO1/VO2 and TO1/TO2 as neural feature dimension maps, there are still other properties of feature dimension maps that need to be tested in the neural correlates. For instance, how do these areas resolve conflict between multiple important stimuli? Since feature dimension maps are supposed to index all prioritized locations in the visual field, when there are competing prioritized locations, then they should all be represented within the appropriate feature dimension map. Further still, the location with the greatest activation must have the highest chance of being selected by attention if these regions are involved in computing priority. Future studies should include multiple stimuli to see if they are simultaneously represented and if the relative activation in these regions reflects what is selected by attention.

A related point is to understand how these regions represent varying levels of stimulus salience. Many studies have shown that gradual increases in stimulus luminance lead to enhanced neural and behavioral responses (Herrmann et al., 2010; Reynolds & Heeger, 2009; Sprague, Itthipuripat, et al., 2018). There is behavioral work evaluating detection thresholds using different levels of feature salience (Nothdurft, 1993a, 1993b, 2000), but there remains minimal work evaluating how graded feature salience emerges neurally (but see Poltoratski et al., 2017). Feature dimension maps should reflect changes in feature contrast. Items that are highly salient should be more strongly represented than items with low levels of salience, as indicated by low local feature contrast. Conducting a study that manipulates the degree to which a feature-salient stimulus varies from the background would be an excellent way to finely assess how these regions compute image salience.

On the other end, these maps provide a unique opportunity to further appraise the computations that occur for prioritizing goal-relevant information. For example, when a specific feature value is task-relevant, there is evidence that modulations can occur in a spatially global manner (Maunsell & Treue, 2006). Feature-based attention (FBA) has been observed in monkey (Martínez-Trujillo & Treue, 2002; Treue & Trujillo, 1999a) and human (O'Craven et al., 1999; Saenz et al., 2002) feature-selective cortices that correspond to the neural color- and motion-dimension maps. Studies investigating FBA have observed enhanced responses at locations containing either a task-irrelevant, or even no stimulus (Serences & Boynton, 2007), when attending to a specific feature value at a different location. Alternatively, some studies have shown that feature-specific modulations only occur at the stimulus location (McMains et al., 2007). The results from Chapter 3 are more in

line with spatially localized feature modulation, as there was a minimal response at stimulus-absent locations. Regardless, there is a large body of work demonstrating that modulations *can* be spatially global. It may be that both accounts are correct under specific circumstances. One way to investigate this is to introduce a second task-irrelevant stimulus when participants perform a task similar to what was used in Chapter 3. That way, the strength of the task-irrelevant stimulus can be assessed as a function of what feature dimension is attended in the relevant stimulus.

Lastly, there is only initial evidence that these regions are modulated by selection history. It was necessary to first show that specific features can be prioritized through repeated encounters, as some models claim only the spatial location of a distracting stimulus can be suppressed through selection history (Theeuwes, 2010; Theeuwes et al., 2003). Contrary to this, Chapter 4 provides concrete evidence that the actual features themselves can be deprioritized. This opens future work to identify how neural feature dimension maps index this computation of priority. By adjusting the frequency by which stimuli are shown in paradigms similar to the ones used for Chapters 2 and 3, the influence of selection history can be evaluated. Additionally, implementing this change can be useful for assessing how selection history interacts with goal relevance and image salience to provide a more comprehensive understanding of the role of feature-selective regions in computing priority.

Conclusions

Across three studies, evidence was presented that regions of visual cortex represent neural feature dimension maps. Specifically, hV4/VO1/VO2 and TO1/TO2 represent neural color and motion dimension maps. This claim is supported by each region tracking the location of

stimuli throughout the visual field, responding more strongly to stimuli that matched the preferred feature of each region, and indexing the location of feature-specific important stimuli. Additionally, results were consistent with selection history acting to suppress specific features through learned statistical relationships. These findings support critical aspects of priority map theory, being the first to validate core structural elements of the model and to definitively demonstrating that learned feature distractors can be suppressed. This body of work explains aspects of feature-selective attention and facilitates a rich body of future research.

References

Adam, K. C. S., & Serences, J. T. (2020). *History-driven modulations of population codes in early visual cortex during visual search* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.09.30.321729

Adam, K. C. S., & Serences, J. T. (2021). History Modulates Early Sensory Processing of Salient Distractors. *Journal of Neuroscience*, *41*(38), 8007–8022. https://doi.org/10.1523/JNEUROSCI.3099-20.2021

Addleman, D. A., & Störmer, V. S. (2023). Distractor ignoring is as effective as target enhancement when incidentally learned but not when explicitly cued. *Attention, Perception, & Psychophysics*, *85*(3), 834–844. https://doi.org/10.3758/s13414-022-02588-y

Albright, T. D. (1993). Cortical processing of visual motion. *Reviews of Oculomotor Research*, *5*, 177–201.

Amano, K., Wandell, B. A., & Dumoulin, S. O. (2009). Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *Journal of Neurophysiology*, *102*(5), 2704–2718.

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: Configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(3), 10.1037/a0027885. https://doi.org/10.1037/a0027885

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*(5), 485–496. https://doi.org/10.3758/BF03205306

Bahle, B., Beck, V. M., & Hollingworth, A. (2018). The architecture of interaction between visual working memory and visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(7), 992.

Bahle, B., Thayer, D. D., Mordkoff, J. T., & Hollingworth, A. (2019). The architecture of working memory: Features from multiple remembered objects produce parallel, coactive guidance of attention in visual search. *Journal of Experimental Psychology: General*.

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*, 295–314. https://doi.org/10.1037/met0000337

Bartels, A., & Zeki, S. (2000). The architecture of the colour centre in the human visual brain: New results and a review. *European Journal of Neuroscience*, *12*(1), 172–193.

Bassett, D. S., & Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in Cognitive Sciences*, *15*(5), 200–209. https://doi.org/10.1016/j.tics.2011.03.006

Basso, M. A., & Wurtz, R. H. (1998). Modulation of Neuronal Activity in Superior Colliculus by Changes in Target Probability. *Journal of Neuroscience*, *18*(18), 7519–7534. https://doi.org/10.1523/JNEUROSCI.18-18-07519.1998

Basso, M. A., & Wurtz, R. H. (2002). Neuronal Activity in Substantia Nigra Pars Reticulata during Target Selection. *Journal of Neuroscience*, *22*(5), 1883–1894. https://doi.org/10.1523/JNEUROSCI.22-05-01883.2002

Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*(8), Article 8. https://doi.org/10.1038/nn1501

Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The role of relational information in contingent capture. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1460–1476. https://doi.org/10.1037/a0020370

Benda, J. (2021). Neural adaptation. *Current Biology*, *31*(3), R110–R116. https://doi.org/10.1016/j.cub.2020.11.054

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. https://doi.org/10.1214/aos/1013699998

Bichot, N. P., & Schall, J. D. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience*, *2*(6), Article 6. https://doi.org/10.1038/9205

Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, *299*(5603), 81–86.

Bisley, J. W., & Goldberg, M. E. (2006). Neural Correlates of Attention and Distractibility in the Lateral Intraparietal Area. *Journal of Neurophysiology*, *95*(3), 1696–1717. https://doi.org/10.1152/jn.00848.2005

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

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, *203*(1), 237–260. https://doi.org/10.1113/jphysiol.1969.sp008862

Bogler, C., Bode, S., & Haynes, J.-D. (2011). Decoding Successive Computational Stages of Saliency Processing. *Current Biology*, *21*(19), 1667–1671. https://doi.org/10.1016/j.cub.2011.08.039

Bogler, C., Bode, S., & Haynes, J.-D. (2013). Orientation pop-out processing in human visual cortex. *NeuroImage*, *81*, 73–80. https://doi.org/10.1016/j.neuroimage.2013.05.040

Bonnefond, M., Kastner, S., & Jensen, O. (2017). Communication between Brain Areas Based on Nested Oscillations. *eNeuro*, *4*(2), ENEURO.0153-16.2017. https://doi.org/10.1523/ENEURO.0153-16.2017

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, *8*(8), 1102–1109.

Britton, M. K., & Anderson, B. A. (2020). Specificity and Persistence of Statistical Learning in Distractor Suppression. *Journal of Experimental Psychology. Human Perception and Performance*, *46*(3), 324–334. https://doi.org/10.1037/xhp0000718

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience*, *29*(44), 13992–14003.

Bruce, C. J., & Goldberg, M. E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *Journal of Neurophysiology*, *53*(3), 603–635. https://doi.org/10.1152/jn.1985.53.3.603

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523.

Burrows, B. E., & Moore, T. (2009). Influence and Limitations of Popout in the Selection of Salient Visual Stimuli by Area V4 Neurons. *Journal of Neuroscience*, *29*(48), 15169–15177. https://doi.org/10.1523/JNEUROSCI.3710-09.2009

Carlisle, N. B. (2023). Negative and positive templates: Two forms of cued attentional control. *Attention, Perception, & Psychophysics*, *85*(3), 585–595. https://doi.org/10.3758/s13414-022-02590-4

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

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*(2), 225–271. https://doi.org/10.1016/0010-0285(90)90017-X

Chang, S., Niebur, E., & Egeth, H. E. (2021). Standing out in a small crowd: The role of display size in attracting attention. *Visual Cognition*, *29*(9), 587–591. https://doi.org/10.1080/13506285.2021.1918810

Chen, X., Zirnsak, M., Vega, G. M., Govil, E., Lomber, S. G., & Moore, T. (2020). Parietal Cortex Regulates Visual Salience and Salience-Driven Behavior. *Neuron*.

Chopin, A., & Mamassian, P. (2012). Predictive Properties of Visual Adaptation. *Current Biology*, *22*(7), 622–626. https://doi.org/10.1016/j.cub.2012.02.021

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*(6), 1407–1409. https://doi.org/10.1016/j.neuropsychologia.2011.01.029

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 224–234. https://doi.org/10.1037/0278-7393.29.2.224

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized Color Modules in Macaque Extrastriate Cortex. *Neuron*, *56*(3), 560–573. https://doi.org/10.1016/j.neuron.2007.10.008

Cook, E. P., & Maunsell, J. H. R. (2002). Attentional Modulation of Behavioral Performance and Neuronal Responses in Middle Temporal and Ventral Intraparietal Areas of Macaque Monkey. *Journal of Neuroscience*, *22*(5), 1994–2004. https://doi.org/10.1523/JNEUROSCI.22-05-01994.2002

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Cosman, J. D., Lowe, K. A., Zinke, W., Woodman, G. F., & Schall, J. D. (2018). Prefrontal Control of Visual Distraction. *Current Biology*, *28*(3), 414-420.e3. https://doi.org/10.1016/j.cub.2017.12.023

Cunningham, C. A., & Egeth, H. E. (2016). Taming the White Bear: Initial Costs and Eventual Benefits of Distractor Inhibition. *Psychological Science*, *27*(4), 476–485. https://doi.org/10.1177/0956797615626564

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

D'Zmura, M. (1991). Color in visual search. *Vision Research*, *31*(6), 951–966. https://doi.org/10.1016/0042-6989(91)90203-H

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5), 14–14. https://doi.org/10.1167/11.5.14

Eckstein, M. P. (2017). Probabilistic Computations for Attention, Eye Movements, and Search. *Annual Review of Vision Science*, *3*(1), 319–342. https://doi.org/10.1146/annurev-vision-102016-061220

Engel, S. A., & Furmanski, C. S. (2001). Selective Adaptation to Color Contrast in Human Primary Visual Cortex. *Journal of Neuroscience*, *21*(11), 3949–3954. https://doi.org/10.1523/JNEUROSCI.21-11-03949.2001

Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially Global Representations in Human Primary Visual Cortex during Working Memory Maintenance. *Journal of Neuroscience*, *29*(48), 15258–15265. https://doi.org/10.1523/JNEUROSCI.4388-09.2009

Failing, M., Feldmann-Wüstefeld, T., Wang, B., Olivers, C., & Theeuwes, J. (2019). Statistical regularities induce spatial as well as feature-specific suppression. *Journal of Experimental Psychology Human Perception and Performance*, *45*(10), 1291–1303. https://doi.org/10.1037/xhp0000660

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Folk, C. L., & Anderson, B. A. (2010). Target-uncertainty effects in attentional capture: Color-singleton set or multiple attentional control settings? *Psychonomic Bulletin & Review*, *17*(3), 421–426. https://doi.org/10.3758/PBR.17.3.421

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimension-weighting" account. *Perception & Psychophysics*, *58*(1), 88–101. https://doi.org/10.3758/BF03205479

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *96*(6), 3314–3319. https://doi.org/10.1073/pnas.96.6.3314

Gao, Y., & Theeuwes, J. (2022). Learning to suppress a location does not depend on knowing which location. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-021-02404-z

Gaspar, J. M., & McDonald, J. J. (2014). Suppression of Salient Objects Prevents Distraction in Visual Search. *Journal of Neuroscience*, *34*(16), 5658–5666. https://doi.org/10.1523/JNEUROSCI.4161-13.2014

Gaspelin, N., Gaspar, J. M., & Luck, S. J. (2019). Oculomotor inhibition of salient distractors: Voluntary inhibition cannot override selection history. *Visual Cognition*, *27*(3–4), 227–246. https://doi.org/10.1080/13506285.2019.1600090

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct Evidence for Active Suppression of Salient-but-Irrelevant Sensory Inputs. *Psychological Science*, *26*(11), 1740–1750. https://doi.org/10.1177/0956797615597913

Gaspelin, N., & Luck, S. J. (2018). Combined Electrophysiological and Behavioral Evidence for the Suppression of Salient Distractors. *Journal of Cognitive Neuroscience*, *30*(9), 1265–1280. https://doi.org/10.1162/jocn_a_01279

Gaspelin, N., & Luck, S. J. (2021). Progress and remaining issues: A response to the commentaries on Luck et al. (2021). *Visual Cognition*, *0*(0), 1–7. https://doi.org/10.1080/13506285.2021.1979705

Geng, J. J. (2014). Attentional mechanisms of distractor suppression. *Current Directions in Psychological Science*, *23*(2), 147–153.

Giesbrecht, B., Woldorff, M., Song, A., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage*, *19*(3), 496–512.

Gonzalez-Castillo, J., & Bandettini, P. A. (2018). Task-based dynamic functional connectivity: Recent findings and open questions. *NeuroImage*, *180*, 526–533. https://doi.org/10.1016/j.neuroimage.2017.08.006

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*(6666), Article 6666. https://doi.org/10.1038/35135

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*(1), 293–321. https://doi.org/10.1016/S0001-6918(01)00019-1

Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K., & Curtis, C. E. (2021). Working memory representations in visual cortex mediate distraction effects. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-24973-1

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

Hayden, B. Y., & Gallant, J. L. (2005). Time Course of Attention Reveals Different Mechanisms for Spatial and Feature-Based Attention in Area V4. *Neuron*, *47*(5), 637–643. https://doi.org/10.1016/j.neuron.2005.07.020

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), Article 12. https://doi.org/10.1038/nn.2669

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward Changes Salience in Human Vision via the Anterior Cingulate. *Journal of Neuroscience*, *30*(33), 11096–11103. https://doi.org/10.1523/JNEUROSCI.1026-10.2010

Hollingworth, A., & Luck, S. J. (2009). The role of visual working memory (VWM) in the control of gaze during visual search. *Attention, Perception, & Psychophysics*, *71*(4), 936–949. https://doi.org/10.3758/app.71.4.936

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*, 599–631. https://doi.org/10.1037/0033-295X.114.3.599

Hughes, A. E., Greenwood, J. A., Finlayson, N. J., & Schwarzkopf, D. S. (2019). Population receptive field estimates for motion-defined stimuli. *NeuroImage*, *199*, 245–260. https://doi.org/10.1016/j.neuroimage.2019.05.068

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *Journal of Neuroscience*, *22*(16), 7195–7205.

Itthipuripat, S., Sprague, T. C., & Serences, J. T. (2019). Functional MRI and EEG index complementary attentional modulations. *Journal of Neuroscience*, *39*(31), 6162–6179.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194.

Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, *4*. https://www.frontiersin.org/articles/10.3389/fnhum.2010.00186

Jerde, T. A., & Curtis, C. E. (2013). Maps of space in human frontoparietal cortex. *Journal of Physiology-Paris*, *107*(6), 510–516.

Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H., & Curtis, C. E. (2012a). Prioritized Maps of Space in Human Frontoparietal Cortex. *The Journal of Neuroscience*, *32*(48), 17382–17390. https://doi.org/10.1523/JNEUROSCI.3810-12.2012

Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H., & Curtis, C. E. (2012b). Prioritized Maps of Space in Human Frontoparietal Cortex. *Journal of Neuroscience*, *32*(48), 17382–17390. https://doi.org/10.1523/JNEUROSCI.3810-12.2012

Jiang, Y. V. (2018). Habitual versus goal-driven attention. *Cortex*, *102*, 107–120. https://doi.org/10.1016/j.cortex.2017.06.018

Jiang, Y. V., Swallow, K. M., Rosenbaum, G. M., & Herzig, C. (2013). Rapid acquisition but slow extinction of an attentional bias in space. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 87–99. https://doi.org/10.1037/a0027611

Jokisch, D., & Jensen, O. (2007). Modulation of Gamma and Alpha Activity during a Working Memory Task Engaging the Dorsal or Ventral Stream. *Journal of Neuroscience*, *27*(12), 3244–3251. https://doi.org/10.1523/JNEUROSCI.5399-06.2007

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, *43*(4), 346–354. https://doi.org/10.3758/BF03208805

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Kastner, S., Schneider, K. A., & Wunderlich, K. (2006). Beyond a relay nucleus: Neuroimaging views on the human LGN. *Progress in Brain Research*, *155*, 125–143.

Katsuki, F., & Constantinidis, C. (2014). Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems. *The Neuroscientist*, *20*(5), 509–521. https://doi.org/10.1177/1073858413514136

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology*, *110*(2), 481–494.

Kay, K. N., Winawer, J., Rokem, A., Mezer, A., & Wandell, B. A. (2013). A Two-Stage Cascade Model of BOLD Responses in Human Visual Cortex. *PLOS Computational Biology*, *9*(5), e1003079. https://doi.org/10.1371/journal.pcbi.1003079

Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin & Review*, *20*(2), 228–242. https://doi.org/10.3758/s13423-012-0359-y

Klink, P. C., Teeuwen, R. R. M., Lorteije, J. A. M., & Roelfsema, P. R. (2023). Inversion of pop-out for a distracting feature dimension in monkey visual cortex. *Proceedings of the National Academy of Sciences*, *120*(9), e2210839120. https://doi.org/10.1073/pnas.2210839120

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227.

Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, *38*(4), 649–662. https://doi.org/10.1016/j.neuroimage.2007.02.022

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00863

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*. https://doi.org/10.3389/fnhum.2014.00005

Larsson, J., & Smith, A. T. (2012). fMRI Repetition Suppression: Neuronal Adaptation or Stimulus Expectation? *Cerebral Cortex*, *22*(3), 567–576. https://doi.org/10.1093/cercor/bhr119

Leavitt, M. L., Mendoza-Halliday, D., & Martinez-Trujillo, J. C. (2017). Sustained Activity Encoding Working Memories: Not Fully Distributed. *Trends in Neurosciences*, *40*(6), 328–346. https://doi.org/10.1016/j.tins.2017.04.004

Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override attentional capture. *Psychonomic Bulletin & Review*, *13*(1), 132–138. https://doi.org/10.3758/BF03193824

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16. https://doi.org/10.1016/S1364-6613(00)01817-9

Liesefeld, H. R., Liesefeld, A. M., & Müller, H. J. (2019). Distractor-interference reduction is dimensionally constrained. *Visual Cognition*, *27*, 247–259. https://doi.org/10.1080/13506285.2018.1561568

Liesefeld, H. R., Liesefeld, A. M., Töllner, T., & Müller, H. J. (2017). Attentional capture in visual search: Capture and post-capture dynamics revealed by EEG. *NeuroImage*, *156*, 166–173. https://doi.org/10.1016/j.neuroimage.2017.05.016

Liesefeld, H. R., & Müller, H. J. (2019). Distractor handling via dimension weighting. *Current Opinion in Psychology*, *29*, 160–167. https://doi.org/10.1016/j.copsyc.2019.03.003

Lin, R., Meng, X., Chen, F., Li, X., Jensen, O., Theeuwes, J., & Wang, B. (2024). Neural evidence for attentional capture by salient distractors. *Nature Human Behaviour*, 1–13. https://doi.org/10.1038/s41562-024-01852-5

Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical Mechanisms of Feature-based Attentional Control. *Cerebral Cortex*, *13*(12), 1334–1343. https://doi.org/10.1093/cercor/bhg080

Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W., & Theeuwes, J. (2020). Progress toward resolving the attentional capture debate. *Visual Cognition*. https://www.tandfonline.com/doi/full/10.1080/13506285.2020.1848949

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews*, *62*, 100–108.

Mackey, W. E., Winawer, J., & Curtis, C. E. (2017). Visual field map clusters in human frontoparietal cortex. *Elife*, *6*, e22974.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672. https://doi.org/10.3758/BF03209251

Martínez-Trujillo, J. C., & Treue, S. (2002). Attentional Modulation Strength in Cortical Area MT Depends on Stimulus Contrast. *Neuron*, *35*(2), 365–370. https://doi.org/10.1016/S0896-6273(02)00778-X

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322.

Mazaheri, A., van Schouwenburg, M. R., Dimitrijevic, A., Denys, D., Cools, R., & Jensen, O. (2014). Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. *NeuroImage*, *87*, 356–362. https://doi.org/10.1016/j.neuroimage.2013.10.052

Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron*, *40*(6), 1241–1250. https://doi.org/10.1016/s0896-6273(03)00764-5

McDermott, K. C., Malkoc, G., Mulligan, J. B., & Webster, M. A. (2010). Adaptation and visual salience. *Journal of Vision*, *10*(13), 17. https://doi.org/10.1167/10.13.17

McMains, S. A., Fehd, H. M., Emmanouil, T.-A., & Kastner, S. (2007). Mechanisms of feature-and space-based attention: Response modulation and baseline increases. *Journal of Neurophysiology*, *98*(4), 2110–2121.

Mendoza, D., Schneiderman, M., Kaul, C., & Martinez-Trujillo, J. (2011). Combined effects of feature-based working memory and feature-based attention on the perception of visual motion direction. *Journal of Vision*, *11*(1), 11. https://doi.org/10.1167/11.1.11

Mirpour, K., & Bisley, J. W. (2012). Dissociating activity in the lateral intraparietal area from value using a visual foraging task. *Proceedings of the National Academy of Sciences*, *109*(25), 10083–10088. https://doi.org/10.1073/pnas.1120763109

Moher, J., & Egeth, H. E. (2012). The ignoring paradox: Cueing distractor features leads first to selection, then to inhibition of to-be-ignored items. *Attention, Perception, & Psychophysics*, *74*(8), 1590–1605. https://doi.org/10.3758/s13414-012-0358-0

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-Based Visual Selection: Evidence From Perceptual Completion. *Psychological Science*, *9*(2), 104–110. https://doi.org/10.1111/1467-9280.00019

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*(6921), 370–373.

Moran, J., & Desimone, R. (1985). Selective Attention Gates Visual Processing in the Extrastriate Cortex. *Science*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Mulckhuyse, M., Belopolsky, A. V., Heslenfeld, D., Talsma, D., & Theeuwes, J. (2011). Distribution of Attention Modulates Salience Signals in Early Visual Cortex. *PLOS ONE*, *6*(5), e20379. https://doi.org/10.1371/journal.pone.0020379

Mullen, K. T. (2019). The response to colour in the human visual cortex: The fMRI approach. *Current Opinion in Behavioral Sciences*, *30*, 141–148. https://doi.org/10.1016/j.cobeha.2019.08.001

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, *57*(1), 1–17. https://doi.org/10.3758/BF03211845

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology*, *83*(5), 2580–2601. https://doi.org/10.1152/jn.2000.83.5.2580

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

Nothdurft, H.-C. (1993a). Saliency effects across dimensions in visual search. *Vision Research*, *33*(5), 839–844. https://doi.org/10.1016/0042-6989(93)90202-8

Nothdurft, H.-C. (1993b). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, *33*(14), 1937–1958. https://doi.org/10.1016/0042-6989(93)90020-W

Nothdurft, H.-C. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research*, *40*(10–12), 1183–1201.

Oberauer, K. (2019). Working memory and attention—A conceptual analysis and review. *Journal of Cognition*, *2*(1).

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–587.

Ogawa, T., & Komatsu, H. (2004). Target Selection in Area V4 during a Multidimensional Visual Search Task. *Journal of Neuroscience*, *24*(28), 6371–6382. https://doi.org/10.1523/JNEUROSCI.0569-04.2004

Ogawa, T., & Komatsu, H. (2006). Neuronal dynamics of bottom-up and top-down processes in area V4 of macaque monkeys performing a visual search. *Experimental Brain Research*, *173*(1), 1–13. https://doi.org/10.1007/s00221-006-0362-5

Olivers, C. N. (2008). Interactions between visual working memory and visual attention. *Frontiers in Bioscience*, *13*(3), 1182–1191.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243–1265. https://doi.org/10.1037/0096-1523.32.5.1243

Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334.

Pearson, D., Watson, P., & Le Pelley, M. E. (2021). How do competing influences of selection history interact? A commentary on Luck et al. (2021). *Visual Cognition*, *0*(0), 1–4. https://doi.org/10.1080/13506285.2021.1912234

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Poltoratski, S., Ling, S., McCormack, D., & Tong, F. (2017). Characterizing the effects of feature salience and top-down attention in the early visual system. *Journal of Neurophysiology*, *118*(1), 564–573. https://doi.org/10.1152/jn.00924.2016

Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. https://doi.org/10.1016/j.neuroscience.2005.06.005

Reynolds, J. H., & Desimone, R. (2003). Interacting Roles of Attention and Visual Salience in V4. *Neuron*, *37*(5), 853–863. https://doi.org/10.1016/S0896-6273(03)00097-7

Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Ringo, J. L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behavioural Brain Research*, *76*(1), 191–197. https://doi.org/10.1016/0166-4328(95)00197-2

Roussy, M., Mendoza-Halliday, D., & Martinez-Trujillo, J. C. (2021). Neural Substrates of Visual Perception and Working Memory: Two Sides of the Same Coin or Two Different Coins? *Frontiers in Neural Circuits*, *15*, 131. https://doi.org/10.3389/fncir.2021.764177

Runeson, E., Boynton, G. M., & Murray, S. O. (2013). Effects of task and attentional selection on responses in human visual cortex. *Journal of Neurophysiology*, *109*(10), 2606–2617.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*(7), 631–632.

Sahan, M. I., Sheldon, A. D., & Postle, B. R. (2020). The Neural Consequences of Attentional Prioritization of Internal Representations in Visual Working Memory. *Journal of Cognitive Neuroscience*, *32*(5), 917–944. https://doi.org/10.1162/jocn_a_01517

Saproo, S., & Serences, J. T. (2014). Attention Improves Transfer of Motion Information between V1 and MT. *Journal of Neuroscience*, *34*(10), 3586–3596. https://doi.org/10.1523/JNEUROSCI.3484-13.2014

Sauter, M., Liesefeld, H. R., Zehetleitner, M., & Müller, H. J. (2018). Region-based shielding of visual search from salient distractors: Target detection is impaired with same- but not different-dimension distractors. *Attention, Perception, & Psychophysics*, *80*(3), 622–642. https://doi.org/10.3758/s13414-017-1477-4

Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, & Psychophysics*, *72*(6), 1455–1470. https://doi.org/10.3758/APP.72.6.1455

Schacter, D. L., & Buckner, R. L. (1998). Priming and the Brain. *Neuron*, *20*(2), 185–195. https://doi.org/10.1016/S0896-6273(00)80448-1

Schall, J. D., & Hanes, D. P. (1993). Neural basis of saccade target selection in frontal eye field during visual search. *Nature*, *366*(6454), Article 6454. https://doi.org/10.1038/366467a0

Schall, J. D., Morel, A., King, D. J., & Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: Convergence and segregation of processing streams. *Journal of Neuroscience*, *15*(6), 4464–4487. https://doi.org/10.1523/JNEUROSCI.15-06-04464.1995

Scotti, P. S., Chen, J., & Golomb, J. D. (2021). *An enhanced inverted encoding model for neural reconstructions* (p. 2021.05.22.445245). https://doi.org/10.1101/2021.05.22.445245

Serences, J. T., & Boynton, G. M. (2007). Feature-Based Attentional Modulations in the Absence of Direct Visual Stimulation. *Neuron*, *55*(2), 301–312. https://doi.org/10.1016/j.neuron.2007.06.015

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214.

Serences, J. T., & Saproo, S. (2012). Computational advances towards linking BOLD and behavior. *Neuropsychologia*, *50*(4), 435–446.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of Voluntary and Stimulus-Driven Attentional Control in Human Cortex. *Psychological Science*, *16*(2), 114–122. https://doi.org/10.1111/j.0956-7976.2005.00791.x

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*(1), 38–45.

Shipp, S. (2003). The functional logic of cortico–pulvinar connections. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1438), 1605–1624. https://doi.org/10.1098/rstb.2002.1213

Shomstein, S. (2012). Object-based attention: Strategy versus automaticity. *Wiley Interdisciplinary Reviews Cognitive Science*, *3*(2), 163–169. https://doi.org/10.1002/wcs.1162

Shomstein, S., & Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting? *Perception & Psychophysics*, *64*(1), 41–51. https://doi.org/10.3758/BF03194556

Shomstein, S., Zhang, X., & Dubbelde, D. (n.d.). Attention and platypuses. *WIREs Cognitive Science*, *n/a*(n/a), e1600. https://doi.org/10.1002/wcs.1600

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*(9), 342–348.

Sprague, T. C., Adam, K. C., 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., Boynton, G. M., & Serences, J. T. (2019). *Inverted encoding models estimate sensible channel responses for sensible models*. https://doi.org/10.1101/642710

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.

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron*, *91*(3), 694–707.

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.

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.

Sreenivasan, K. K., & D'Esposito, M. (2019). The what, where and how of delay activity. *Nature Reviews Neuroscience*, *20*(8), Article 8. https://doi.org/10.1038/s41583-019-0176-7

Stilwell, B., Egeth, H., & Gaspelin, N. (2022). *Electrophysiological Evidence for the Suppression of Highly Salient Distractors*.

Stilwell, B., & Gaspelin, N. (2021). Attentional Suppression of Highly Salient Color Singletons. *Journal of Experimental Psychology Human Perception & Performance*, *47*. https://doi.org/10.1037/xhp0000948

Stilwell, B. T., Bahle, B., & Vecera, S. P. (2019). Feature-based statistical regularities of distractors modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 419.

Stilwell, B. T., & Vecera, S. P. (2022). Testing the underlying processes leading to learned distractor rejection: Learned oculomotor avoidance. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-022-02483-6

Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual Topography of Human Intraparietal Sulcus. *The Journal of Neuroscience*, *27*(20), 5326–5337. https://doi.org/10.1523/JNEUROSCI.0991-07.2007

Thayer, D. D., Bahle, B., & Hollingworth, A. (2021). Guidance of attention from visual working memory is feature-based, not object-based: Implications for models of feature binding. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001116

Thayer, D. D., Miller, M., Giesbrecht, B., & Sprague, T. C. (2022). Learned feature regularities enable suppression of spatially overlapping stimuli. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-022-02612-1

Thayer, D. D., & Sprague, T. C. (2023). *Feature-specific salience maps in human cortex* (p. 2023.03.29.534828). bioRxiv. https://doi.org/10.1101/2023.03.29.534828

Thayer, D., & Sprague, T. (2023). *Interactions between working memory and attention depend on remembered feature dimension*. https://doi.org/10.31234/osf.io/sx6d5

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, *50*(2), 184–193. https://doi.org/10.3758/BF03212219

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. https://doi.org/10.3758/BF03211656

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2023a). *The Attentional Capture Debate: When Can We Avoid Salient Distractors and When Not?* (1). *6*(1), Article 1. https://doi.org/10.5334/joc.251

Theeuwes, J. (2023b). *The Attentional Window, Search Difficulty and Search Modes: A Reply to Commentaries on Theeuwes (2023)* (1). *6*(1), Article 1. https://doi.org/10.5334/joc.305

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In *Attention and Performance*. MIT Press.

Theeuwes, J., de Vries, G.-J., & Godijn, R. (2003). Attentional and oculomotor capture with static singletons. *Perception & Psychophysics*, *65*(5), 735–746. https://doi.org/10.3758/BF03194810

Tootell, R. B. H., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The Retinotopy of Visual Spatial Attention. *Neuron*, *21*(6), 1409–1422. https://doi.org/10.1016/S0896-6273(00)80659-5

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156–177. https://doi.org/10.1016/S0734-189X(85)80004-9

Treisman, A. (1998). The perception of features and objects. In *Visual attention* (pp. 26–54). Oxford University Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treue, S., & Trujillo, J. C. M. (1999a). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579.

Treue, S., & Trujillo, J. C. M. (1999b). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579.

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *eLife*, *7*, e38677. https://doi.org/10.7554/eLife.38677

Vatterott, D. B., & Vecera, S. P. (2012). Experience-dependent attentional tuning of distractor rejection. *Psychonomic Bulletin & Review*, *19*(5), 871–878. https://doi.org/10.3758/s13423-012-0280-4

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), Article 7067. https://doi.org/10.1038/nature04171

Wade, A. R., Brewer, A. A., Rieger, J. W., & Wandell, B. A. (2002). Functional measurements of human ventral occipital cortex: Retinotopy and colour. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *357*(1424), 963–973. https://doi.org/10.1098/rstb.2002.1108

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383.

Wandell, B. A., & Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends in Cognitive Sciences*, *19*(6), 349–357.

Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture independent of search strategy. *Attention, Perception, & Psychophysics*, *80*(7), 1763–1774. https://doi.org/10.3758/s13414-018-1562-3

Wang, L., Huang, L., Li, M., Wang, X., Wang, S., Lin, Y., & Zhang, X. (2022). An awareness-dependent mapping of saliency in the human visual system. *NeuroImage*, *247*, 118864. https://doi.org/10.1016/j.neuroimage.2021.118864

Wark, B., Lundstrom, B. N., & Fairhall, A. (2007). Sensory adaptation. *Current Opinion in Neurobiology*, *17*(4), 423–429. https://doi.org/10.1016/j.conb.2007.07.001

Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science*, *1*(Volume 1, 2015), 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

White, A. L., Runeson, E., Palmer, J., Ernst, Z. R., & Boynton, G. M. (2017). Evidence for unlimited capacity processing of simple features in visual cortex. *Journal of Vision*, *17*(6), 19–19.

White, B. J., Berg, D. J., Kan, J. Y., Marino, R. A., Itti, L., & Munoz, D. P. (2017). Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms14263

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X

Winawer, J., & Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps. *Visual Neuroscience*, *32*, E020. https://doi.org/10.1017/S0952523815000176

Witkowski, P., & Geng, J. J. (2019). Learned feature variance is encoded in the target template and drives visual search. *Visual Cognition*, *27*(5–8), 487–501. https://doi.org/10.1080/13506285.2019.1645779

Witkowski, P. P., & Geng, J. J. (2023). Prefrontal Cortex Codes Representations of Target Identity and Feature Uncertainty. *Journal of Neuroscience*, *43*(50), 8769–8776. https://doi.org/10.1523/JNEUROSCI.1117-23.2023

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 483–502. https://doi.org/10.1037/0096-1523.29.2.483

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1–8. https://doi.org/10.1038/s41562-017-0058

Wöstmann, M., Störmer, V. S., Obleser, J., Addleman, D. A., Andersen, S., Gaspelin, N., Geng, J., Luck, S. J., Noonan, M., Slagter, H. A., & Theeuwes, J. (2021). *Ten simple rules to study distractor suppression*. PsyArXiv. https://doi.org/10.31234/osf.io/vu2k3

Wöstmann, M., Störmer, V. S., Obleser, J., Addleman, D. A., Andersen, S. K., Gaspelin, N., Geng, J. J., Luck, S. J., Noonan, M. P., Slagter, H. A., & Theeuwes, J. (2022). Ten simple rules to study distractor suppression. *Progress in Neurobiology*, *213*, 102269. https://doi.org/10.1016/j.pneurobio.2022.102269

Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology*, *13*(2), 187–193.

Yildirim, F., Carvalho, J., & Cornelissen, F. W. (2018). A second-order orientation-contrast stimulus for population-receptive-field-based retinotopic mapping. *NeuroImage*, *164*, 183–193. https://doi.org/10.1016/j.neuroimage.2017.06.073

Yu, X., & Geng, J. J. (2019). The attentional template is shifted and asymmetrically sharpened by distractor context. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 336–353. https://doi.org/10.1037/xhp0000609

Yu, X., Zhou, Z., Becker, S. I., Boettcher, S. E. P., & Geng, J. J. (2023). Good-enough attentional guidance. *Trends in Cognitive Sciences*, *27*(4), 391–403. https://doi.org/10.1016/j.tics.2023.01.007

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the New York Academy of Sciences*, *1339*(1), 154.

Zhang, X., Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural Activities in V1 Create a Bottom-Up Saliency Map. *Neuron*, *73*(1), 183–192. https://doi.org/10.1016/j.neuron.2011.10.035

Zhang, Z., Sahatdjian, R., & Carlisle, N. B. (2022). Benefits from negative templates in easy and difficult search depend on rapid distractor rejection and enhanced guidance. *Vision Research*, *197*, 108031. https://doi.org/10.1016/j.visres.2022.108031

Zhaoping, L., & May, K. A. (2007). Psychophysical Tests of the Hypothesis of a Bottom-Up Saliency Map in Primary Visual Cortex. *PLOS Computational Biology*, *3*(4), e62. https://doi.org/10.1371/journal.pcbi.0030062