COMPUTATIONAL LINKING MODELS OF HUMAN SELECTIVE VISUAL ATTENTION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Daniel Birman
August 2019

This dissertation is online at: http://purl.stanford.edu/rd566sf5779

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Justin Gardner, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Kalanit Grill-Spector

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Anthony Norcia

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

To sample the important parts of the visual world observers make saccades, moving the high-resolution and color-sensitive fovea to informative locations. Choosing to make a saccade requires sampling the periphery and identifying potentially important parts of the visual scene. This *covert attention*, without eye movement, is essential to selecting information in an efficient manner. At an intuitive level covert attention is a focusing on a feature or a location in the visual world and a suppression of other irrelevant features and locations. When operationalized into the laboratory, cueing an observer with covert attention can be shown to result in improved detection, smaller thresholds of discrimination, faster reaction times, and suppression of distractors. These changes are known to be in part the result of small tweaks to the representation of visual stimuli in sensory cortex, but are also the result of context-dependent selection occurring after sensory processing has gone to completion. How attention implements this balance of sensory change and selection is a central problem for the neuroscience of vision.

Acknowledgments

This dissertation would not have been possible without the incredible patience of the nearly one hundred volunteers who participated in my experiments. I am grateful for their willingness to see a ten-hour experiment as a fun challenge and for the resolve of many of them to return over and over throughout my time at Stanford.

I have developed a habit during graduate school of glancing down the hall to see if Justin is in, or out. Early on in graduate school Justin left his door open which made it easy to walk in and tell him about my successes, failures, and ideas. I have been extraordinarily lucky to have an advisor who still leaves their door open after five years of endless interruptions. Thank you Justin, for teaching me how to read papers, pay attention to the small details, make figures that convey stories, and for showing me how to create experiments that connect vast datasets in compelling and understandable ways.

I have a deep gratitude for all of the vision science community at Stanford, in particular the faculty and students who have been a part of the Vision seminars and have helped guide my research. Kalanit Grill-Spector and Tony Norcia in particular watched out for me in my first year in a brand new lab and have continued to provide guidance at every stage of my projects. Tirin Moore's encouragement to connect my findings outside of the fMRI world have been critical to the final thesis and to my plans for the future. Bill Newsome's support (and wisdom) has been invaluable at every stage of data collection. Helping organize the Vision community's seminar with Mareike Grotheer was a highlight of my time at Stanford.

My views on vision science changed a lot during a summer course in my fourth year and I am deeply thankful to the instructors and students who I shared my time with at Cold Spring Harbor. I feel like I have a community in vision science now, which is something that I consider extremely valuable. I want to extend a special thanks to Marlene Cohen and E.J. Chichilnisky for convincing me that it might be possible to study attention in other organisms than humans and to Anne Urai for expanding my network of possible postdoc advisors. I am excited about the future path you all have pushed me toward.

I have been lucky to be able to explore the mountains of California with all of my climbing partners as well as my non-PhD-student friends while in graduate school. Andrew Lampinen, Mona

Rosenke, and all of the other Stanford Psychology climbers have been a wonderful source of deep research conversations on car rides and while sitting on ledges. When not in the mountains I have experienced the Bay Area with my cohort and the other students in the Psychology department. Particular thanks to Sophie Bridgers, MH Tessler, and Kara Weisman for inviting me to join their apartment in my fourth year and to Andrew Lampinen for joining us the year after. It's impossible to convey how helpful it is to come home to a community of shared experiences. Perhaps most importantly every idea, data point, and figure in this dissertation, even the possible options for fonts, were in part the product of my office mate Ian Eisenberg. My memories of our office are filled with rainbows, soylent, puzzles, and rockets.

I could not have imagined the possibility of being here without the support and example of my family. My parents and grandparents have made it easy for me to follow my interests wherever they lead and supported me at every step of the process. I am extraordinarily lucky to have grown up in the intellectual and scientific environment that I have always been surrounded by.

Finally, I want to express my endless thanks to Allison Ong for sharing with me the joys and struggles of this experience and many others, past and future.

Author contributions

All chapters were written by D. Birman and describe research designed and executed solely by him, with the following exceptions:

Chapter 2, 3, and 4 describe research designed in collaboration with J. L. Gardner.

Chapters 2 and 3 were written in collaboration with J. L. Gardner.

Chapter 2 is adapted from a published paper: Birman and Gardner (2018). A quantitative framework for motion visibility in human cortex. *Journal of neurophysiology*, 120(4), 1824-1839.

Chapter 3 is adapted from a published paper: Birman and Gardner (2019). A flexible readout mechanism of human sensory representations. *Nature Communications*, 10(1), 1-13.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

In this thesis, I investigate the processes underlying selective visual attention in human cortex. The overarching goal is to establish how information is selected from visual representations for use during adaptable behavior. To do this, I start by extending an existing linking model of contrast, which controls the visibility of images, to a second feature: coherence. These two ways of manipulating the visibility of motion require a new framework to be built, demonstrating how visual cortex is sensitive to each of these features and whether they interact. I then connect this framework to perception. This connection uses a computational linking model and allows me to test different hypotheses about how perception depends on sensory representations. I show with these models that the scale of sensory changes due to attention are insufficient to capture perception, implicating a downstream readout process as a key component of selective visual attention. I expect these results to hold true for other forms of selection, e.g. by spatial location, color, or motion direction, and move toward demonstrating this in the final section of the thesis.

1.1.1 Aim 1: A flexible readout mechanism of human sensory representations

The prevailing view is that attention is implemented by modifications of sensory representations. Without a computational linking model it is impossible to know whether these sensory changes are the only changes occurring during selective visual attention.

In Chapter 2, I build a quantitative framework for features which control the visibility of motion. Contrast, motion coherence, and duration all control the visibility of motion and their representations in human visual cortex are similar. This makes them excellent tools to study whether sensory change alone can account for the behavioral effects of attention. In this chapter, I

measure and quantify how these features are represented in human cortex and lay the ground work for constructing a linking model.

In Chapter 3, I use a linking model to show that sensory change is insufficient to account for all the behavioral effects of attention. I first validate that a computational linking model of motion visibility can be constructed, based on the quantitative framework in Chapter 2. Then, using the validated linking model I show that the sensory change occurring during selective visual attention is insufficient to account for behavioral changes – flexible readout must be a necessary component.

1.1.2 Aim 2: Comparing different forms of sensory selection on a shared perceptual metric

One question we asked ourselves after completing Aim 1 was whether our results would be consistent in other features. Is selection by contrast and coherence similar to selection by spatial location, color, or motion direction? Although we know that different visual features are processed in different ways, this does not necessarily mean that selection by these features requires different computational resources and implementations.

In Chapter 4, I address this aim by building a psychophysical task with which the strength of spatial and feature-based attention can be compared on a common metric. I demonstrate in these experiments that the sensitivity (i.e. the variability in response error) is similar for different forms of selection. This suggests that selection by color, motion direction, and location may all be implemented by similar mechanisms.

Before going into detail about my experiments and computational models, I will review what is known about selective visual attention. In particular, I will focus on the history of how selection might be implemented as a computation in the human brain. I will also cover existing uses of computational linking models.

1.2 Selective attention

We all know, more or less intuitively, what it is like to attend to something that we see. When stopped at an intersection, a driver might be motivated by their context to focus on the direction of nearby cars, the movement of pedestrians, and the stop signs or traffic lights in their vicinity (Fig. 1.1a). The ease with which we can deploy attention hides a complex set of changes which occur inside the brain and which change our sensory perceptions. By moving such experiences into the laboratory we can operationalize them (Fig. 1.1b). That is to say, we can break down a complex process such as attention into discrete measurable quantities and control them using the parameters of a task. For example, attention might improve reaction times, detection, or the ability to discriminate between stimuli, but not necessarily all of these and not necessarily all to an equal extent. Once operationalized, we can begin to track the results of attention back to its roots inside the brain.

1.2.1 A brief history of selective attention research

Early attention research was based on introspective observations of the experimenter's conscious experience. One of these early observations was that percepts become more "intense" or "clear" when focused upon (Helmholtz, 1924; James, 1981; Kuelpe, 1902; Titchener, 1908). Although intuiting changes in perceptions comes with many problems (Helmholtz, 1924), these observations were not without merit. Spatial attention toward visual stimuli does lead to a perceptual change (Carrasco & Barbot, 2018). When carefully measured, it can be shown that these enhancements in perception trade off with a loss of information about unattended stimuli. An early example of this is a set of experiments in which observers were asked to echo speech from one ear at a time (Cherry, 1953). Observers can do this, but they often recall little to no information about the ear they are asked to ignore. Some low-level details leak through such as the pitch of the speakers voice (Cherry, 1953). Together what these early findings revealed is that there is a common notion of *attention* and it leads to a profound change in our conscious experience. A vast literature has since been devoted to understanding how these changes occur as a function of neural activity in the brain.

One of the steps necessary to go from intuition to an implementation-level understanding was to shift research toward objective measurement. The transition from introspection to objective measurement can be highlighted in the way that the experiment mentioned above was introduced, quoting from Cherry (1953): "the 'subject' under test (the listener) is regarded as a transducer

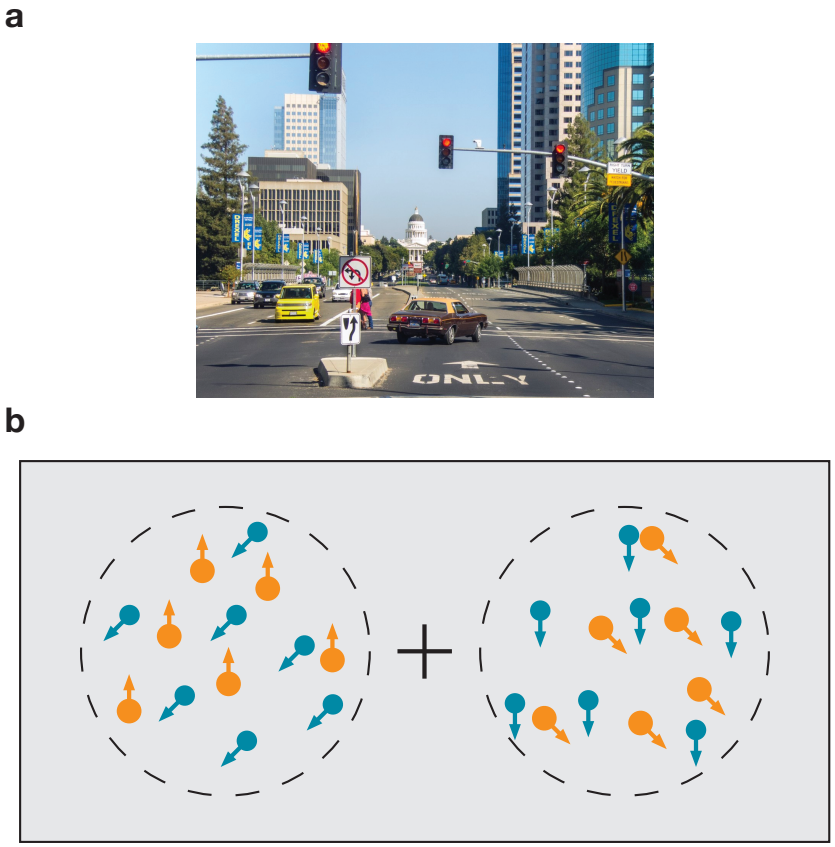

Figure 1.1: Natural and operationalized tasks which engage selective visual attention. (a) A driver, stopped at an intersection, puts more contextual weight on information such as the direction of nearby cars, the movement of pedestrians, and stop signs or traffic lights in their vicinity. While those features of the visual environment might be enhanced, irrelevant information, like the colors of the sky or houses, might be filtered out of their perceptual experience. (b) To operationalize selective visual attention we simplify complex tasks like that presented in (a), to forms where the stimulus can be precisely manipulated. In this example task, attention could be used to select two of the four dot patches and report their direction of motion. Tasks like this allow us to compare different forms of attention: for example, selecting the two dot patches by color (blue or orange) or selecting the two dot patches according to their location (left or right).

whose responses are observed when various stimuli are applied, whereas his subjective impressions are taken to be of minor importance". This approach of seeing an observer as a kind of computational processor has been effective. The study made an abrupt move toward understanding sensory systems as processing units and differentiating between the computations a system might employ versus the implementation of those computations (Marr & Vision, 1982). The rough framework for how sensory processing occurs and how attention might interface with this is as follows: any given stimulus might

be processed in a serial or parallel manner and processing might go to 'completion' or be halted at a certain point, based on an observer's attentional state. Measuring the knowledge of an observer about an attended or ignored stimulus therefore assesses the extent of processing which has occurred as well as the extent of conscious access to that processed representation. For example, for the echoing task described earlier, an observer might be asked whether they had retained knowledge about an unechoed (and therefore, presumably unattended) voice. If they could recall simple low-level features of the unattended voice, for example that the speaker's voice was higher-pitched, then a researcher could conclude that parallel processing of both streams had occurred up to pitch processing (Cherry, 1953). From the same data a second conclusion can also be reached: processing of the ignored stimulus must have been halted before the point of complete semantic understanding. These kinds of results about auditory attention coincide neatly with early findings about the human visual cortex. Early stages of visual processing are thought to occur in a parallel manner, in which simple sensory detectors are repeated across retinotopic space (Kuffler, 1953; Hubel & Wiesel, 1962, 1968). These ideas led researchers to begin to distinguish between stages of processing: an early parallel stage in which incoming sensory information is processed without immediate limits in capacity, and a second limited capacity serial stage from which complex decisions are made.

Based on the ideas above, early theories of attention focused on when attentional selection might occur relative to the parallel and serial stages of processing. As described above, some features of auditory stimuli (and other senses) are available for decision making regardless of the observers focus. Based on this, researchers suggested a bottleneck during processing and proposed that this was the mechanistic implementation of selective attention. An early example of this was Broadbents Filter theory (Broadbent, 1958) which includes an early bottleneck. In Filter theory, visual information is processed in parallel until low-level features (location, intensity, frequency) are resolved. At this point, parallel processing gives way to a serial complete processing of object identity, form, etc. An alternative theory, late selection, suggests that processing occurs up to semantics in an unconscious parallel manner (Deutsch & Deutsch, 1963). An example best illustrates the distinctions between each theory. Again, if an observer is echoing one speaker while ignoring a second, a late selection account predicts that a substantial amount of information is nevertheless available about the ignored voice. This is because late selection predicts that semantic-level processing of speech will occur regardless of sensory selection, leaving high-level information available to an observer. Evidence for this comes from experiments in which observers orient to highly salient but also very high-level features such as their own name, even when focused on other tasks (Moray, 1959). These two theories have since been reconciled by suggesting that not all features are processed in identical ways and that selection is graded (Treisman, 1960) or has variable capacity limits at different stages (Kahneman, 1973).

This thesis is primarily focused on visual attention, but the results are likely to be broadly applicable to sensory selection. Selective attention has been most heavily studied in the visual

domain, in part because of the close relationship between selecting information in the world and orienting of the eyes. There are two ways to orient visual attention. An observer can make an overt movement of the eyes or they can covertly attend, without moving the eyes. Like all forms of selective attention, covert attention can accelerate responses (Eriksen & Hoffman, 1972; Posner, Snyder, & Davidson, 1980), improve detection performance, and increase discrimination sensitivity (Carrasco, 2011). Cues about important locations can both be imposed externally on an observer (Posner, 1980) or the result of internal guiding of attention toward a cued location. For the remainder of this introduction I will focus on covert visual attention and how researchers have suggested it might be implemented in human visual cortex.

1.3 Organization of the human (and primate) visual cortex

The selection of visual information is primarily thought to occur in cortex and not in the retina or in thalamic relay areas. Briefly, visual perception begins when light reaches the cones and rods in the retina. These photoreceptors transduce photons into electrical signals while maintaining spatial precision and implicitly coding information about wavelength. Substantial processing occurs in parallel within the retina by the many different retinal ganglion cells and related interneurons (Field & Chichilnisky, 2007). These cells then project their outputs to the lateral geniculate nucleus (LGN), a relay area within the thalamus. Again, considerable processing is known to occur within the LGN, and already in this second visual processing area modulation by attention has been measured (O'Connor, Fukui, Pinsk, & Kastner, 2002). It is in the LGN that visual information becomes contra-lateralized, i.e. the inputs from the two visual fields in each retina are separated and sent to the opposite hemispheres. The LGN sends its outputs to multiple areas, but primarily into striate cortex area V1. Area V1 contains a full retinotopic map of the contralateral visual field and neurons in this area are sensitive to low-level features: contrast, spatial orientation and frequency, and color, among other simple visual features.

Area V1 sends its projections on to a multitude of retinotopic areas (V2, V3, etc) which together are referred to as "early visual cortex". These areas have progressively larger receptive fields (Dumoulin & Wandell, 2008) and are sensitive to more complex features. Eventually, visual processing differentiates into at least two streams (Mishkin, Ungerleider, & Macko, 1983) which encode primarily for object identity or temporal information like motion. The ventral stream (hV4 and regions in ventral temporal cortex) is thought to encode information about the form (Desimone & Schein, 1987), color, and texture of objects (Okazawa, Tajima, & Komatsu, 2015). The dorsal stream (MT and posterior parietal cortex) is thought to encode information about spatial processing, depth, and motion (Britten, Shadlen, Newsome, & Movshon, 1993). These broad definitions are based on lesion studies which showed a double dissociation where ventral temporal cortex lesions lead to deficits in pattern and object recognition but not grasping behaviors, and vice versa. All modern models of

these processing streams acknowledge that they are far from distinct and are deeply interconnected.

In this thesis, I focus in particular on human cortical areas that are important for the processing of low-level visual features, such as contrast (area V1, orange shading Fig. 1.2) and motion coherence (area hMT+, purple shading Fig. 1.2). I will demonstrate that these areas are parametrically related to each of these properties, respectively, and that their responses are scaled in a manner which reflects human perception to each feature. Other areas that are likely also involved in representing these features include V2, V3, V3A and V3B, hV4, and V7 – all retinotopic areas in early visual cortex (Wandell, Dumoulin, & Brewer, 2007). In Aim 2, I focus on features such as color and motion, which also implicate similar parts of human visual cortex.

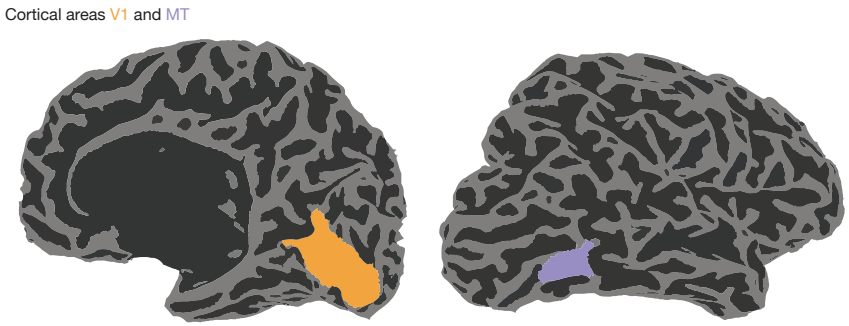

Figure 1.2: Human cortical areas implicated in the representation and perception of motion visibility. Two brain areas in the human visual cortex: V1 and MT, are shown highlighted on a reconstructed cortical surface. These two regions are critical areas in the processing of motion visibility.

1.4 Implementations of selective visual attention

Selective attention is a balancing act for the brain, which must weigh the possibility of needing unattended information against the strength of sensory selection. In early selection theories information is being thrown out before complete processing – which could be potentially disadvantageous if a stimulus later becomes important for behavior (Mack & Rock, 1998). Any modification of sensory representations to enhance one visual feature will have a cost for others. Attention research has always been aware of this and in a few particularly dramatic demonstrations (Haines, 1991; Mack & Rock, 1998; Neisser, 1979; Simons & Chabris, 1999) observers can be shown to entirely lose access to otherwise highly salient information. These effects of selection require observers to be performing a task with significant cognitive load (Lavie, 2005; Lavie, Hirst, de Fockert, & Viding, 2004; Rees, Frith, & Lavie, 1997). What this demonstrates is that the implementation of selective visual attention involves the gating of information transfer between cortical areas, where some information is passed on for additional processing while other information is not.

Modern neuroscience now deploys a multitude of neural recording techniques to understand how

sensory selection might be implemented by the brain. Both at a coarse scale in humans and at the level of individual neurons in primates and, very recently, rodents. In humans and non-human primates attention has been shown to alter the response gain of neurons in the visual system, including in the LGN (O'Connor et al., 2002), in V1 (Motter, 1993), V2 (Buffalo, Fries, Landman, Liang, & Desimone, 2010; Luck, Chelazzi, Hillyard, & Desimone, 1997; Motter, 1993), V3 (Liu, Larsson, & Carrasco, 2007b; Pestilli, Carrasco, Heeger, & Gardner, 2011; Saenz, Buracas, & Boynton, 2002; Silver, Ress, & Heeger, 2007), V4 (Buffalo et al., 2010; Connor, Gallant, Preddie, & Van Essen, 1996; Luck et al., 1997; McAdams & Maunsell, 1999; Moran & Desimone, 1985; Motter, 1993; Reynolds, Pasternak, & Desimone, 2000; Spitzer, Desimone, & Moran, 1988), V3A (Serences & Boynton, 2007), MT (Beauchamp, Cox, & DeYoe, 1997; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997; Saenz et al., 2002; Seidemann, Poirson, Wandell, & Newsome, 1999; Serences & Boynton, 2007; Treue & Martínez Trujillo, 1999; Treue & Maunsell, 1996) and MST (O'Craven et al., 1997; Treue & Maunsell, 1996), and in IT cortex (Chelazzi, Duncan, Miller, & Desimone, 1998; Moran & Desimone, 1985). Using BOLD imaging these changes can be observed simultaneously throughout almost all of early visual cortex (Liu et al., 2007b; Pestilli et al., 2011; Saenz et al., 2002; Silver et al., 2007), ventral temporal cortex (Baldauf & Desimone, 2014), and in some tasks throughout association areas as well (Çukur, Nishimoto, Huth, & Gallant, 2013). Changes to sensory representations occur for spatial attention tasks (Klein, Harvey, & Dumoulin, 2014; McAdams & Maunsell, 1999; Mitchell, Sundberg, & Reynolds, 2009; Pestilli et al., 2011; Womelsdorf, Anton-Erxleben, Pieper, & Treue, 2006) and are thought to be linked to preparatory signals (Tolias et al., 2001; Moore, Armstrong, & Fallah, 2003; Moore & Fallah, 2001) originating in the frontal eye fields prior to saccades (Moore & Armstrong, 2003). Many of the examples above involved tasks in which spatial attention was not deployed, but instead observers shift feature-based attention (Baldauf & Desimone, 2014; Harel, Kravitz, & Baker, 2014; Huk & Heeger, 2000; Jehee, Brady, & Tong, 2011; Saenz et al., 2002; Sàenz, Buraĉas, & Boynton, 2003; Serences & Boynton, 2007; Treue & Martínez Trujillo, 1999; Çukur et al., 2013).

Several general hypotheses describe how sensory representations might be modified during selective visual attention (Fig. 1.3). One early hypothesis was that neurons sensitive to an attended feature might become more sensitive during a selection behavior (Reynolds et al., 2000; Serences & Boynton, 2007; Snyder, Yu, & Smith, 2018; Treue & Martínez Trujillo, 1999) (Fig. 1.3a). Such a change is often modeled as a multiplicative gain, i.e. where $R'(s) = \alpha R(s)$, with α being a parameter fit to measurements of neural activity. Another possibility is that neurons increase their baseline response (Buracas & Boynton, 2007; Chen & Seidemann, 2012; Fang, Boyaci, Kersten, & Murray, 2008; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Li, Lu, Tjan, Dosher, & Chu, 2008) (Fig. 1.3b). Baseline shifts are modeled as an additive offset, i.e. $R'(s) = R(s) + \beta$. A shift in baseline response for some neurons, but not others, could be used by downstream mechanisms to select out the attended stimulus (Pestilli et al., 2011; Hara & Gardner, 2014). A third possibility is

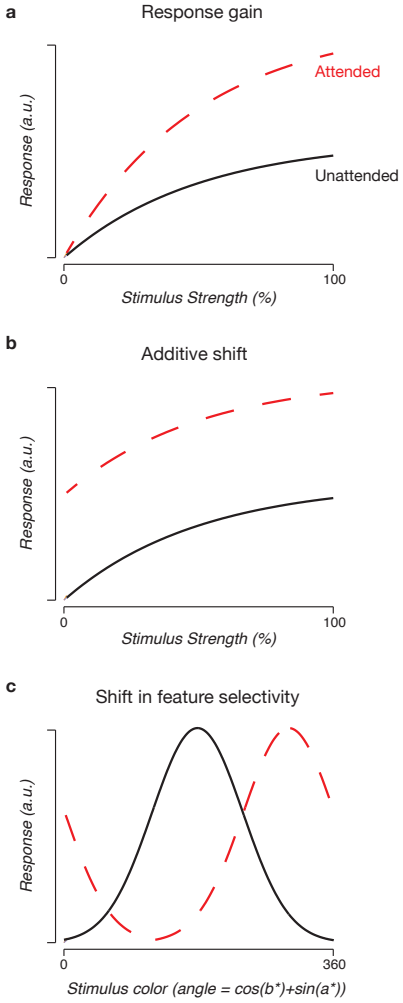

Figure 1.3: Implementations of attention in sensory representations. (a) Multiplicative gain of the sensitivity of neurons during attention to a preferred feature. (b) An additive shift in the neural response, which could be used by a downstream mechanism to differentiate attended and unattended features. (c) A shift in the preferred feature of a neuron to a different feature.

that neurons change sensitivity, so that more, or different, neurons in the population begin to code for the relevant stimulus feature (Çukur et al., 2013; David, Hayden, Mazer, & Gallant, 2008; Kastner, De Weerd, Desimone, & Ungerleider, 1998; Klein et al., 2014; Spitzer et al., 1988; Womelsdorf et al., 2006; Womelsdorf, Anton-Erxleben, & Treue, 2008) (Fig. 1.3c). A shift can be represented as $R'(s) = R(s - \gamma)$, where again γ is a free parameter to be fit according to the data. Finally, the population of neurons might not change their response characteristics, but instead top-down signals might modify the structure of stimulus-driven and noise correlations between neurons (Cohen & Maunsell, 2009; Mitchell et al., 2009; Ruff & Cohen, 2016; Verhoef & Maunsell, 2017). These

changes in the population code are hypothesized to make it easier to linearly decode the relevant stimulus-driven signals from internal noise (Snyder et al., 2018; Ecker, Denfield, Bethge, & Tolias, 2016; Rabinowitz, Goris, Cohen, & Simoncelli, 2015). Many of these changes may reflect a single implementation (Reynolds & Heeger, 2009) with apparent differences being the result of the exact combination of a particular task and stimulus.

Our understanding of the neural implementation has advanced dramatically thanks to recordings of neuronal populations from animal models. But animal models have also held back research in important ways. Animals and humans must learn selective visual attention tasks in dramatically different ways (Birman & Gardner, 2015) which makes it difficult to compare results between species. In general it is also not possible to probe an animals memory about unattended stimuli. Because of these differences it is often complex to synthesize the results of research in different model systems. Recently, mice have been shown to be able to exhibit selective attention (McBride, Lee, & Callaway, 2019; Wang & Krauzlis, 2018; Nakajima, Schmitt, & Halassa, 2019). This is promising for understanding the role of different brain areas in attentional behaviors, but also concerning. The mice in these studies exhibit a bias which resembles selective attention, but they also have high lapse rates compared to human observers. This may be because mice continue to explore the experiment space (Pisupati, Chartarifsky-Lynn, Khanal, & Churchland, 2019) whereas humans who learn from rules do not. Correctly taking these differences into account, perhaps by linking animal research directly to parallel human research, has a good chance of overcoming these issues.

1.4.1 Which features can survive inattention?

Although attending to a stimulus often results in changes in sensory representation, there are a few instances in which visual information seems to be processed no matter what. In vision, scene gist can survive inattention, perceptually (Li, VanRullen, Koch, & Perona, 2002) and as decodable information from measurements of BOLD signal in visual cortex (Peelen, Fei-Fei, & Kastner, 2009). This is also true in audition, where highly salient features like names may pop out for subjects despite a demanding task (Moray, 1959).

One of the easiest operationalized tasks in which to observe that different features are affected by attention in different ways is during search (Wolfe, 1994). In a search task, an observer will be cued in advance about the properties of an item and must find its location or detect its presence. The item will be hidden among a set of distractors whose properties determine the difficulty of the task. When the target stimulus differs from the distractors along certain key dimensions the task is trivial. Trivial, in this case, means that the processing required to solve the task occurs in parallel and the incongruent target will “pop out” of the array. The features which pop out happen to coincide with the visual properties that are encoded by neurons in the earliest visual areas (Barlow, Fitzhugh, & Kuffler, 1957; Hubel & Wiesel, 1962, 1959). The parallel processing of detectors that are topographically mapped across visual space allows this behavior. Differences in the strength

of signals across these maps, for each feature, can then result in pop out of the relevant stimulus (Nothdurft, 1993; Treisman, 1985). Difficult search tasks involve conjunctions of stimulus properties (Egeth, Virzi, & Garbart, 1984) and appear to require attention be directed in a serial manner to each item (Treisman & Gelade, 1980).

The physiology of early visual cortex, in particular the repetitive small receptive fields in early visual cortex, suggest one explanation for why some features are processed regardless of attentional state. Behavioral relevance and past experience may explain why scene gist (Li et al., 2002; Peelen et al., 2009) and names (Moray, 1959) are also processed in the absence of attention.

1.5 Computational linking models

Measurements of the neural effects of selective attention are not sufficient to understand its implementation, they must be linked correctly to behavior. To reconcile changes in cortical activity with behavior cognitive neuroscientists link these with computational linking models (Barlow, 1972; Brindley, 1960; Cohen & Maunsell, 2010; Newsome, Britten, & Movshon, 1989; Pestilli et al., 2011; Cook & Maunsell, 2002). One assumption underlying much of cognitive neuroscience is that when we make a measurement of cortical activity, we are seeing the same signals that the brain is using to solve sensory decision making. This is only an assumption; it is possible that sensory decision making (and other forms of neural processing) are based on subsets of signals, or population codes, which remain harder to measure. To avoid making errors in inference it is important to make our hypotheses (and assumptions) about possible implementations explicit in a form which can be tested. Here I propose to do this by build computational models which lay out the steps from sensory signal to sensory decision. We refer to these as "linking models", as they link together perceptual and cortical measurements. Linking models are valuable because they force researchers to be explicit about the scale of behavioral and neural effects and to ensure that these match. It is not sufficient to find a change in a neural representation during attention: it must match the size of the corresponding behavioral change. In this last section, I will briefly summarize a few examples of linking models which have shown promise in connecting selective attention behaviors to their neural implementations.

When primates exert covert spatial attention at a location neurons with receptive fields near that retinotopic location show a shift in their tuning (Klein et al., 2014; Womelsdorf et al., 2008; Womelsdorf et al., 2006; Connor et al., 1996). This is likely related to how the brain implements changes in spatial selectivity prior to and during saccades (Tolias et al., 2001; Moore & Fallah, 2001; Moore et al., 2003). Other changes in sensory representation also occur, e.g. in the correlations between neural firing across populations in visual cortex (Cohen & Maunsell, 2009). Meanwhile, covert spatial attention improves task performance. Recent work has begun to connect these measurements with linking models, suggesting that the shifts in receptive fields and in neural firing patterns can

directly account for the behavioral effects (Klein, Paffen, Pas, & Dumoulin, 2016; Vo, Sprague, & Serences, 2017; Cohen & Maunsell, 2011, 2009). Other modeling work has shown that these receptive field shifts are broadly consistent with gain changes in early visual cortex (Baruch & Yeshurun, 2014; Miconi & VanRullen, 2016). Taken together, these results demonstrate a direct computational link between the changes in sensory representations in visual cortex and the behavioral effects of covert spatial attention.

In more complex tasks these direct link sometimes breaks down. In a recent paper Pestilli et al. (2011) found that a simple linking model of response gain during spatial attention was quantitatively insufficient to explain behavior. In that work, the authors found that to explain behavior two steps were necessary: a change in sensory representation, combined with a particular form of readout. Together these were able to capture how the seemingly small changes in sensory representation could lead to large improvements in perceptual sensitivity. Because the authors observed a correlated sensory change during attention an incorrect conclusion could easily have been made here. The linking model was necessary to demonstrate that the scale of changes in the neural representation were incompatible with the scale of behavior enhancement during attention.

Linking models have also been used to try to isolate which sensory responses are of the right magnitude and shape to explain behavioral performance. For example, contrast discrimination can be linked to representations in early visual cortex, because of their corresponding scales (Boynton, Demb, Glover, & Heeger, 1999). This approach has also been used to identify neurons that may individually contribute to perceptual decisions (Newsome et al., 1989).

Chapter 2

A quantitative framework for motion visibility in human cortex

2.1 Introduction

Much of the neural basis of perception has been revealed by manipulations that control the visibility of motion stimuli. For example, global motion direction of random-dot stimuli is made less visible by decreasing motion coherence, i.e., the percentage of dots moving in the same direction. At lower visibility levels, small changes in cortical signals manifest in measurable behavioral effects, thus documenting direct links between cortical physiology and perception (Britten, Shadlen, Newsome, & Movshon, 1992; Newsome et al., 1989) and uncovering neural signals supporting evidence accumulation (Huk & Shadlen, 2005; Katz, Yates, Pillow, & Huk, 2016; Roitman & Shadlen, 2002; Shadlen, Britten, Newsome, & Movshon, 1996; Shadlen & Newsome, 2001). Making stimuli brief also renders them less visible, aiding, for example, the study of information integration across eye movements (Melcher & Morrone, 2003). Increasing image contrast, the average difference between bright and dark (Bex & Makous, 2002), makes stimuli more visible and cortical responses monotonically larger allowing links to be made between cortical response and perception (Boynton et al., 1999; Ress, Backus, & Heeger, 2000; Ress & Heeger, 2003), disambiguating mechanisms for spatial attention (Carrasco, Penpeci-Talgar, & Eckstein, 2000; Hara, Pestilli, & Gardner, 2014; Hara & Gardner, 2014; Pestilli et al., 2011), uncovering neural correlates of conscious perception (Lumer, Friston, & Rees, 1998; Wunderlich, Schneider, & Kastner, 2005), and revealing the effects of putative priors (Stocker & Simoncelli, 2006; Vintch & Gardner, 2014). While each of these manipulations has been used extensively in the human perceptual literature, they can have greatly different effects on human neural response. Given the central importance of motion visibility, a quantitative model of response across human visual cortex is required to provide a framework for interpreting and building upon

these various findings.

Such a population response model must quantitatively account for the shape of the relationship between motion visibility and cortical response. The response function for contrast has been characterized as a sigmoidal function for measurements in single units (Albrecht & Hamilton, 1982; Sclar, Maunsell, & Lennie, 1990) and populations (Avidan et al., 2002; Boynton, Engel, Glover, & Heeger, 1996; Boynton et al., 1999; Gardner et al., 2005; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Olman, Ugurbil, Schrater, & Kersten, 2004; Tootell et al., 1998a). Increasing motion coherence typically results in linear increases in response (Aspell, Tanskanen, & Hurlbert, 2005; Britten et al., 1993; Händel, Lutzenberger, Thier, & Haarmeier, 2007; Rees, Friston, & Koch, 2000; Simoncelli & Heeger, 1998) although this may depend on the exact stimulus parameters (Ajina, Kennard, Rees, & Bridge, 2015).

A population response model must also quantify the variable sensitivity to visibility parameters across cortical areas. The earliest cortical areas have a larger dynamic range for contrast compared with later areas which are more invariant (Avidan et al., 2002; Cheng, Hasegawa, Saleem, & Tanaka, 1994; Rolls & Baylis, 1986; Sclar et al., 1990). Less is known about motion coherence sensitivity except that the neural response to coherent compared with incoherent motion or blank evokes a large response in the human middle temporal area (hMT+, referred to as MT) with some sensitivity reported in earlier visual cortical areas (Ajina et al., 2015; Costagli et al., 2014; Dupont, Orban, De Bruyn, Verbruggen, & Mortelmans, 1994; Heeger, Boynton, Demb, Seidemann, & Newsome, 1999; Watson et al., 1993; Zeki et al., 1991) and parietal and ventral regions (Braddick et al., 2001).

Finally this model must account for stimulus duration effects. Hemodynamic responses to visual stimuli are approximately temporally linear except when durations (Boynton et al., 1996; Boynton, Engel, & Heeger, 2012) or inter-stimulus intervals (Huettel & McCarthy, 2000) are brief. The divergence from linearity may differ across cortical areas (Birn, Saad, & Bandettini, 2001) and motion-sensitive regions may be most sensitive to transient changes (Stigliani, Jeska, & Grill-Spector, 2017).

Here we measured blood-oxygen-level dependent (BOLD) (Ogawa, Lee, Kay, & Tank, 1990) response in human observers to a large range of contrast, coherence and duration of motion stimuli, and built a quantitative model linking these visibility properties with physiological response in retinotopically defined visual areas. Sensitivity to these parameters varied significantly across areas, although all were sensitivity to both contrast and coherence without interaction. While perceptual experiments have often used different means of affecting visibility interchangeably our results provide a reference model that underscores the differences in response to each manipulation of visibility across cortical areas, thus providing a quantifiable way to interpret experiments that link cortical response to perception.

2.2 Methods

2.2.1 Observers

In total, 11 observers (8 female, 3 male; mean age 26 y; age range 1936 y) were subjects for the experiments. All observers except one (who was an author) were naive to the intent of the experiments. Observers were scanned three times, in 2 two-hour sessions of the experiment and a one hour retinotopy session. Procedures were approved in advance by the Stanford Institutional Review Board on human participants research and all observers gave prior written informed consent before they participated in the experiment. When necessary, observers wore corrective lenses to correct their vision to normal.

2.2.2 Hardware setup for stimulus and task control

Visual stimuli were generated using MATLAB (The MathWorks) and MGL (Gardner, Merriam, Schluppeck, & Larsson, 2018a) (http://gru.stanford.edu/mgl). Stimuli were backprojected via an Eiki LC-WUL100L projector (resolution of 1,9001,200, refresh rate of 100 Hz) onto an acrylic sheet mounted inside the scanner bore near the head coil. Visual stimuli were viewed through a mirror mounted on the head coil and responses were collected via an MRI-compatible button box. Output luminance was measured with a PR650 spectrometer (Photo Research) and a neutral density filter used to set the average screen luminance to 300 cd/m^2. The gamma table was then dynamically adjusted at the beginning of each trial to linearize the luminance display such that the full 10-bit output resolution of the gamma table could be used to display the maximum contrast needed. Other sources of light were minimized during scanning.

2.2.3 Eye tracking

Prior to the experiment subjects were extensively trained on a behavioral task requiring precise fixation. Eye tracking was performed using an infrared video-based eye-tracker at 500 Hz (Eyelink 1000; SR Research). Calibration was performed throughout each session to maintain a validation accuracy of less than 1°average offset from expected using either a 10-point or 13-point calibration procedure. Trials were canceled online when observers eyes moved more than 1°away from the fixation cross for more than 300 ms. After training, canceled trials consisted of fewer than 0.1% of all trials. Due to technical limitations eye tracking was not performed inside the scanner.

2.2.4 Experimental design

Motion stimuli consisted of two patches of moving dots and a central cross ($1 \times 1°$) on which observers maintained fixation. The dot patches were rectangular regions extending from 3.5 to 12°horizontal and 7 to 7°vertical. Each patch was filled with 21 dots/deg^2, 50% brighter and 50% darker than

the gray background (300 cd/m^2). Both patches maintained a constant baseline in between trials of 25% contrast and incoherent motion. During a trial, the patches increased in either or both contrast and coherence. To minimize involuntary eye movements, the coherent dot motion direction was randomized to be horizontally inward or outward from fixation on each trial, such that each patch moved in opposite direction. All dots moved at 6°/s updated on each video frame. Motion strength was adjusted by changing motion coherence; that is, the percentage of dots that moved in a common direction with all other dots moving in random directions. Dots were randomly assigned on each video frame to be moving in the coherent or random directions.

We measured the cortical response to a wide range of brief increments of stimulus contrast and coherence of variable duration while observers performed an independent and asynchronous task at fixation (Fig. 2.1). Each scan began with a 30-s baseline period (25% contrast, 0% coherence) to allow visual cortex to adapt. Each trial consisted of a brief increment in either or both the contrast and motion coherence of the dot patches. The dot patches then returned to baseline (25% contrast, 0% coherence) for an inter—trial interval of 2 to 11 s (mean 6.5 s) randomly sampled from an exponential distribution. The next trial then began synchronized to the next volume acquisition of the magnet. Stimulus increments were chosen to be +0, +25, +50, or +75% above the baseline 25% contrast and +0, +25, +50, +75, or +100% above the baseline 0% coherence and lasted for 250, 500, 1,000, 2,000, 2,500 or 4,000 ms (or as close to these durations as the display frame refresh would allow). We presented trials in two sets; a complete cross set in which all combinations of contrast and coherence changes at 2,500 ms duration were presented (4 contrasts5 coherences = 20 conditions) and a duration set in which a subset of the contrast and coherence combinations (+25 or +75 contrast and +25 or +100 coherence) were presented for variable stimulus durations (4 contrast and coherence combinations × 5 stimulus durations = 20 conditions). Thus, across the complete cross and duration sets, there was a total of 40 conditions (20 each in the complete cross and duration sets). For each condition we acquired a minimum of 20 repeated presentations throughout the scan sessions of each observer, resulting in a minimum of 800 trials total. The two trial sets were presented in separate scans interleaved within sessions. Condition order within each scan, for both trial sets, was randomized independently for the stimulus on the left and right such that in every block of 40 trials all conditions were presented in both dot patches.

While these stimuli were being presented for the passive viewing condition, the observer was required to perform a luminance decrement task on the fixation cross. The fixation cross decremented twice in luminance for 400 ms, separated by an 800-ms interstimulus interval and the observer reported with a button press which decrement interval appeared darker (see Gardner, Merriam, Movshon, and Heeger (2008) for details). Decrement amplitude was adjusted according to a staircase procedure to maintain 82% correct.

2.2.5 MRI acquisition and preprocessing

Visual area mapping and cortical measurements were obtained using a multiplexed sequence on a 3 Tesla GE Discovery MR750 (GE Medical Systems) with a Nova Medical 32ch head coil. Functional images were obtained using a whole-brain T2*-weighted two-dimensional gradient-echo acquisition (FOV = 220 mm, TR = 500 ms, TE = 30 ms, flip angle = 46°, 7 slices at multiplex 8 = 56 total slices, 2.5 mm isotropic). In addition, two whole-brain high-resolution T1-weighted 3D BRAVO sequences were acquired (FOV = 240 mm, flip angle = 12°, 0.9 mm isotropic) and averaged to form a canonical anatomical image which was used for segmentation and surface reconstruction and session-to-session alignment. A T2*-weighted scan with the phase encoding direction reversed was collected in each session and used in combination with the FSL function TOPUP to correct for distortions due to high multiplex factors (Andersson, Skare, & Ashburner, 2003). In each functional session, we also obtained a session anatomical image for alignment with the canonical anatomy using a T1-weighted 3D BRAVO sequence (FOV = 240 mm, flip angle = 12°, 1.2 × 1.2 × 0.9 mm). Analysis was performed using custom MATLAB software (Gardner, Merriam, Schluppeck, Besle, & Heeger, 2018b).

Session anatomies were aligned to the canonical anatomy and data were displayed on flattened cortical surfaces for visualization and for defining visual areas. Gray matter and white matter segmentation was performed on the canonical anatomy using FreeSurfer (Dale, Fischl, & Sereno, 1999) and flattened triangulated surfaces used for displaying data. Each session anatomy, was aligned to the canonical anatomy using image-based registration (Nestares & Heeger, 2000) so that the location of mapped cortical visual areas could be projected into each sessions space. All data analysis was performed in the native coordinate of the functional scan without transformation.

Cortical visual area mapping was performed using a population receptive field mapping technique (Dumoulin & Wandell, 2008). Observers performed the fixation task described above while a moving-bar stimulus moved across the visual field in different directions. The measured responses were used to estimate the voxel-wise population receptive field and then the eccentricity and polar angle of each receptive fields was projected onto a flattened representation of the cortical surface where visual areas were identified according to published criteria by hand (Gardner et al., 2008; Wandell et al., 2007). Each moving bar stimulus scan lasted 4 min and the same randomization sequence was repeated and averaged eight times to improve the signal-to-noise ratio. The stimulus was a full contrast 3°width bar spanning the entire display. Inside the bar a full contrast cross-hatch pattern of black and white rectangles moved continuously to minimize adaptation. Each of the 4-min scans began with a 12 s blank followed by eight 24-s cycles in which the bar swept across the entire screen in one of the eight cardinal or oblique directions. Two additional 12 s blanks occurred after the third and sixth bar sweeps to help estimate large population receptive fields. The bar swept across the visual field at 2°/s. The screen was crescent shaped and extended 25°vertical and 50°horizontal. Beyond the screen boundaries the image was blacked out to prevent artifacts from reflecting on the scanner

bore. We were able to consistently map V1-hV4, V3A/B, V7 (IPS0) and hMT+ (referred to as MT; see Huk, Dougherty, and Heeger (2002), Amano, Wandell, and Dumoulin (2009)) in all observers. Areas LO12, VO12, and IPS13 were not consistently identified and were therefore excluded from analysis.

Motion correction, linear trend removal, filtering, and averaging across cortical visual areas were performed to obtain a single time course for each cortical area for each observer. T2*-weighted images were motion corrected with a rigid body alignment using standard procedures (Nestares & Heeger, 2000). Scans within each session were linearly detrended, high-pass filtered with a cutoff frequency of 0.01 Hz to remove low-frequency drifts, converted to percent signal change by dividing each voxels time course by its mean image intensity within each scan, and then concatenated across scans.

Analyses of responses of cortical areas were conducted by averaging the time series of voxels whose trial-triggered response across all conditions accounted for the highest amount of variance within each retinotopically defined visual area. Specifically, we performed an event-related analysis to recover the response evoked by each trial (regardless of condition), using the following equation to model voxel responses

$$y = x\beta + \epsilon \tag{2.1}$$

where y is an $n \times 1$ array representing the time-series of BOLD response for n volumes from a single voxel. X is an $n \times k$ stimulus convolution matrix in which the first column contains a one for the volume when each trial began and zeros elsewhere. Each subsequent column is shifted downwards by one to form a Toeplitz matrix and k was set to 81 to model responses as occurring from the time of stimulus presentation through 40.5 s later. Each voxel is assumed to have additive Gaussian noise with variance . By computing the least-squares estimate of the column vector β, we obtained the finite impulse response evoked by all trials, that is, the average response after a trial accounting for linear response overlap. We computed r^2, the amount of variance accounted for by this model (Gardner et al., 2005). We then averaged the time series of the top 25 voxels per cortical area sorted by r^2. While we chose this voxel selection criterion to produce high signal-to-noise estimates of each cortical areas response, our conclusions did not depend on its use. Repeating the complete analysis using either all voxels in each cortical area, the top two voxels, or all voxels weighted by their receptive field overlap with the stimulus results in a change in the signal-to-noise in the data but did not qualitatively change the key findings.

To examine how the hemodynamic response for each cortical area changed as a function of stimulus condition (Fig. 2.2), we computed the finite impulse response for each condition in the passive viewing experiment. That is, we computed the finite impulse response as above, but allowed for a separate response for each of the 20 conditions in the cross set and 20 in the duration set. Our complete stimulus convolution matrix therefore had 3,240 columns (81 volumes by 40 conditions),

while each observers data consisted of at minimum 13,440 time points and up to 30,000 time points in some observers. Solving for the least squares solution results in hemodynamic response for each of the 40 conditions in the experiment which we call the measured cortical response.

2.2.6 Population response functions

Overview

Using the measured cortical responses we then estimated the population response functions for contrast and coherence in each cortical visual area. Our model framework and measurements are available online, as a tool for experiment design and comparison with existing results (Birman & Gardner, 2018). Following previous work examining the relationship between contrast or coherence and BOLD response (Avidan et al., 2002; Boynton et al., 1996; Boynton et al., 1999; Gardner et al., 2005) (Heeger, Huk, Geisler, & Albrecht, 2000; Logothetis et al., 2001; Olman et al., 2004; Rees et al., 2000; Tootell et al., 1998a) we assumed that there was a smooth functional form (linear, exponential or sigmoidal, see details below) between the contrast and coherence of the stimulus and the magnitude of neural response. For each trial, the magnitude of neural response was computed as the linear sum of the response to contrast and coherence predicted by these smooth functions and a trial onset response that was the same across all conditions (interaction terms between contrast and coherence were tested and compared against simpler models by cross-validated variance explained). The neural magnitude was used to scale the magnitude of a boxcar function of the appropriate duration exponentially scaled (see below) to account for nonlinear effects of duration. The resulting time series was then convolved with a canonical hemodynamic response function estimated from the data. The parameters of the population response functions and magnitude of the trial onset response were then adjusted to best fit the event-related responses in the least squares sense through nonlinear fitting routines (active-set algorithm implemented in *lsqnonlin* in MATLAB). To avoid over-fitting and to compare models with different numbers of parameters, we evaluated models according to the cross-validated r^2 by performing a leave-one-condition out cross-validation, using 39 of the 40 stimulus conditions to train the model while predicting on the left out condition. We proceeded with this analysis in two steps: characterizing the canonical hemodynamic response and duration effects, and then fitting the population response functions parameters.

Canonical hemodynamic response function and duration effects

We first fit parameters of the canonical hemodynamic response function and duration effects, ignoring the effect of contrast and coherence. To do so we fit the population response model with arbitrary scaling factors (beta weights) for each of the 40 conditions. This approach allowed us to determine the shape parameters of the hemodynamic response function and temporal non-linearity without being biased by magnitude differences across conditions.

We characterized the shape of the canonical hemodynamic response function for each observer with a difference of two gamma functions:

$$r_{canonical}(t) = \Gamma_1(t) - \Gamma_2(t) \tag{2.2}$$

$$\Gamma(t) = \begin{cases} \frac{\alpha[\frac{t-t_0}{\tau}]^{n-1} e^{\frac{-1}{\tau}}}{\tau(n-1)!}, & t >= 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

Where α is the amplitude, t_0 is the time lag such that when $t < t_0$ the function is zero, and n and τ control the shape of the function. The parameter α was set such that the peak response to a 500-ms stimulus was 1. Thus the reported percent signal change in the population response functions are relative to a 500-ms stimulus.

We accounted for nonlinear effects of temporal summation in the BOLD response by allowing responses to be exponentially scaled. Small variations in duration are known to scale in an approximately linear manner (Boynton et al., 1996) whereas across large variation in stimulus durations the response to longer durations is less than expected by a linear system (Boynton et al., 2012). We are agnostic to the source of this effect, which could result from either neural adaptation (Buxton, Uludağ, Dubowitz, & Liu, 2004) or due to saturation of the BOLD signal (Friston, Josephs, Rees, & Turner, 1998). We took the response of the 500 ms duration stimulus as the baseline and scaled shorter and longer responses according to the inverse ratio of the durations raised to a fit parameter δ (i.e., a 1,000-ms stimulus has a ratio of $\frac{1000}{500}^{-\delta} = 2^{-\delta}$). This final value corresponds to the proportion of a linear response that occurred and the boxcar of appropriate duration was scaled by this value.

Altogether we fit the parameters for two gamma functions in the canonical hemodynamic response function (α, t_0, τ), the duration effect δ and 40 beta weights for stimulus conditions to the event-related responses. The canonical hemodynamic response function parameters and the duration parameter were then used in the estimation of the population response functional forms while the beta weights were discarded.

Functional forms

To characterize the population responses of each visual area to changes in contrast and motion coherence we fit functional forms to the underlying neural population response functions. We assumed that these population response functions would be monotonically increasing for both contrast and coherence. For contrast, we parameterized the relationship between contrast and neural response as a sigmoidal function (Naka & Rushton, 1966) following previous work (Albrecht & Hamilton, 1982):

$$R_{con}(s_{con}) = \alpha_{con}(\frac{s_{con}^{1.9}}{s_{con}^{1.6} + \sigma^{1.6}}) \tag{2.4}$$

where α is the maximum amplitude of the function and σ controls the shape of the function. We fixed the exponent parameters of the Naka-Rushton to 1.9 and 1.6 based on previous work (Boynton et al., 1999). To avoid making assumptions about the coherence response function we assumed that the form would either be linear or a saturating non-linearity motivated by previous work (Rees et al., 2000; Simoncelli & Heeger, 1998). The saturating non-linearity was an exponential function but can interpolate smoothly between a linear and nonlinear function.

$$R_{coh}(s_{coh}) = \alpha_{coh}(1 - e^{\frac{s_{coh}}{\kappa}}) \tag{2.5}$$

In the exponential function the parameter κ controls the shape of the function by setting the point at which the exponential function reaches 63% of its maximum and α controls the amplitude. Large values of κ combined with large values of α make this function approach linear in the range [01] in which the stimulus strength s_{coh} is bounded.

To assess whether and to what extent contrast and motion coherence interact we included an additional parameter in the population response function model. The parameter $\beta_{interaction}$ scaled the multiplicative effect of contrast and motion coherence according to the following equation:

$$R_{interaction}(s_{con}, s_{coh}) = \beta_{interaction} R_{con}(s_{con}) R_{coh}(s_{coh}) \tag{2.6}$$

The full model of neural response was computed as the sum of the contrast and coherence response, the interaction term, and a constant stimulus onset effect R_{onset}.

$$R_{neural}(s_{con}, s_{coh}) = R_{con}(s_{con}) + R_{coh}(s_{coh}) + R_{interaction}(s_{con}, s_{coh}) + R_{onset} \tag{2.7}$$

We evaluated the fit of the full model with and without the additional interaction parameter by comparing the cross-validated variance explained. We also fit an alternative interaction model in which different population response functions were allowed to fit for conditions in which only one feature changed (i.e., the first column and last row of the grid in Fig. 2.2A) compared with conditions in which both features changed (other parts of the grid in Fig. 2.2A).

We fit the free parameters of the population response functions by constraining the fits on each observers cortical measurements (Fig. 2.4). To do this we computed the neural response R_{neural} and then scaled this by the boxcar of appropriate duration for each stimulus condition. The boxcar was additionally scaled according to the duration parameter. Finally we convolved this scaled boxcar with the canonical hemodynamic response resulting in a predicted hemodynamic response for each stimulus condition.

To evaluate whether the parameters we fit differed across subjects and across cortical areas we fit a linear model for each parameter. We first performed model comparison to establish whether each parameter was better explained by a model with only an intercept, a per-subject effect, a per-area effect, or a per-subject and per-area effect. For each parameter we fit all four models

(using the function *fitlme* in MATLAB) and retained the most complex model which resulted in a statistically significant improvement in prediction, assessed via partial F-test. For each parameter we then investigated which observers and cortical areas showed statistically significant differences relative to the mean parameter value as reported in Tables 2.1—2.3.

2.2.7 Computing stimulus sensitivity

For each cortical area we computed various measures of sensitivity to contrast and motion coherence. In particular, we examined the contrast parameter, which controls the maximum response of the Naka-Rushton function. Because in the range we measured the slopes are approximately linear and the R_{onset} term absorbs the stimulus-independent response, contrast tracks the slope of the relationship between contrast and response and therefore is a measure of sensitivity to contrast. The parameters of the exponential form of the coherence function we used are not interpretable in isolation so instead we took the population response functions for coherence and measured their response range by performing a linear fit. We report the slope of that fit as the sensitivity to coherence.

The measurements of sensitivity which we report will be sensitive to the signal-to-noise of our measurements. This could be particularly problematic because signal magnitude and variability may depend on whether there are sinuses or large draining veins in a cortical region which are known to have large signals with high variability. Also, differences in signal-to-noise that are due to proximity to receiver coils or partial voluming effects may bias our measurements of sensitivity, particularly making comparisons across different areas problematic. In addition if variance is proportional to the mean as it is expected to be for single neurons or Poisson-like processes (Softky & Koch, 1993), then measures of population sensitivity would need to be scaled appropriately as response magnitude grows. We therefore examined the variability of response in each cortical visual area. First, we fit a canonical hemodynamic response function to all trials as described above. We then fit a general linear model using this canonical hemodynamic response and allowed each trial to have a separate beta weight. That is, we found the scale factor (beta weight) for every single trial which best fit the measured time course in the least squares sense, accounting for linear overlap across trials, for each observer for every cortical area. To avoid response variance associated with different stimulus strengths, we grouped the scale factors by condition (20 contrast and coherence; 20 duration) and computed the standard deviation. This results in 3,520 measurements of standard deviation (11 observers $\times$ 8 cortical areas $\times$ 40 conditions) each of which was computed from 25 trials. If the microvasculature, coil proximity, or partial voluming in different cortical areas resulted in differences in variability, or if contrast or coherence caused the variability to increase, we would expect that these measurements of standard deviation would consistently vary with those parameters. We tested for this by fitting a series of linear models in which the standard deviation depended on either an intercept alone, each conditions contrast, coherence, cortical area, or random effect of subject, and

all the effects together. We also tested models in which the contrast and coherence effects could differ by area. We performed model comparison by testing for improvement over the intercept-only model via partial F-test.

2.3 Results

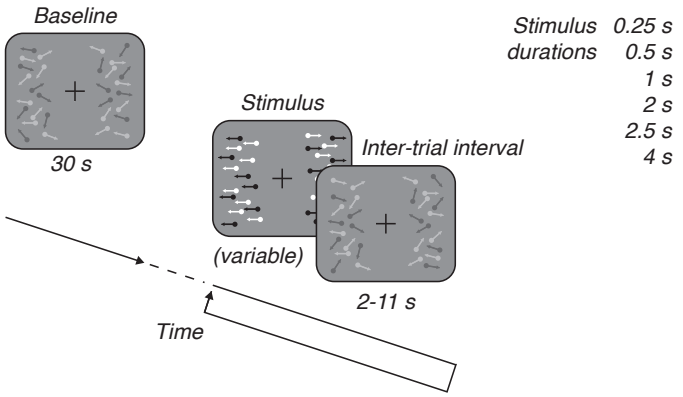

Figure 2.1: Cortical measurement experiment. Observers were shown patches of moving dots that increased in contrast and motion coherence on each trial. A 30-s baseline period preceded each scan with 25% contrast dots and incoherent motion and the baseline dots persisted between trials. On each trial the contrast increased by 0, 25, 50, or 75% and the coherence by 0, 25, 50, 75, or 100% for a stimulus duration of 250 to 4,000 ms. Observers performed an asynchronous task at fixation throughout the experiment.

2.3.1 Measuring cortical responses to contrast and motion coherence.

We characterized human cortical responses to changes in contrast and motion coherence of patches of dynamic random-dot stimuli by measuring BOLD responses while observers passively viewed two patches of moving dots (Fig. 2.1). Each scan began with 30 s of baseline stimulus presentation (0% coherence, 25% contrast) after which trials consisting of brief increments (0.254 s) in either or both coherence and contrast before returning back to baseline for a random length intertrial interval (211 s) (see methods for full details). In total observers were shown 40 conditions: 20 consisted of combinations of changes in contrast (+0, +25, +50, and +75%) and changes in motion coherence (+0, +25, +50, +75, and +100%) for 2,500 ms each, the remaining 20 were a subset of these combinations combined with variable stimulus durations (250, 500, 1,000, 2,000, and 4,000 ms). To minimize task-dependent effects and maintain a consistent level of engagement, observers performed an independent fixation task during viewing. We computed hemodynamic responses to each stimulus condition for each observer using an event-related analysis for retinotopically defined visual areas V1, V2, V3, hV4, V3A, V3B, V7, and MT. We begin by describing responses in visual

areas V1 (Fig. 2.2A) and MT (Fig. 2.2B), as they are well known to be sensitive to contrast (Avidan et al., 2002; Boynton et al., 1996; Gardner et al., 2005; Logothetis et al., 2001; Olman et al., 2004; Tootell et al., 1995; Tootell et al., 1998a) and motion coherence (Britten et al., 1993; Händel et al., 2007; Rees et al., 2000; Simoncelli & Heeger, 1998), respectively.

We observed clear parametric sensitivity to increases in contrast in V1 but weaker sensitivity in cortical area MT. Our measurements in V1 confirm previous results (Gardner et al., 2005; Logothetis et al., 2001; Tootell et al., 1995; Tootell et al., 1998a). The contrast sensitivity of V1 can be appreciated as monotonically increasing response magnitudes for higher levels of contrast increments (top left orange traces, Fig. 2.2A). These traces are for a stimulus duration of 2.5 s collapsing across motion coherence increments, i.e., averaging each row in the full response grid. While MT was also sensitive to increments of contrast, the monotonic increase appeared less pronounced compared with V1 (top left orange traces, Fig. 2.2B), consistent with other reports that have noted MT as having near maximal responses to small changes to contrast (Sclar et al., 1990; Tootell et al., 1995).

For motion coherence, we found the opposite pattern: MT was much more sensitive to increments in motion coherence compared with V1. MT showed clear monotonic increasing responses with increasing motion coherence (bottom right purple traces, Fig. 2.2B). These traces are again for a stimulus duration of 2.5 s averaged over contrast increments, i.e., collapsing each column in the full response grids. In V1 there was little difference in response amplitude as a function of motion coherence, i.e., weak sensitivity to coherence (bottom right purple traces Fig. 2.2A).

While V1 showed little parametric sensitivity to difference in coherence and MT little sensitivity to difference in contrast, both show a large response to the smallest increment of these parameters. This consistent trial-by-trial response, which we call the stimulus-onset response, appears unrelated to our parametric manipulations. For example, despite showing little sensitivity to different levels of coherence all of the responses for V1, including the one induced by the least change in coherence (+25%), induced a large response relative to the baseline (purple traces, Fig. 2.2A). Similarly, for MT and contrast as can be appreciated by noting that increasing contrast by 25% (orange traces, Fig. 2.2B) resulted in a large response. Part of this apparently large response is due to the fact that these responses for contrast or coherence are averaged over changes in the other parameter. That is, increases in contrast are shown averaged over coherence and vice versa. However this is not the complete story as can be appreciated by examining the grid of responses to each parameter separately (small bold black traces in grid, Fig. 2.2A and B). V1 can be seen to respond to a small change in coherence (+25, along horizontal) when there is no change (+0, along vertical) in contrast and vice versa for MT. These relatively large responses, to a feature each area is not strongly sensitive to, suggests that there is a response to stimulus onset regardless of condition.

Motion visibility is also adjusted by reducing the duration of stimuli, often in conjunction with reduced contrast and coherence. Along with the measurements described above, for which the stimulus duration was 2.5 s, we tested a large array of different durations from 0.25 to 4 s. As

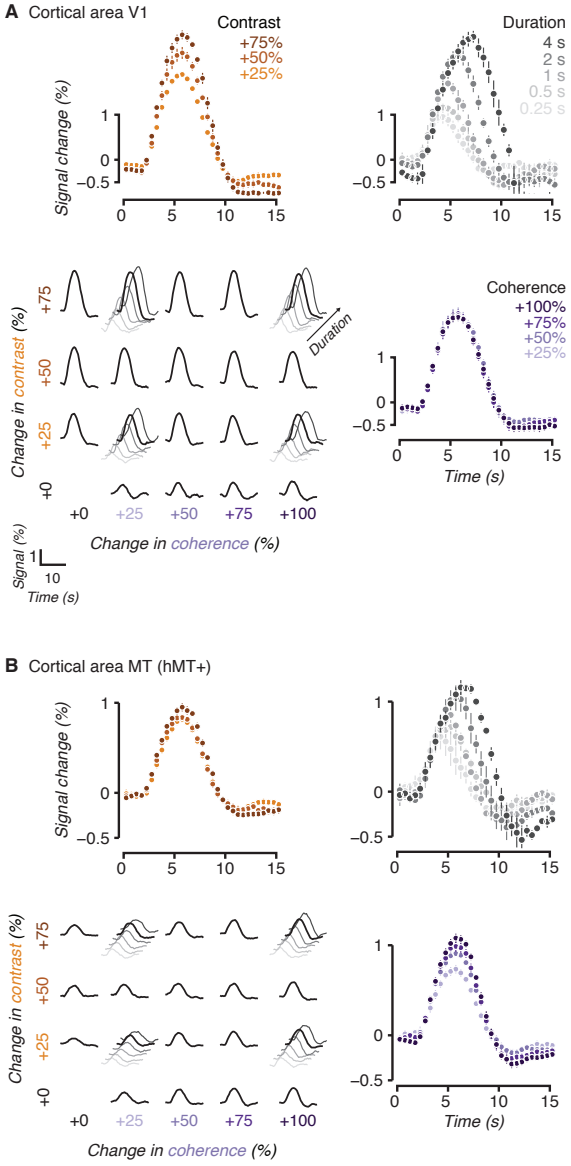

Figure 2.2: Measurements of event-related responses in cortical areas V1 and MT. A: cortical area V1. To obtain the individual responses shown here we performed an event-related analysis on our time series. In total we included 40 conditions in the experiment: 20 consisted of a full cross of changes in contrast and/or coherence presented for 2,500 ms (shown in bold in the grid in the bottom left) and 20 were a subset of the full cross conditions presented for various durations (shown in diagonal for the four conditions with additional durations recorded). We measured cortical responses to changes in contrast (top left) where each trace is averaged over changes in coherence, i.e., each response is the average of a row in the bottom left grid. We also measured responses to changes in coherence (bottom right), each trace is averaged over changes in contrast, i.e., each response is the average of a column in the grid. We made additional measurements across a large range of stimulus durations (top right) also shown in the grid. B: as in A for cortical area MT (hMT+). In all panels the event-related responses are averaged across observers and error bars indicate the bootstrapped 95% confidence interval; some error bars may be hidden. Note that for visualization event-related responses are only shown out to 15 s but the analysis used a window of 40.5 s.

expected of an approximately linear system (Boynton et al., 2012) we observed that responses scaled with stimulus duration in both cortical areas V1 and MT (top right gray traces, Fig. 2.2A,B).

Across the rest of the visual areas that we were able to retinotopically define in all subjects (V2, V3, hV4, V3A, V3B, and V7) we found similar parametric sensitivity to contrast, motion coherence and stimulus duration (Fig. 2.3). In general, and in concordance with previous reports (Avidan et al., 2002) we found less parametric sensitivity to changes in contrast for visual areas higher up in the visual hierarchy in the range we measured (+25 to +75% contrast). Sensitivity to coherence was observed in a number of the visual areas, although MT and to a lesser extent V3A were the clear stand-outs in showing monotonically increasing responses to this parameter. These observations will be quantified below.

2.3.2 Fitting population response functions to cortical responses.

To quantify the parametric sensitivity to contrast and coherence of each visual area we fit the event-related responses with a population response model using idealized functional forms for the relationship between contrast and coherence and neural response (Fig. 2.4). Based on previous work we expected that the population response to contrast would be a sigmoidal function (Albrecht & Hamilton, 1982; Sclar et al., 1990; Boynton et al., 1999) with the form of a Naka-Rushton equation (Fig. 2.4B, orange curve) (Naka & Rushton, 1966). To avoid overfitting, we fixed the exponents in the equation based on previous work (Boynton et al., 1999) and only allowed σ and α_{con} to vary. For motion coherence, we allowed for a functional form that can smoothly interpolate between linear (Britten et al., 1992, 1993; Simoncelli & Heeger, 1998; Rees et al., 2000) and a saturating exponential (Fig. 2.4B, purple curve). Finally, we included an onset term to capture the portion of response that did not vary across all conditions which presumably reflects stimulus onset and not parametric variation of stimulus parameters.

To predict the BOLD response from the modeled contrast and coherence response functions, we employed a linear-systems approach (Heeger et al., 2000; Rees et al., 2000; Logothetis et al., 2001). To account for different durations of stimuli, we multiplied the response magnitude predicted by the onset, contrast, and coherence functions with a boxcar function of appropriate length (Fig. 2.4C). As it is known that brief stimuli evoke response larger than expected by linearity (Boynton et al., 1996; Boynton et al., 2012), we also scaled the boxcar magnitude with an exponential that accounted for this nonlinearity in response. This scaled boxcar was then convolved with a hemodynamic response function (Fig. 2.4D) whose parameters were adjusted to best fit the event-related responses across all conditions (Fig. 2.4E). All together, we fit the model parameters for the contrast function (α_{con}, σ), coherence function (α_{coh}, κ) and temporal effects (δ, R_{onset}), and the parameters for the hemodynamic response function (t_0, τ_0, t_1, τ_1, α_1) for each observer for each visual area by minimizing the sum of least squares between the output of the model and the event-related responses for each of the 40 conditions.

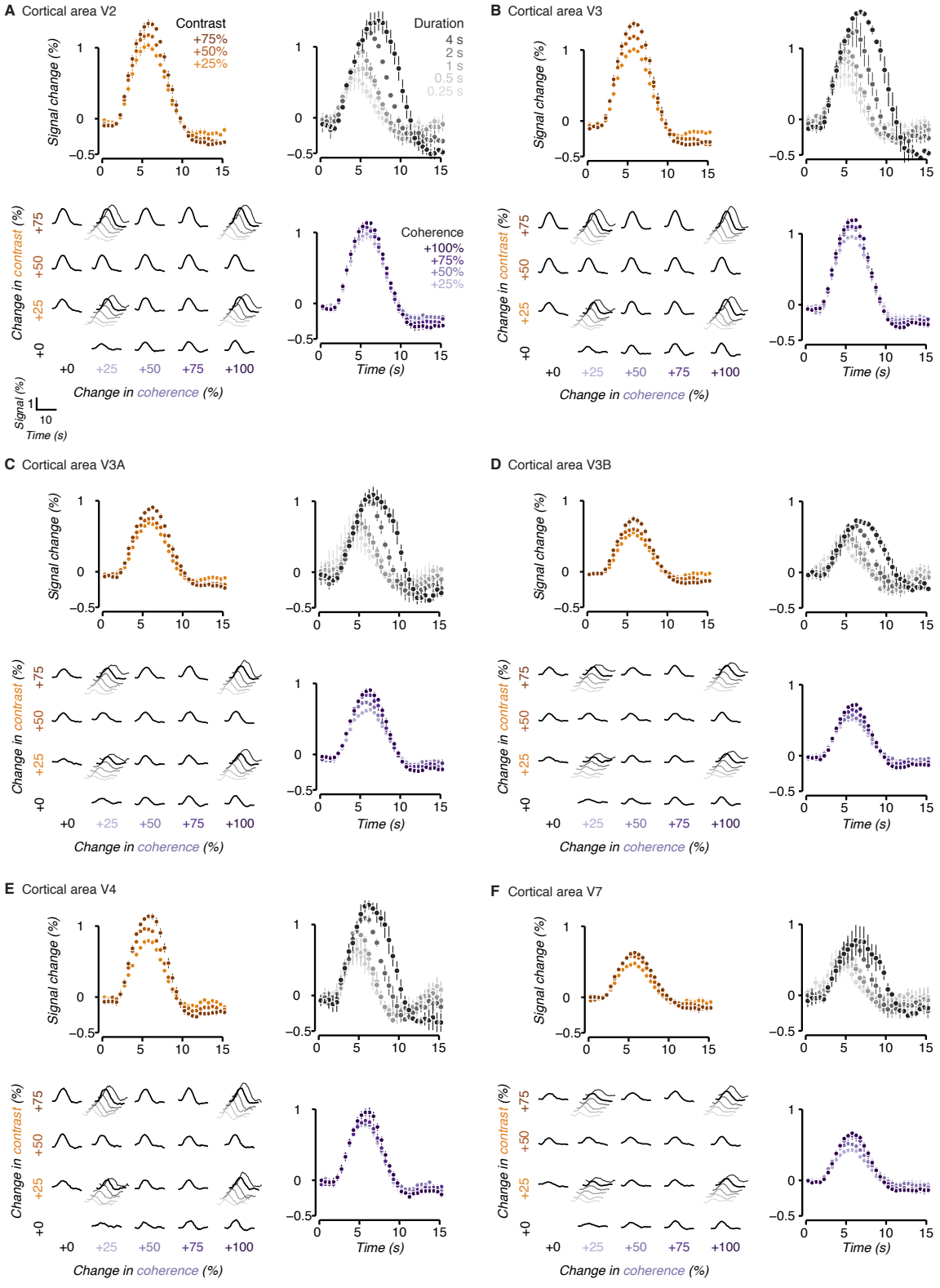

Figure 2.3: Measurements of event-related responses in cortical areas V2—V7. A—F: conventions are the same as in Fig. 2.2

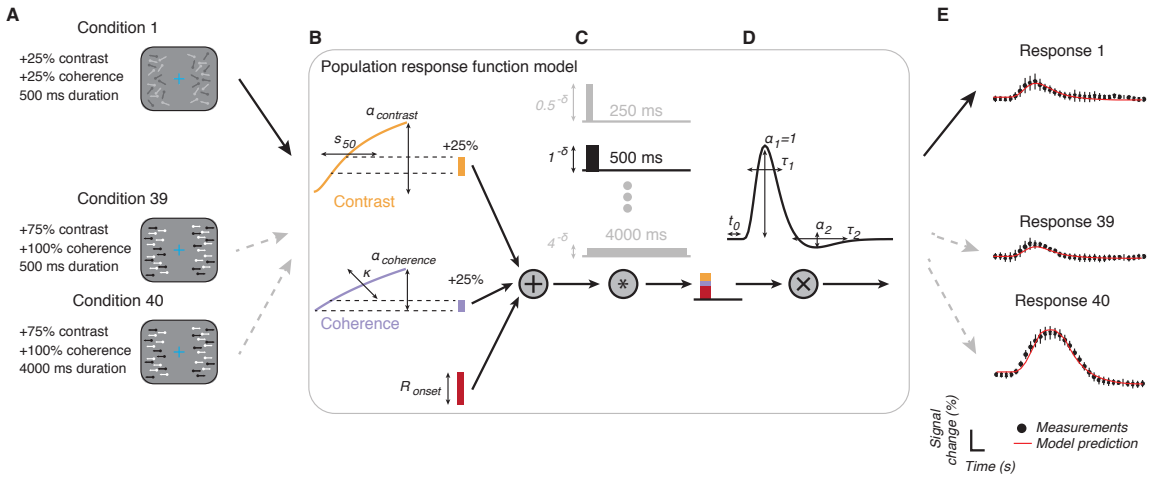

Figure 2.4: Population response function model. A: Each condition in the experiment was defined by three parameters: the increment in contrast above baseline (+0, +25, +50, or +75%), the increment in coherence above baseline (+0, +25, +50, +75, +100%), and the stimulus duration (250, 500, 1,000, 2,000, or 4,000 ms). As an example we use condition 1 to demonstrate the model. B: to estimate the response to each feature within a condition we first find the change in response due to the corresponding change in stimulus intensity according to the population response functions. For contrast the population response function is a Naka-Rushton with two free parameters: α_{con} controlling the amplitude and σ the shape. For coherence the response function was a saturating non-linearity with two free parameters: coherence controlling the amplitude and the shape. We added the resulting change in response together (while testing for interaction effects, see methods) and included an onset parameter to account for stimulus response that did not vary parametrically with the stimulus features. C: the total response, including onset, was used to scale a boxcar function whose length matched the stimulus duration. The boxcar was additionally scaled by a parameter to account for the nonlinear effect of stimulus duration. D: the resulting boxcar was convolved with a canonical hemodynamic response function fit separately for each observer. E: the model outputs a prediction for each condition about the expected event-related response (red lines). The parameters within the population response function model were then optimized to minimize the sum of squared errors between the data (black markers) and the model responses.

	Hemodynamic response			Contrast		Coherence		Onset
Area	tau_0	tau_1*	α_1	α_{con}*	σ*	α_{coh}	κ	R_{onset}*
V1	0.51 (0.05)	1.80 (0.42)	-0.24 (0.06)	1.68 (1.18)	0.35 (0.14) (p=0.006)	13.94 (32.04)	2.11 (6.79)	0.42 (0.24) (p<0.001)
V2	0.52 (0.06)	1.79 (0.44)	-0.24 (0.08)	0.69 (0.57)	0.40 (0.13)	21.27 (27.42)	0.40 (1.06)	0.28 (0.16)
V3	0.52 (0.07)	1.79 (0.59)	-0.24 (0.09)	0.63 (0.34)	0.43 (0.13)	13.23 (26.11)	-0.61 (3.13)	0.30 (0.12) (p=0.013)
V4	0.52 (0.07)	1.83 (0.93)	-0.24 (0.09)	0.61 (0.24)	0.47 (0.12)	14.05 (22.42)	-1.83 (5.98)	0.21 (0.13)
V3A	0.52 (0.07)	1.79 (0.59)	-0.24 (0.09)	0.35 (0.26)	0.48 (0.11)	3.41 (12.46)	1.05 (1.71)	0.15 (0.1) (p=0.005)
V3B	0.52 (0.07)	1.79 (0.59)	0.24 (0.09)	0.24 (0.13) (p=0.003)	0.43 (0.17)	7.99 (16.15)	0.29 (0.41)	0.14 (0.06) (p=0.002)
V7	0.52 (0.07)	2.15 (0.87) (p=0.032)	-0.23 (0.07)	0.32 (0.25) (p=0.022)	0.53 (0.22)	9.80 (14.21)	1.11 (1.87)	0.09 (0.07) (p<0.001)
MT	0.51 (0.04)	1.81 (0.63)	-0.22 (0.09)	0.22 (0.25) (p=0.001)	0.58 (0.32)	6.87 (17.60)	0.92 (1.07)	0.24 (0.09)

Table 2.1: Variability in parameter estimates across cortical areas

	Hemodynamic response			Onset
Observer	τ_0	τ_1	α_1	R_{onset}
1	0.53 (0.03)	2.12 (0.09) (p=0.041)	0.24 (0.01)	0.31 (0.11) (p=0.013)
2	0.50 (0.01)	2.09 (0.15) (p=0.020)	0.23 (0.02)	0.28 (0.13)
3	0.53 (0.03)	1.80 (0.36)	0.20 (0.03) (p<0.001)	0.06 (0.13) (p<0.001)
4	0.41 (0.03) (p<0.001)	1.29 (0.25) (p<0.001)	0.13 (0.03) (p<0.001)	0.22 (0.18)
5	0.45 (0.02) (p<0.001)	3.09 (0.53) (p<0.001)	0.10 (0.02) (p<0.001)	0.29 (0.13)
6	0.54 (0.03) (p=0.005)	2.41 (0.15) (p<0.001)	0.22 (0.02)	0.18 (0.11)
7	0.51 (0.03)	1.77 (0.12)	0.27 (0.04) (p<0.001)	0.12 (0.08) (p=0.001)
8	0.61 (0.05) (p<0.001)	1.47 (0.12) (p<0.001)	0.33 (0.04) (p<0.001)	0.33 (0.17) (p=0.004)
9	0.54 (0.03) (p=0.002)	2.05 (0.54) (p<0.001)	0.22 (0.03) (p=0.032)	0.32 (0.25) (p=0.009)
10	0.54 (0.03)	1.18 (0.24) (p<0.001)	0.34 (0.05) (p<0.001)	0.20 (0.15)
11	0.52 (0.02)	1.02 (0.14) (p<0.001)	0.33 (0.03) (p<0.001)	0.24 (0.11)

Table 2.2: Variability in hemodynamic response and onset parameter estimates across observers

We report the main fit parameters of the hemodynamic response function and population response function model across cortical areas (Table 2.1) and observers (Tables 2.2 and 2.3). We assessed whether between-observer variability existed by fitting a linear model predicting each parameter with observers as categorical predictors and used the same procedure to assess for within-observer variability across cortical areas (see methods). We found that there was statistically significant between-observer variability across all of the parameters but only significant variability within-observer (i.e., across cortical areas) for the shape parameter of the hemodynamic response τ, the magnitude and shape parameters of the contrast response function α_{con} and σ, the parameters of the coherence response function α_{coh} and κ, and the onset parameter R_{onset} (significance established by a partial F-test comparing linear regression models with and without each group of additional parameters at the $p = 0.05$ threshold). Note that the κ and α_{coh} parameters which together control both the shape and magnitude of the coherence response are hard to interpret in isolation.

	Contrast		Coherence	
Observer	α_{con}	σ	α_{coh}	κ
1	0.73 (0.46)	0.47 (0.14)	20.64 (19.35)	0.00 (0.00)
2	0.23 (0.37) (p<0.010)	0.32 (0.18) (p=0.004)	5.68 (18.19)	1.30 (1.85)
3	1.31 (1.65) (p<0.001)	0.50 (0.02)	1.77 (3.82)	2.04 (1.55)
4	0.56 (0.45)	0.26 (0.20) (p<0.001)	6.81 (15.88)	1.19 (5.85)
5	0.64 (0.92)	0.50 (0.07)	15.81 (22.26)	2.66 (5.01) (p=0.005)
6	0.66 (0.23)	0.47 (0.25)	-12.93 (11.24) (p<0.001)	3.82 (7.77) (p<0.002)
7	0.42 (0.18)	0.56 (0.20) (p=0.040)	3.88 (9.76)	0.56 (0.80)
8	0.40 (0.43)	0.44 (0.10)	34.68 (28.94) (p<0.001)	0.03 (0.29)
9	0.45 (0.32)	0.58 (0.18) (p=0.014)	12.60 (20.76)	0.11 (0.82)
10	0.52 (0.31)	0.39 (0.16)	7.12 (9.65)	0.77 (1.02)
11	0.61 (0.18)	0.56 (0.20) (p=0.016)	28.44 (28.99) (p<0.001)	0.02 (0.02)

Table 2.3: Variability in population response function parameter estimates across observers

The population response model was able to capture the majority of variance in each observers event-related responses and a significant portion of this explained variance was accounted for by the population response functions. We assessed variance explained as the squared correlation between the model predictions and the actual event-related responses for held-out conditions. For V1, r^2=0.69, 95% CI [0.63 0.75]; V2, r^2=0.63, 95% CI [0.58 0.68]; V3, r^2=0.62, 95% CI [0.56 0.68]; hV4, r^2=0.44, 95% CI [0.35 0.53]; V3A, r^2=0.42, 95% CI [0.35 0.50]; V3B, r^2=0.38, 95% CI [0.31 0.46]; V7, r^2=0.32, 95% CI [0.24 0.40]; MT, r^2=0.49, 95% CI [0.43 0.56]. Part of the variance accounted for by the model is simply due to the stimulus-onset term and hemodynamic response, but the population response functions also captured significant variance. We assessed this by comparing our results to a model fit to the same measurements but where the condition labels were permuted. This corresponds to keeping the variance explained by stimulus onset and the hemodynamic response but randomizes the relationship between condition and response. We repeated this permutation test procedure 100 times per observer and cortical area. On average across observers and areas the variance explained by fitting to the measured data set (average cross-validated r^2=0.508) exceeded the variance explained in the permuted data set (average cross-validated r^2=0.340) with $p < 0.001$, Δr^2=0.164, 95% CI [0.162, 0.165].

Across cortical visual areas the model captured the response to changes in contrast and motion coherence as well as the amplitude effects due to duration. To visualize the fit of the population model to each variable we scaled the canonical hemodynamic response function for each observer to fit the event-related responses in the conditions with either no change in contrast or no change in coherence. This results in a single scaling factor for each of these conditions (circles, Fig. 2.5) which we compared with the model predictions (lines, Fig. 2.5). Examination of the magnitude of the population model fit to the event-related response peaks for changes in contrast (orange

curves, Fig. 2.5A) and coherence (blue curves) shows good correspondence. This is particularly notable given that the model is fit across all conditions containing different response lengths, as well as combinations of contrast and coherence changes, while the displayed data are for changes in contrast and coherence in isolation. This visualization displays a model fit to all the data, i.e., not on held-out data, but with similar explained variance to the cross-validated model (difference between cross-validated and full fit, Δr^2 ==0.005, 95% CI [0.004 0.006]). The population response functions echoed the qualitative results described above for the event-related responses: V1, V2, V3, and hV4 showed strong response to contrast with relatively weak response to coherence. Only MT showed stronger response to motion coherence than to contrast. Moreover, the amplitude of responses as a function of duration (Fig. 2.5B) were similarly well captured by the population response model. As noted earlier the amplitude of responses due to doubling in duration do not appear to scale in a linear manner.

The form of the contrast response function has been extensively studied (Albrecht & Hamilton, 1982; Boynton et al., 1999; Sclar et al., 1990) while the motion coherence response function has received much less attention. Single-unit studies have found a linear response function, whereas BOLD measurements in humans have found some non-linearity of response, particularly outside of MT (Rees et al., 2000). We therefore tested for non-linearity in the population response functions to motion coherence and found that responses were generally best characterized as linear, with a small deviation from linearity for MT. We quantified this comparison as the difference in cross-validated variance explained between the saturating exponential and a linear form for the coherence response function. In MT we found a small difference in favor of the nonlinear model r^2=0.004 (95% CI [0.001 0.007]) while all other cortical areas confidence intervals overlapped with zero. This difference is visible as the saturation of the MT coherence response to large changes in coherence (Fig. 2.2B and Fig. 2.5A, MT).

While population responses to each motion feature could interact, i.e., a change in contrast might influence the response to a change in coherence or vice versa, we found no evidence for this. We tested for interactions by adding an additional beta weight to the model accounting for the effect of multiplicative changes in contrast and coherence (see methods section Population responses: functional forms). Including this term reduced the cross-validated variance explained by on average 6.67% (95% CI [13.42, 0.08]) across cortical areas, suggesting overfitting compared with the no-interaction model. One observers data was particularly strongly overfit. Removing that observer resulted in an average reduction in variance explained of 0.08% (95% CI [0.25 0.09]) and for individual areas, V1: 0.18%, V2: 0.16, V3: 0.17, hV4: 0.13, V3A: 0.07, V3B: 0.14, V7: 0.07, MT: 0.11.

Visual inspection of the response grids (bottom left panels in Fig. 2.2 and 2.3) suggest an alternative kind of interaction in which the response to contrast and coherence might be stronger in the absence of the other feature changing. Take for example the response to contrast compared with

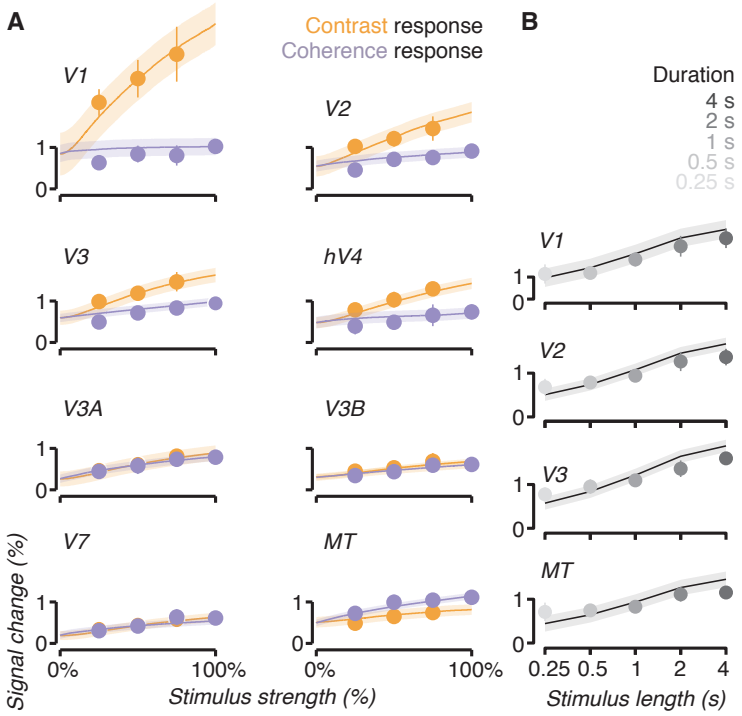

Figure 2.5: Population response functions. A: the population response functions fit to each cortical area V1-MT (hMT+) are shown compared with the magnitude of the event-related response for the conditions in which only one feature changed. These correspond to the conditions in the first column and last row of each event-related response grid in Figs. 2 and 3. To make the functions comparable to the data in an easy to interpret space we reduced each event-related response to a single magnitude value which was obtained by finding the linear scaling of the canonical hemodynamic response to that condition. The model outputs predictions for all 40 conditions but we are only showing the subset where either contrast or coherence changed alone. Note that the predictions here are not out of sample (i.e., these are not the cross-validation results) but we show the full fit to better visualize the response functions. B: as in A but for the variable duration conditions in which contrast and coherence changed maximally (+75% contrast, +100% coherence). In all plots markers indicate the average across observers and error bars the bootstrapped 95% confidence interval.

coherence in V1. The contrast response in V1 is so much larger than the response to coherence that its possible it "washes out" any visible effect due to coherence. To test for this possibility we fit a model with different population response functions for conditions in which only a single feature changed vs. when both features changed. We found that these models were also not statistically better than the simplest model with no interactions: average reduction in cross-validated variance explained 5.34% 95% CI [9.13, 1.56] and without the overfit observer 0.20% 95% CI [0.31, 0.08]. Although statistically the models were similar in our data set we did find that in the interaction model the population response functions to contrast had a higher maximal response when the coherence was not simultaneously changed, but the reverse was not true. On average across subjects and cortical areas we found an increase in sensitivity of 50% in the contrast response when no simultaneous change in coherence occurred (average parameter change 1.68 95% CI [0.58, 2.78], significantly different from zero as assessed by bootstrap over observers, P = 0.007).

The population response model fits (Fig. 2.5) replicate earlier reports showing that contrast responses have a smaller dynamic range and saturate more quickly in higher visual cortical areas (Avidan et al., 2002), and add the finding that coherence sensitivity peaks in MT. To assess this we plotted the maximum of the contrast response function (the contrast parameter) against the linear slope of the coherence response function (the response range measured as the slope of a linear fit, see methods) for each cortical area (Fig. 2.6A). As expected we found stronger sensitivity to motion coherence in V3A and MT compared with area V1 (Dupont et al., 1994; Tootell et al., 1998a; Watson et al., 1993; Zeki et al., 1991). The difference in coherence sensitivity between V3A and V1 was 0.167, P ¡ 0.001 and between MT and V1 0.251, P ¡ 0.001. But we also observed significant sensitivity to changes in coherence in all regions measured (Fig. 2.6B): V1 = 0.12, V2 = 0.19, V3 = 0.18, hV4 = 0.13, V3A = 0.25, V3B = 0.15, V7 = 0.22, MT = 0.36, slopes in % signal change/unit coherence, all P ¡ 0.001 assessed by bootstrap across observers. All cortical visual areas showed statistically significant parametric sensitivity to changes in contrast (Fig. 6C) assessed as a nonzero α_{con} parameter by bootstrap across observers, all P ¡ 0.001 except MT, P = 0.002. The maximum contrast response dropped quickly for regions higher in the visual hierarchy (V1 = 2.00, V2 = 0.87, V3 = 0.68, hV4 = 0.63, V3A = 0.35, V3B = 0.24, V7 = 0.33, MT = 0.20, units in % signal change/unit contrast).

Although we fit a Naka-Rushton function to the contrast response our measurements were limited to only a few points (no change in contrast, +25, +50, and +75%). This meant that the data did not strongly constrain a sigmoidal fit. We assessed whether in our data set the results would be equally well fit by a linear model and found that this was the case for all areas except V7, with an average improvement of 0.32% in cross-validated variance explained. Therefore, the contrast parameter which fits the maximal response to contrast in each region tracks the slope of the relationship between contrast and response and can therefore be used as a measure of the sensitivity to contrast, in the range of contrasts we measured. The linear models improvement in variance explained for

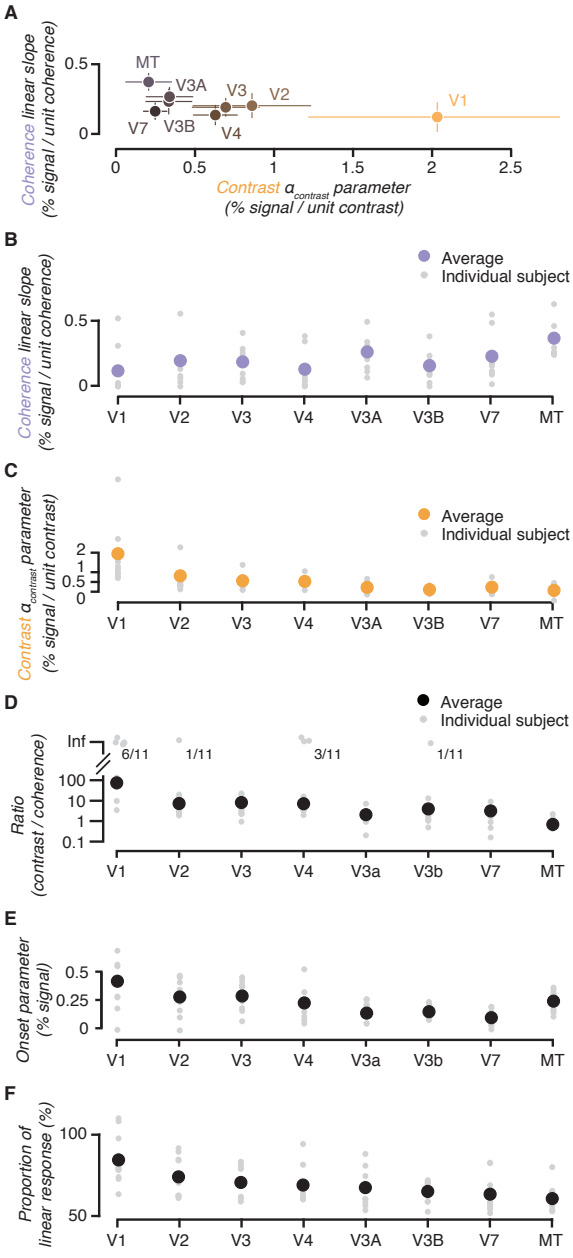

Figure 2.6: Cortical sensitivity to contrast and motion coherence. A: to obtain a qualitative estimate of cortical sensitivity to each motion visibility feature across the cortical visual areas we plotted the contrast parameter from the Naka-Rushton function against the slope of a linear fit of the coherence functions. B: the slope of the coherence functions fits as in A replotted with individual subjects. C: the contrast parameter as in A replotted with individual subjects shown for each cortical area. D: we plot the ratio of the sensitivity parameters as an unbiased additional comparison because the amplitude parameters could be sensitive to the signal-to-noise ratio of the measurement in different cortical areas. Note that for some subjects the slope of the coherence response was near zero in some cortical areas, we note these as a ratio of infinity (Inf). The means are calculated excluding infinite values. E: the stimulus-onset response parameter R_{onset} indexes the portion of the response that was not parametrically modulated by contrast or coherence. F: for each doubling in stimulus duration the proportion of response increase is shown by cortical area where 100% would indicate that responses increased linearly with duration. In all panels markers indicate the mean and error bars the bootstrapped 95% confidence interval. Error bars are omitted in panels (BE) for visualization.

individual areas were V1 0.66, 95% CI [0.16 1.15]; V2 0.79, 95% CI [0.14 1.45]; V3 0.51, 95% CI [0.04 0.98]; V4 0.27, 95% CI [0.09 0.45]; V3a 0.24, 95% CI [0.06 0.42]; V3b 0.17, 95% CI [0.03 0.31]; V7 0.04, 95% CI [0.14 0.06]; MT 0.21, 95% CI [0.07 0.34].

We found that the variability in our measurements did not differ significantly across different cortical areas or according to the stimulus strength. We performed this analysis to test whether various nuisance variables could have altered our measurements, e.g., proximity to the coils and partial voluming might affect signal-to-noise in different cortical areas, or the variability in our measurements might increase with response magnitude as contrast and coherence cause populations of neurons to be more active. To do this we estimated the response magnitude of every trial and grouped these by condition and cortical area, then fit a series of linear models to see whether variability differed. We found that none of the additional variables improved the model fit over the intercept-only model at the $p = 0.05$ significance threshold. Importantly, the model that allowed separate values for each cortical area did not improve the model fit, suggesting that response variability did not significantly differ between cortical areas (mean cortical area standard deviation was 1.50 percent signal change; V1 1.75, 95% CI [1.33 2.17]; V2 1.08, 95% CI [0.91 1.25]; V3 1.28, 95% CI [1.06 1.49]; V4 1.37, 95% CI [1.03 1.72]; V3a 5.71, 95% CI [1.28 10.15]; V3b 1.14, 95% CI [0.96 1.31]; V7 1.32, 95% CI [0.97 1.66]; MT 1.30, 95% CI [0.95 1.66]). In addition we found that there was no statistically significant change in variability in the slope of the relationship between variability and stimulus strength (even when separate slopes were allowed for different cortical areas), suggesting that noise in our measurements was additive, i.e., did not increase with increasing response magnitude. Fitting the model with a slope for contrast and coherence (shared across areas) results in a slope of 1.23 percent signal change per unit contrast, $t(2,557) = 1.27$, $P = 0.21$, and a slope of 0.58 percent signal change per unit coherence, $t(2,557) = 0.78$, $P = 0.44$.

Although our measurements do not suggest that any bias is introduced by potential signal-to-noise differences across areas, we computed the ratio of the contrast and coherence slope parameters as an additional unbiased analysis (Fig. 6D). This ratio allows for between region comparison of the sensitivity to contrast and coherence because the ratio reports how sensitive each region is to contrast compared with coherence and not overall sensitivity. That is, the ratio should be invariant to differences in signal-to-noise, under the assumption that contrast and coherence sensitivity are equally affected. In line with our previous results we found that V1 has a ratio of contrast to coherence sensitivity that is at least an order of magnitude more than the other areas. In addition MT was found to have a ratio near 1 and lower than the other cortical areas, reflecting its stronger relative sensitivity to coherence.

We found that the portion of the BOLD response that did not vary parametrically with contrast or coherence, the stimulus-onset response Ronset did vary across cortical areas (Fig. 6E and Table 1). On average the onset response was 0.23 percent signal change across observers and cortical areas. The stimulus-onset response in V1 and V3 were larger than average at 0.42 percent signal change,

95% CI [0.28, 0.55], while areas V3A, V3B, and V7 were smaller than average, 0.14, 95% CI [0.07, 0.20]; 0.15, 95% CI 0.11, 0.18]; 0.09, 95% CI [0.06, 0.13], respectively. The other cortical areas onset effects were V2 0.28, 95% CI [0.19, 0.37]; V3 0.22, 95% CI [0.22 0.36]; V4 0.23, 95% CI [0.15, 0.31]; MT 0.24, 95% CI [0.19, 0.29].

Finally, we found that the effect of increasing stimulus duration was not consistent across cortical areas (Fig. 2.6F). We found that early visual cortex, V1 in particular, was significantly more sensitive to changes in duration than later visual areas, especially MT. The effect of a doubling in duration on the population response, as a proportion of that expected from a linear model, was 68.56%. On average across subjects we found that V1 and MT differed significantly from the average. We found that the effect of a doubling in duration in V1 was 83% of the linear model, 95% CI [76.00, 92.27], suggesting that V1 is more sensitive to stimulus duration. By contrast in MT the effect was only 62% of the linear model, 95% CI [57.06, 66.08], suggesting that MT may have a more transient response. The effects in other areas were not significantly different from the average: V2 73%, 95% CI [67.20, 79.70]; V3 70%, 95% CI [65.05, 74.95]; V4 70%, 95% CI [64.21, 76.59]; V3A 67%, 95% CI [61.24, 73.31]; V3B 64%, 95% CI [60.45, 67.64]; V7 64%, 95% CI [58.07, 70.02].

2.4 Discussion

We have developed a quantitative framework for modeling human cortical response to motion visibility as parameterized by image contrast, motion coherence, and duration. Our results provide a comprehensive view of the variability in cortical sensitivity to these features, each of which is a critical component of visual stimuli often manipulated in experiments designed to understand visual perception and decision-making. Our measurements show that the range of responses to different levels of contrast was larger in early visual cortex, especially V1, and the range of responses for coherence larger in V3A and MT (hMT+). Nonetheless, a change in either feature caused a cortical response in all the retinotopic areas we mapped. Our results weigh on various other findings in the literature: the precise shape of population response functions, the influence of stimulus duration on cortical signals, and whether or not sensory representations for different features interact. Finally, we believe that this parameterized model, and parametric models in general, suggest mechanisms for the read out of sensory representations from population responses and have therefore made our measurements and framework available online as a resource (see methods).

We studied changes in contrast, coherence, and duration to measure human cortical response within a range where typical human perceptual experiments are performed. One choice we made was to measure contrast from a relatively high baseline. Because the contrast response function is known to adapt to the current background stimulus without altering the form of parametric modulation (Ohzawa, Sclar, & Freeman, 1982, 1985; Sclar, Ohzawa, & Freeman, 1985; Sclar, Lennie, & DePriest, 1989; Gardner et al., 2005) the relative sensitivities we measured should hold at other baselines. With

this design we were also able to show that sensitivity to changes in contrast and coherence do not interact. The interaction analysis would be impossible in stimuli where the dots appear from a black or gray background such that both contrast and coherence always change together (Britten et al., 1993; Rees et al., 2000). When designing the dot motion stimulus we also had to ensure that there were sufficient dots and a large enough aperture to be clearly visible and generate a reliable coherence response. At low dot densities the response to changes in coherence are negligible (Smith, Wall, Williams, & Singh, 2006) and small aperture sizes can cause changes in coherence to result in decrements in response (Ajina et al., 2015; Becker, Erb, & Haarmeier, 2008; Costagli et al., 2014). By creating a large stimulus with high density we guaranteed that our dot motion would blanket the population receptive fields of all the cortical areas measured.

We set our stimulus to move at a constant rate of 6°/s, within the peak range of speed tuning in visual cortex, and used a dot stimulus rather than gratings to avoid having spatial frequency tuning affect our measurements. Although individual V1 and MT neurons in the macaque differ greatly in their speed tuning the average tuning of the population is quite similar and centered near 6°/s with ranges that extend far above and below that (Priebe, Lisberger, & Movshon, 2006). Measurements of speed tuning in humans evidence broad variability across all of visual cortex but our chosen speed is within the peak range (Singh, Smith, & Greenlee, 2000; Hammett, Smith, Wall, & Larsson, 2013). One common concern with speed tuning in gratings is that spatial frequency tuning differs across cortex and directly impacts sensitivity to other stimulus properties, such as image contrast (Priebe, Cassanello, & Lisberger, 2003; Priebe et al., 2006). We used a random dot stimulus with a wide range of spatial frequency components rather than gratings with a specific spatial frequency to avoid this confound. In principle our stimulus drives neurons with a wide range of tunings and by averaging over voxels in each cortical area we reduce the impact of columnar and other local microstructure in each area (Sun et al., 2007; Liu & Newsome, 2002).

We reported here several parameters which together defined the population response functions, but which of these represents a good measure of the sensitivity of a region? We use the term sensitivity to capture parametric differences in response magnitude with differences in contrast or coherence. Thus, an area with high contrast or coherence sensitivity is one in which the response to the lowest and highest values of these parameters evoke the largest difference in response (see methods for how the reported parameters correspond to this). This measure can be used to compare with human behavioral contrast or coherence discrimination performance since signal detection theory predicts that perceptual sensitivity, d, is directly proportional to this difference (Boynton et al., 1999; Newsome et al., 1989; Pestilli et al., 2011; Tolhurst, Movshon, & Dean, 1983; Zenger-Landolt & Heeger, 2003). However, d is also inversely proportional to the standard deviation of response which could vary across different areas, particularly for measurement related reasons that would therefore distort our measures of sensitivity. Our analysis of the variability of response across different areas did not find differences, thus suggesting that our measures are an accurate reflection of contrast and coherence

sensitivity. Moreover, we used a selection criterion to analyze a subset of voxels that show consistent trial-to-trial responses to reduce the effect of measurement noise but our parametrization will still be sensitive to any noise that remains.

Response variability might also change with response amplitude as it is known to do for single-unit responses. Although occasionally single neurons can be found that match perception (Britten et al., 1992), groups of neurons (Tolhurst et al., 1983) or larger populations (Averbeck, Latham, & Pouget, 2006; Zohary, Shadlen, & Newsome, 1994), depending on the correlation structure in the population, are likely to more closely reflect perceptual reports. Supporting the idea that populations are used for perceptual readout is evidence from human work where at the coarse resolution of the BOLD signal, which pools over large numbers of neurons, cortical measurements closely track perception under an assumption of additive noise (Boynton et al., 1999; Hara & Gardner, 2014; Pestilli et al., 2011; Sapir, d'Avossa, McAvoy, Shulman, & Corbetta, 2005). In line with this the variance of population responses measured with voltage-sensitive dyes do not change with magnitude of response in V1, i.e., they are additive (Chen, Geisler, & Seidemann, 2006). Our own measurements support the hypothesis that populations are subject to additive noise: we found that as contrast and coherence increased and caused larger magnitudes of response we found no evidence that trial-by-trial variability changed. Together our data and previous results suggest that measures of the slope in the BOLD signal population response function are indeed measures of sensitivity and leaves us with a testable prediction: if parameters measure sensitivity (i.e., signal-to-noise ratio) then they should be relatable to human perception under additive noise but not noise which scales with response magnitude.

We observed a saturation of the cortical response to motion coherence that differs from recordings of a linear response in MT in human (Händel et al., 2007; Rees et al., 2000) and monkey (Britten et al., 1993). Saturation of the contrast response function is thought to be the result of normalization, a canonical computation in cortex (Baker & Wade, 2017; Carandini & Heeger, 2011). If the response to motion coherence is linear, it might suggest that similar normalization does not apply. In fact, models of the V1 to MT circuitry include explicit normalization (Simoncelli & Heeger, 1998) and the normalization strength alters whether the model predicts linear or saturating responses. This may account for the discrepancies of results; i.e., normalization may result in weak saturation of coherence response as we have found, in line with evidence from both humans (Costagli et al., 2014; Rees et al., 2000) and monkeys (Britten et al., 1993). In support of this idea is evidence that in the absence of a normal input from V1 the coherence response function in MT becomes more linear, possibly reflecting an increased input from subcortical regions whose coherence response is linear (Ajina et al., 2015). To clarify this we can again turn to behavior. Because the MT response has been linked to behavior (Katz et al., 2016) our model makes a testable prediction: under the assumption that the visual system performs signal detection subject to additive noise (Boynton et al., 1999) a saturating coherence function would predict worse discriminability of coherence at higher base levels

of coherence.

To build out our quantitative framework we measured responses to stimuli of varying durations, down to those typically used in psychophysical experiments (e.g., 0.25 s) as well as at durations more typically used for BOLD measurement (e.g., 4 s). Our results confirm many previous results showing that there exists a nonlinearity in the BOLD response, such that shorter stimuli have a larger response than expected by temporal linearity (Boynton et al., 1996; Boynton et al., 2012). Modeling our responses, we found that on average across cortical areas a doubling of the stimulus duration was associated with an increase in response of only 67% of the expectation of a linear model. Whether or not this is due to neural adaptation (Buxton et al., 2004) or saturation of the BOLD signal (Friston et al., 1998) cannot be determined from our data. We also observed a slight difference in the duration effect across cortical areas. In V1 increasing duration results in a larger effect on the population response whereas MT showed a smaller than average response, which could be a result of the more transient response in MT (Stigliani et al., 2017).

We also noted that any change in the stimulus, regardless of the type, amplitude, or duration resulted in what we refer to as a stimulus-onset response in all cortical areas. What is the nature of this response? Early recordings comparing BOLD responses to electrophysiological recording suggest that the BOLD signal may be thresholded at some minimum response even though neural activity continues to be modulated below that threshold (Logothetis et al., 2001). Another possibility is that the stimulus-onset response may be the result of a consistent trial structure causing anticipatory responses (Cardoso, Sirotin, Lima, Glushenkova, & Das, 2012). In the latter case fitting a separate stimulus onset parameter to absorb this trial-structure related variance is appropriate to correctly estimate the population response from the BOLD signal.

Our approach to making a parametric model of cortical response to motion visibility contrasts with more complex models, such as Gabor wavelet pyramids and deep convolutional networks (Kay, Naselaris, Prenger, & Gallant, 2008; Kay & Yeatman, 2017; Yamins et al., 2014) that are typically image-computable and thus can make detailed predictions of cortical response properties directly from images. A complete image-computable model would implicitly contain our parametric model within it and seemingly obviate the need to parameterize stimulus visibility and its relationship to cortical response. Building such complex models is a worthy goal, however, we would note that much success in understanding visual cortex function has come from experiments which parametrically altered visual features, in particular features related to visibility. Consider the result of stimulus combination. When two gratings with different luminance contrast are presented the evoked response is not well captured by simple rules such as linear summation or winner-take-all (Busse, Wade, & Carandini, 2009). Instead, across a large range of parameter combinations the evoked response is well explained by normalization (Carandini & Heeger, 2011). The canonical rule that an evoked response should be scaled by the response of a neighboring region of cortex is easily understood in a parametric model, but far less intuitive in a complex one. Another low-dimensional parametric model

is the population receptive field (Dumoulin & Wandell, 2008; Wandell & Winawer, 2015), which has been widely used to map and interpret the properties of retinotopic visual cortex, largely because of its simplicity. In general, low-dimensional quantitative frameworks like the one we have built can parameterize cortical response to key stimulus properties and by doing so, serve to make testable predictions for perceptual function. For example, our framework suggests that small variation in sensitivity across cortical areas might be used to separately determine the visibility of motion for different parameters. That is, a read-out of the visual representations could take advantage of the differences in feature sensitivity by differentially weighting V1 and MT for contrast discrimination and vice versa for coherence discrimination.

Each parameter of motion visibility that we studied has been separately used to uncover the neural basis of different aspects of perception and perceptual decision making. The quantitative framework that we have proposed here shows that despite their similar effects on perception, contrast, coherence, and duration have distinct cortical representation at the level of populations. In studies of perception, the effects of these parameters on cortical response should not be considered to be interchangeable. With our reference framework one can now make changes in one parameter or the other and predict how this will affect human cortical response. In this way our predictive model is a key tool in furthering the goal of linking cortical response to perceptual behavior.

Chapter 3

Aim 1: A flexible readout mechanism of human sensory representations

3.1 Introduction

Humans can flexibly attend to different aspects of the environment when their goals require it. This can be operationalized by asking human observers to report about one feature of a visual stimulus while ignoring other features. Such context-dependent judgments could be supported by a cortical implementation which increases sensitivity or selectivity for the sensory representations of reported features while suppressing others. A second and potentially complimentary implementation is to maintain stable sensory representations while flexibly changing the downstream readout of these.

A great deal of evidence exists for the former possibility of changing representations to accommodate behavioral demands. Behavioral manipulations of spatial attention (Klein et al., 2014; Mitchell et al., 2009; Pestilli et al., 2011; Womelsdorf et al., 2006), feature-based attention (Baldauf & Desimone, 2014; Harel et al., 2014; Huk & Heeger, 2000; Jehee et al., 2011; Serences & Boynton, 2007; Cohen & Tong, 2013; Treue & Martínez Trujillo, 1999), and stimulus expectations (Kok, Jehee, & de Lange, 2012; Kok, Brouwer, van Gerven, & de Lange, 2013) all have been associated with changes in sensory representations. These changes may occur very early in the visual hierarchy (Ling, Pratte, & Tong, 2015) and take the form of changes in sensitivity (Reynolds et al., 2000; Serences & Boynton, 2007; Snyder et al., 2018; Treue & Martínez Trujillo, 1999), shifts in feature selectivity (Çukur et al., 2013; David et al., 2008; Kastner et al., 1998; Klein et al., 2014; Spitzer et al., 1988; Womelsdorf et al., 2006; Womelsdorf et al., 2008), increases in baseline response (Buracas & Boynton, 2007; Chen, Palmer, & Seidemann, 2012; Kastner et al., 1999; Ress et al., 2000; Li

et al., 2008; Murray, 2008) useful for efficient selection (Hara et al., 2014; Pestilli et al., 2011), and changes in the structure of stimulus-driven and noise correlations (Cohen & Maunsell, 2010, 2011; Mitchell et al., 2009; Ruff & Cohen, 2016; Verhoef & Maunsell, 2017).

However, flexible readout rather than change in sensory representation can be a behaviorally advantageous implementation of task demands. Although changing sensory representations can be beneficial, there cans be associated behavioral costs to suppressing ignored features in sensory representations (Gazzaley, Cooney, McEvoy, Knight, & D'esposito, 2005; Mesgarani & Chang, 2012; Rees et al., 1997) when these are actually relevant to behavior. In many dramatic demonstrations (Haines, 1991; Mack & Rock, 1998; Neisser, 1979; Simons & Chabris, 1999) observers have been made blind to salient events when reporting about other aspects of a visual scene. This suggests a potential advantage to maintaining stable sensory representations and using flexible sensory readouts to enable adaptable behavior (Bugatus, Weiner, & Grill-Spector, 2017; Mante, Sussillo, Shenoy, & Newsome, 2013; Peelen et al., 2009).

Finding changes in sensory representations across different task conditions is not enough to demonstrate that these changes are large enough to explain perceptual behavior. Instead, linking models are needed. Quantitative linking models (Barlow, 1972; Brindley, 1970; Cohen & Maunsell, 2010; Cook & Maunsell, 2002; Newsome et al., 1989; Pestilli et al., 2011; Teller, 1984; Hara & Gardner, 2014) connect measurements of cortical activity to behavior by modeling the presumed process by which sensory activity gives rise to perceptual behavior. Such linking models are explicit hypotheses that can be falsified if they are unable to quantitatively link sensory representational changes to behavioral performance across different task conditions.

Here we used a linking model to study human reports of motion visibility and to understand whether sensory change or flexible readout implement this behavior. We first established that observers could independently report about either the contrast (luminance difference between dark and bright dots) or motion coherence (percentage of dots moving in a coherent direction) of random dot patches while ignoring the other feature. We then extended a well-established linking model of human contrast perception (Boynton et al., 1999; Foley & Legge, 1981; Gardner, 2015; Ling & Carrasco, 2006; Nachmias & Sansbury, 1974; Pestilli, Ling, & Carrasco, 2009; Pestilli et al., 2011) to account for behavioral performance during these tasks. Because in individual cortical areas the response to motion visibility is mixed, we allowed the model to weight retinotopic areas according to their sensitivity to the two features. The critical step to understand behavioral flexibility was to measure BOLD signal while observers performed each discrimination task. If sensory representations changed enough, then a linking model with a fixed readout of sensory areas should account for behavior (i.e. that used the same weighting of cortical responses for both tasks). Implementing such a fixed readout model showed that sensory changes alone were insufficient in magnitude to explain perception. Instead, in addition to the sensory change, a change in readout between different task conditions was necessary (i.e. a flexible readout). A benefit of flexible readouts is that sensory

representations can retain information about the unattended feature. In line with this, we show that observers can re-map their reports unexpectedly.

3.2 Methods

3.2.1 Observers

In total 29 observers were subjects for the experiments. All observers except one (who was an author) were naive to the intent of the experiments. Eight observers were excluded during initial training sessions due to inability to maintain appropriate fixation (see eye-tracking below). All of the remaining 21 observers (13 female, 8 male; mean age 28 y; age range 18-55) performed the motion visibility behavioral experiment outside of the scanner. Observers performed up to six one-hour sessions on separate days for an average of 2467 trials each (range 1167-3652, standard deviation 497). Ten of the observers (7 female, 3 male; mean age 26 y; age range 19-36) repeated the motion visibility experiment inside the scanner. Observers were scanned in two 90-minute sessions, each consisting of eight 7-minute runs, and a third one-hour scan which included retinotopy and anatomical images. Procedures were approved in advance by the Stanford Institutional Review Board on human participants research and all observers gave prior written informed consent. Observers wore corrective lenses to correct their vision to normal when necessary.

3.2.2 Hardware setup for stimulus and task control

Visual stimuli were generated using MATLAB (The Mathworks, Inc.) and MGL (Gardner et al., 2018a). During scanning, stimuli were displayed via an Eiki LC-WUL100L projector (resolution of 1920x1080, refresh-rate of 100 Hz) on an acrylic sheet mounted inside the scanner bore near the head coil. Visual stimuli were viewed through a mirror mounted on the head coil and responses were collected via an MRI-compatible button box. Outside the scanner, stimuli were displayed on a 22.5 inch VIEWPixx LCD display (resolution of 1900x1200, refresh-rate of 120 Hz) and responses collected via keyboard. Output luminance was measured for both the projector and the LCD display with a PR650 spectrometer (Photo Research, Inc.). The gamma table for each display was dynamically adjusted at the beginning of each trial to linearize the luminance display such that the full resolution of the 8-bit table could be used to display the maximum contrast needed. Other sources of light were minimized during behavior and scanning.

3.2.3 Eye tracking

Eye-tracking was performed using an infrared video-based eye-tracker at 500 Hz (Eyelink 1000; SR Research). Calibration was performed throughout each session to maintain a validation accuracy of less than 1 deg average offset from expected using either a ten-point or thirteen-point calibration

procedure. Trials were canceled on-line when an observers eye position moved more than 1.5 deg away from the center of the fixation cross for more than 300 ms. During training and before data collection, observers were excluded from further participation if we were unable to calibrate the eye tracker to an error of less than 1 deg of visual angle or if their canceled trial rate did not drop to near zero. After training canceled trials consisted of fewer than 0.1% of all trials. Due to technical limitations eye tracking was not performed inside the scanner.

3.2.4 Experimental design

Stimulus

Motion stimuli consisted of two patches of random dot stimuli flanking a central fixation cross (1 x 1 deg). The random dot stimulus patches were rectangular regions extending from 3.5 to 12 deg horizontal and -7 to 7 deg vertical on either side of the fixation cross. Each patch was filled with 21 dots / deg2, 50% brighter and 50% darker than the gray background (300 cd / m^2 in the scanner and 46 cd / m^2 during behavior. All dots moved at 6 deg / s updated on each video frame. Motion strength was adjusted by changing motion coherence: the percentage of dots that moved in a common direction with all other dots moving in random directions. Dots were randomly reassigned on each video frame to be moving in the coherent or random directions. Both patches maintained a constant baseline in between trials of 25% contrast and incoherent motion. To minimize involuntary eye movements, the coherent dot motion direction was randomized to be horizontally inward or outward from fixation on each trial, such that the two patches moved in opposite directions.

Contrast and coherence tasks

Observers performed a two-alternative forced choice judgment about the visibility of the two dot patches (Fig. 3.1). At the start of each run observers were shown the word contrast or motion cueing them to report which side had the higher contrast or motion coherence, respectively. Each run began with a 5 s baseline period during behavioral measurements or 30 s during scanning (25% contrast, 0% coherence) to allow time for adaptation to occur. Trials consisted of a 0.5 s increment in either or both the contrast and motion coherence of the dot patches, a variable delay of 0.5 - 1 s, and a response period of 1s. The dot patches then returned to baseline for an inter-trial interval of 0.2 to 0.4 s randomly sampled from a uniform distribution (2 to 11 s, sampled from an exponential distribution during scanning). The base stimulus strength increments were chosen to be +7.5, +15, +30, and +60% contrast above the baseline 25% contrast and +15, +30, +45, and +60% coherence above the baseline 0% coherence. On every trial one dot patch was chosen as the target for contrast and incremented by an additional small delta, and the same was done independently for coherence. The target increment for the uncued feature was randomly chosen from [0.0, 1.8, 2.5, 3.5, 4.9, 6.9, 9.5, 13.3, 18.5%] for contrast and [0.0, 5.0, 6.9, 9.6, 13.4, 18.6, 25.9, 36.1, 50.2%]

for coherence. The relevant target increment was chosen by a PEST staircase (Taylor & Creelman, 1967) to maintain 82% correct on the cued task for each base strength (4 base strengths $\times$ 2 task conditions = 8 total staircases). Observers indicated with a button press which side contained the delta increment of the cued feature. An observer would be at chance performance if they reported on the wrong feature. Staircases were initialized on the first run (after training) at 25% and 85% for contrast and coherence, respectively. The staircases were maintained across sessions, but the step size was reset to one third the threshold every third run to allow for long-term fluctuation. Before data collection observers trained on the task until their performance at all base stimulus strengths was measurable (i.e. their threshold converged to less than 1 minus the base strength), on average one hour of training. Behavioral runs lasted four minutes and observers took breaks as needed. Observers performed up to 6 one-hour sessions of behavioral runs spanning multiple days.

On a subset of the motion visibility experiment runs (two of every five runs) observers were occasionally asked to report about the non-cued feature (trial probability 1/7, randomized). We refer to these as catch trials. Stimulus presentation occurred as normal on catch trials but after stimulus presentation and a fixed 0.5 s delay, a letter replaced the fixation cross to indicate that the observer needed to recall and respond about the un-cued feature. The length of the delay periods in both catch and regular trials (0.5 s and 0.5 - 1 s, respectively) were chosen to ensure observers could not rely on iconic memory to complete the task (Sperling, 1960) and to avoid observers getting into a rhythm and responding before the post-cue could appear. On contrast runs the post-cue letter was an M indicating that observers should recall about motion coherence and on coherence runs a C to indicate contrast. To improve our statistical power in estimating perceptual sensitivity during catch runs we used a single base stimulus increment: +30% contrast and +40% coherence. These base increments were used both for catch and regular trials on these runs.

3.2.5 Behavioral data analysis

To assess whether the perceptual data could be well characterized by a signal detection framework we tested the fit of cumulative normal distributions to the measured psychometric functions. We collapsed data from all observers across the four base stimulus strengths and separated trials in which observers discriminated contrast or coherence. We binned data according to the difference in stimulus strength for each task and computed the probability of making a rightward choice in each bin (filled circles, Figure 3.2a, b). We fit the binned data with a cumulative normal distribution (three parameters: the mean, μ, standard deviation, σ, and a lapse rate, λ which scaled the function so that it spanned the range $\frac{\lambda}{2}$ to $1-\frac{\lambda}{2}$) and evaluated the cross-validated fit on a held-out observer using the pseudo r^2:

$$r^2_{pseudo} = 1 - \frac{\log(\mathcal{L}_{model})}{\log(\mathcal{L}_{null})} \tag{3.1}$$

where L_{model} is the likelihood of the model given the data and L_{null} is the likelihood of an intercept-only model.

Just-noticeable difference (threshold) estimation

To assess perceptual sensitivity we obtained just-noticeable differences (or thresholds) by fitting a Weibull function (Wichmann & Hill, 2001) to each observers data using maximum likelihood estimation:

$$P_{correct}(x) = \gamma + (1 - \gamma - \lambda)(1 - e^{-[\frac{x}{\tau}]^{\beta}} \tag{3.2}$$

Where x is the difference in signal (either contrast or coherence) between dot patches, γ is the guess rate, λ is the lapse rate, β controls the slope of the function, and τ the value of x at which the function reaches 63% of its maximum. For this two-alternative forced choice task the guess rate was 50% while threshold (when $d^{'} = 1$) corresponds to 76% correct. In total we fit twelve Weibull functions for each observer: eight for the contrast and coherence task (4 base strengths $\times$ 2 task conditions), two for the cued tasks in catch runs (1 base strength $\times$ 2 tasks), and two for the catch trials (1 base strength $\times$ 2 tasks).

3.2.6 Cortical measurement during task performance

We measured how contrast and coherence response functions changed in gain or offset compared to passive fixation in different retinotopically defined cortical visual areas as ten observers performed the contrast or the coherence discrimination task. Our general strategy was based on previous work (Birman & Gardner, 2018) in which we have shown that the relationship between contrast or coherence and BOLD response can be independently parameterized with functional forms, as described below. The analysis proceeded in the following steps. We first used population receptive field measurements (Dumoulin & Wandell, 2008) to determine the location of cortical visual areas in each individual subject. We then took the time series of data averaged across each visual area (for each hemisphere and subject) and performed an event-related analysis to compute the average response to the stimulus presented in the contralateral visual field for each of the 16 combinations of base contrast and coherence and 2 task conditions. We computed the amplitude of response by fitting these event-related responses to a canonical hemodynamic response measured during passive viewing. We had at least 42 measurements (21 repeats in 2 hemispheres) of each base stimulus combination for each task condition in each subject. Consistent with our overall conclusion of flexible readout, comparing these response magnitudes directly between conditions showed weak if any change between conditions. The 95% confidence interval of the differences between tasks included zero for almost all conditions (amplitudes were higher during the contrast task compared the coherence task for 4/16 conditions, averaging over observers). This analysis does not separate out

the independent effects of contrast and coherence across task conditions. So, to gain statistical power and to establish how these BOLD responses reflect difference in contrast and coherence response between task conditions, we used the response magnitudes to scale and shift by additive offset the contrast and coherence response functions, originally based on data from passive viewing. These 6 parameter fits (2 gain parameters and 1 offset parameter for each of the 2 task conditions) were based on 672 (16 base contrast and coherence conditions $\times$ 42 repeats) trial measurements which provided sufficient statistical power and are reported in the main results. Note, for one subject the contrast and coherence values in the conditions differed: only 12 out of 16 conditions were run and with slightly different contrast and coherence values), we were still able to fit the population response function models to this smaller dataset.

All BOLD imaging and data analysis procedures including imaging protocol, preprocessing, data registration across sessions, retinotopic definition of visual areas using population receptive field measurements, and extraction of mean time series from each visual area followed procedures described in detail in Birman and Gardner (2018). Briefly, visual area mapping and cortical measurements were obtained using a multiplexed sequence on a 3 Tesla GE Discovery MR750 (GE Medical Systems) with a Nova Medical 32ch head coil. Functional images were obtained using a whole-brain T2*-weighted two-dimensional gradient-echo acquisition (FOV = 220mm, TR = 500 ms, TE = 30 ms, flip angle = 46 deg, 7 slices at multiplex 8 = 56 total slices, 2.5 mm isotropic). In addition, two whole-brain high-resolution T1-weighted 3D BRAVO sequences were acquired (FOV=240mm, flip angle=12 deg, 0.9 mm isotropic) and averaged to form a canonical anatomical image which was used for segmentation, surface reconstruction, session-to-session alignment, and projection of data onto a flattened cortical surface. Pre-processing was performed using mrTools (Gardner et al., 2018b) and included linear trend removal, high pass filtering (cutoff of 0.01Hz), and motion correction with a rigid body alignment using standard procedures (Gardner et al., 2008). Visual cortical areas V1-V4, V3A/B, V7 (IPS0), and MT (hMT+) were identified using the population receptive field method (Dumoulin & Wandell, 2008) and standard criteria (Wandell et al., 2007). Average time courses were obtained for each cortical visual area by averaging the top twenty-five task-responsive voxels per area. As documented in Birman and Gardner (2018), repeating the analysis using either all voxels, the top two voxels, or all voxels weighted by their population receptive field overlap with the stimulus results in a change in the signal-to-noise in the data, but did not change the relative sensitivities across cortical areas.

To compute event-related responses we assumed that overlapping hemodynamic events sum linearly, an assumption that has been validated explicitly for visual responses (Boynton et al., 1996; Dale & Buckner, 1997). We used a randomized inter-trial interval to avoid cognitive (Zarahn, Aguirre, & D'Esposito, 1997) and hemodynamic (Sirotin & Das, 2009) anticipatory effects and to increase the efficiency of our design (Dale et al., 1999; Liu & Frank, 2004). As violations of linearity have been noted with shorter inter-trial intervals, we chose a mean inter-trial interval of 6 s, sampled

from an exponential with a range of 2 to 11 s, intended to minimize the overlap in the main positive lobe of the hemodynamic response between different events. Moreover, we used a balanced design in which each trial was equally likely to be followed by a trial with any of the other base stimulus strengths to minimize any systematic mis-estimation. We confirmed that the probability of each condition being followed by any other was roughly equal, i.e. $\chi^2(r, 15) > 0.05$, where r was the test statistic computed by comparing the distribution of trial types following each individual trial type against a uniform distribution. No catch trials were run during scanning.

We computed event-related responses for each trial type using a finite-impulse response model (Zarahn et al., 1997) following standard procedures (Gardner et al., 2005). We assumed each combination of different base strengths for contrast and coherence evoked a different hemodynamic response and responses that overlapped in time summed linearly. Because each visual stimulus was lateralized in one half of the visual field, we assumed that they evoked a response only in contralateral retinotopic areas. There were four base increments for contrast (+7.5, +15, +30, and +60%) and four base increments for coherence (+15, +30, +45, and +60%) which were independently manipulated, resulting in 32 total conditions (4 contrast $\times$ 4 coherences $\times$ 2 task conditions). To model these data, we used the following equation:

$$y = X\beta + \epsilon \tag{3.3}$$

Where y was an $n \times 1$ column vector (n = number of volumes) containing the measured hemodynamic response for one hemifield of one visual area in a single observer. X was an $n \times (k \times c)$ stimulus convolution matrix (c = number of conditions, k = length in volumes of hemodynamic response to calculate), β was a $(k \times c) \times 1$ column vector to be estimated, and ϵ the residual variance (assumed to be 0 mean Gaussian). Each block of k columns in X corresponded to one of the c conditions. These blocks contained a one in the first column at the starting volume of each occurrence of a trial of that condition and zeroes elsewhere. Each of the subsequent k columns was then shifted downwards by one to form a Toeplitz matrix for that condition. In total X had n rows, equal to the length of the BOLD timeseries (for most observers n was 13,184), and 2592 columns (k=81 $\times$ c=32, where k was chosen to compute 40.5 s of response and the c conditions were the 4 contrast base strengths $\times$ 4 coherence base strengths $\times$ 2 tasks). By computing the least-squares estimate of the column vector β we obtained the estimated event-related response to each condition accounting linearly for overlap in time. On every trial one dot patch was at a base strength and one had an additional increment. To equate difficulty throughout the task we allowed the additional increments to vary continuously via staircasing. To simplify the estimation problem and to improve statistical power we rounded the base + increment values to the nearest base strength. The choice of number of volumes of response k to compute did not change the result as long as it was sufficiently large to capture the full hemodynamic response. The Pearsons correlation of the first 41 volumes between an analysis with k=41 (20.5 s of response) and k=81 (40.5 s of response) was r=0.97. Because we randomized

trial presentation, we assessed multicollinearity by checking that the stimulus convolution matrices (see below) were full rank and that the off-diagonal elements of the covariance matrix were small (less than 0.1% of off-diagonal elements were larger than 10% of the on-diagonal elements).

To obtain a response magnitude, we fit a scaled canonical hemodynamic response function measured during passive viewing to the event-related responses. We used a canonical hemodynamic response function that was measured in previous work when observers passively viewed the same stimulus (Birman & Gardner, 2018). This function took the form of a difference-of-gamma function whose maximum amplitude was set to one. We fit a single magnitude per condition which scaled this canonical function to minimize the sum of squared error between the event-related response and the scaled canonical function. For each condition (4 contrast base strengths $\times$ 4 coherence base strengths $\times$ 2 tasks) this gave us a scalar response amplitude for the evoked activity in each cortical area.

The response magnitudes for each contrast, coherence, and task condition were next used to estimate how population response functions for contrast and coherence in different visual areas changed in gain and offset during task performance. In our previous work we parametrized the population response to contrast as a sigmoid function (Albrecht & Hamilton, 1982; Naka & Rushton, 1966):

$$R_{con}(s_{con}) = \alpha_{con}(\frac{s_{con}^{1.9}}{s_{con}^{1.6} + \sigma^{1.6}}) \tag{3.4}$$

Where α was the maximum amplitude and σ controlled the shape of the function. The exponents in the function were chosen according to previous work (Boynton et al., 1999). The population response function to coherence was parameterized to be a saturating nonlinear function:

$$R_{coh}(s_{coh}) = \alpha_{coh}(1 - e^{\frac{s_{coh}}{\kappa}}) \tag{3.5}$$

Where the parameter κ controls the shape of the function by setting the point at which the exponential function reaches 63% of its maximum and α_{coh} controls the amplitude. Large values of α_{coh} combined with large values of κ make this function approach linear in the range [0 1] in which the stimulus strength s_{coh} is bounded. Because α_{coh} and κ are not interpretable on their own, we instead report the linear slope of the coherence response functions as a measure of sensitivity (see Birman and Gardner, 2018, for rationale).

We fit the population response functions for each cortical area to the 32 measurements of response magnitude (4 base contrasts $\times$ 4 base coherences $\times$ 2 task conditions) during task performance:

$$R_{area}(s_{con}, s_{coh}) = R_{area,con}(s_{con} + R_{area,coh}(s_{coh}) + \alpha_{task} \tag{3.6}$$

We added the α_{task} parameter to fit additive offset while allowing the α_{con} (Eq. 3.4) and α_{coh} (Eq. 3.5) parameters to change to fit multiplicative gain. The parameters for the response functions

were initialized according to the passive viewing data in Birman and Gardner (2018) with the σ and κ parameters held constant such that response functions maintained their shape. For reference, the initial α_{con} parameter in V1 was 1.68, V2: 0.69, V3: 0.63, V4: 0.61, V3A: 0.35, V3B: 0.24, V7: 0.32, and MT: 0.22. The initial slope of the coherence response function in V1 was 0.07% signal change / unit coherence, V2: 0.16, V3: 0.18, V4: 0.11, V3A: 0.25, V3B: 0.14, V7: 0.20, MT: 0.34. For each cortical area there were six free parameters (3 parameters $\times$ 2 task conditions) fit by minimizing the sum of squared error using the MATLAB function *lsqnonlin*.

3.2.7 Linking model

To link cortical responses to the perception of motion visibility, we modeled the decision process of an observer as a comparison of linearly weighted responses from retinotopically defined visual cortical areas subject to additive Gaussian noise. The model assumed the form of a probit regression in which the difference in weighted cortical responses for the two stimuli were passed through a probit function to make a trial-by-trial prediction of a choice for the stimulus on the right (Fig 3.4). The response to each visual stimulus for each cortical visual area was calculated from the parametric forms of population response functions for contrast and coherence, as defined above. When validating the model assumptions such as additive noise and lack of choice history terms, we used the parameters for the population response functions that were fit to passive viewing data (Birman and Gardner, 2018). To test whether fixed or flexible readouts were needed to explain task performance, we used parameters for the population response functions fit to BOLD data collected during task performance, as described above. The linking model parameters that were fit by maximum likelihood estimation to the behavioral data were the weights for each visual area (in different versions of the model we either fit all 8 visual areas or subsets of visual areas) and a bias term to account for any propensity to choose one side over the other. For the fixed readout there was one set of cortical weights for both tasks and for the flexible readout there were two sets of weights, one for each task. We describe in more detail the specifics of the model below.

We used the population response functions (Eq. 3.4, 3.5) to simulate the trial-by-trial response of visual cortical areas to the stimulus in either hemifield (Eq. 3.6). The parameters of the functions were either from the fit to passive viewing data or during task performance. Summing the response for contrast and coherence assumes that the responses to contrast and coherence are independent of each other, which we showed to be the case in Birman and Gardner (2018).

To obtain the 'readout' of this representation from multiple cortical areas we proceeded by linearly weighting the area responses (Fig. 3.4). The full readout with all visual areas was computed with the following equation:

$$R_{patch}(s_{con}, s_{coh}) = \beta_{V1}R_{V1}(s_{con}, s_{coh}) + \beta_{V2}R_{V2}(s_{con}, s_{coh}) + ... + \beta_{MT}R_{MT}(s_{con}, s_{coh}) \quad (3.7)$$

Where the response for each area on the right side of the equation is computed according to Eq 3.6. Each β was a free parameter which set the weight assigned to cortical areas in the readout process. We use the phrase fixed readout to refer to a model in which there are 8 cortical readout weights in total (one for each cortical area) shared across the two task conditions. Implicitly the fixed readout model therefore assumes that the measured cortical responses must differ between task conditions to accommodate changes in behavior. We use the phrase flexible readout when 16 weights were allowed, i.e. a separate weight for each task for each cortical area. In addition to the 8 cortical area models we also fit models in which we only used the response of areas V1 and MT, the most contrast and coherence sensitive human cortical areas, respectively (Birman & Gardner, 2018).

To compute the probability of an observer choosing the stimulus on the right we passed the difference in response to the two stimuli through a cumulative normal distribution (Bliss, 1934):

$$\begin{aligned} P_{right}(s_{(con,left)}, s_{(con,right)}, s_{(coh,left)}, s_{(coh,right)}) = \\ \Phi(R_{right}(s_{(con,right)}, s_{(coh,right)}) - R_{left}(s_{(con,left)}, s_{(coh,left)}) + \beta_{bias}) \end{aligned} \tag{3.8}$$

Where R_{right} and R_{left} are the weighted cortical responses to the two stimuli on each trial, as calculated using Eq. 3.7. β_{bias} accounts for any bias to one side or another and Φ is the cumulative probability of a normal distribution with $\mu = 0$ and $\sigma = 1$.

In the linking model, we allowed an additional parameter λ to capture the observer's lapse rate, modifying Eq. 3.8:

$$P_{right}(s, ..., \lambda) = \frac{\lambda}{2} + (1 - \lambda)\Phi(R_{right}(s...) - R_{left}(s..) + \beta_{bias}) \tag{3.9}$$

We empirically estimated the lapse rate by finding the rate of observer errors on trials with a stimulus strength far above threshold (Prins, 2012). Because we occasionally reset the step size in the staircases we were able to record a non-negligible number of trials with large stimulus increments, from these we selected trials in which the increment was at least 15% for contrast or 40% for coherence, which corresponded to increments of at least 2× threshold (15% and 40% also correspond to the maximum increment which could be shown at the highest base strength of contrast and coherence, respectively). Computed in this way λ varied from 0 - 7% (mean 3.0%, 95% CI [1.94, 4.56]).

We fit all variants of the linking model with maximum likelihood estimation using the active-set algorithm as implemented by the function fmincon in MATLAB. To avoid getting trapped in local minima we randomized the starting parameters and repeated the fitting procedure multiple times.

We fit the linking model both within observers and across observers to test for generalization. Several observers were involved in both the experiments reported here as well those reported in Birman and Gardner (2018) and so their linking models could be fit within-observer. To ensure generalization we also computed the average population response functions and used those to fit the

linking model to the individual perceptual measurements from each of the 21 observers, including those who did not have within-subject measurements of cortical responses. For the population response functions estimated from passive viewing data the averaged-physiology and within-subject models had similar cross-validated log-likelihoods, $\log(\frac{\mathcal{L}_{average}}{\mathcal{L}_{within}})$= -2.22, 95% CI [-8.18, 4.71]. This suggests that the population response functions were similar across subjects and that noise in the physiological measurements is reduced by averaging across observers. For the measurements during task performance there was a large improvement from using averaged-physiology data, $\log(\frac{\mathcal{L}_{average}}{\mathcal{L}_{within}})$= 34.6, 95% CI [-6.4, 185.4], presumably due to the lower signal-to-noise ratio in those data because the stimulus was limited to 0.5 s.

Linking model variants

To capture bias due to past choices (Abrahamyan, Silva, Dakin, Carandini, & Gardner, 2016; Fründ, McCann, & Williams, 2016) we tested models with additional stay/switch bias parameters. We added four additional parameterstwo which absorbed bias after correct responses (usually found to be a bias toward the same side) and a second which absorbed bias after incorrect responses (usually found to be switching after errors). For clarity we show Eq. 3.8 modified, but this model was still fit with the lapse rate (Eq. 3.9):

$$\begin{aligned} P_{right}(s...) = \Phi(R_{right}(s_{con}, s_{coh}) - R_{left}(s_{con}, s_{coh}) + \beta_{bias} + \beta_{(left,correct)}C_{left} + \\ \beta_{(right,correct)}C_{right} + \beta_{(left,incorrect)}I_{left} + \beta_{(right,incorrect)}I_{right} \end{aligned} \tag{3.10}$$

Where C and I are binary variables set by whether the last trial was correct or incorrect, respectively, and had a response on the corresponding side (i.e. $C_{left} = 1$ if the observer chose left on the last trial and was correct).

We also fit an efficient selection variant of the linking model where responses are weighted according to their magnitude during active viewing (Hara & Gardner, 2014; Pestilli et al., 2011). In this version of the model the responses in each cortical area were raised to an exponent ρ, multiplied by the cortical readout weights, and then the exponent root was taken before passing through the cumulative normal. The effect of this transformation is that an area which has a larger base response, through the exponential, will dominate the final signal. Again, for clarity we show this modification for Eq. 3.8 but the full model included lapse rates (Eq. 3.9):

$$P_{right}(s...) = \Phi(\sqrt[\rho]{R_{right}(s_{con}, s_{coh})^{\rho} - R_{left}(s_{con}, s_{coh})^{\rho} + \beta_{bias}}) \tag{3.11}$$

The linking model described so far makes the assumption that sensory noise limiting perception is additive, i.e. independent of stimulus strength, but we also tested a variation with noise that increased with response strength. If readout was limited by the variability of individual or small

groups of correlated neurons, we might expect sensitivity to be subject to noise which increases with response. We tested this Poisson variant of the model by setting the variance (i.e. σ^2 in Eq. 3.8) of the noise to the average population response in the two dot patches, prior to being passed through the readout weights. Following the equations above this computation is done by averaging the response across areas for each dot patch:

$$\sigma^2_{patch} = \frac{R_{V1} + R_{V2} + ... + R_{MT}}{N} \tag{3.12}$$

Where N is the number of areas averaged and R_{area} is computed using Eq. 3.6. We based the noise on the signal prior to readout under the assumption that Poisson noise would be generated by spiking variability occurring in the sensory system.

3.2.8 Interpreting linking model parameters

Using the fit model parameters, we were able to determine an estimate of the magnitude of noise limiting an observers perceptual sensitivity in units of BOLD percent signal change. Because we set $\sigma = 1$ in the cumulative normal function of Eq. 3.7 we can estimate the noise in the sensory representation from the weight parameters. According to Eq. 3.8, a unit input difference between R_{right} and R_{left} will allow the observer to achieve threshold performance. It follows then that the β weights (Eq. 3.7) can be interpreted as scaling the raw BOLD responses such that a unit difference in weighted response gives rise to threshold performance. Assuming a standard signal-detection model where perceptual sensitivity (d) is equal to the difference in responses divided by the standard deviation of the noise, a small β weight would suggest a large amount of noise is limiting perception as it would take a very large difference in response to get threshold performance. Conversely a large β weight would suggest the opposite, that only small differences in response are needed for threshold performance. More formally, if one considers just one area, such as V1:

$$\text{Threshold performance}(d' = 1) = \frac{(R_{V1,right} - R_{V1,left})}{\sigma_{V1}} = \beta_{V1}(R_{V1,right} - R_{V1,left}) \tag{3.13}$$

Therefore, the β weights are inversely proportional to the implied neural noise, σ, of the representation which limits perception.

To recover the model's just-noticeable differences (Fig. 3.2) we proceeded analytically. As described above, because we fit the additive noise model with the noise parameter $\sigma = 1$ the population response functions, after scaling by the beta weights, are in units of standard deviations. To find the just-noticeable difference relative to a base stimulus strength we simply calculated the increment in signal needed to increase the readout response by one, equivalent to $d^{'} = 1$. This is because when $\sigma = 1$ we can reduce:

$$d' = 1 = \frac{R(base + increment) - R(base)}{\sigma} \tag{3.14}$$

to simply:

$$R(base + increment) - R(base) = 1 \tag{3.15}$$

Model Comparison

To compare the different variants of the linking model we used the cross-validated log-likelihood ratio and Tjurs coefficient of discrimination (Tjur, 2009). Each variation of the linking model was fit in a 10-fold cross-validation procedure. 10% of the data was reserved for validation while the remaining 90% used to train. The log-likelihood was computed for each validation set and summed across all ten folds. To compare any two variations of the linking model we computed their likelihood ratio (i.e. the difference in total log-likelihood). The cross-validated log-likelihood ratio is similar in principle to measures of information criterion and sometimes referred to as the cross-validated information criterion (McLachlan & Peel, 2000). When the difference in this statistic between two models is large, e.g. greater than 10 (Burnham & Anderson, 2004), it indicates a substantial improvement in model fit. We use the cross-validated log-likelihood ratio rather than other information criterions (e.g. AIC, or BIC) because the cross-validation procedure already penalizes models with additional parameters for over-fitting. Although the cross-validated log-likelihood is useful for model comparison it is difficult to interpret its absolute magnitude in isolation. To help with interpretation we also report the cross-validated coefficient of discrimination CD.

$$CD = \mu_{right} - \mu_{left} \tag{3.16}$$

Where μ_{right} is the models average predicted likelihood of a rightward choice for validation trials when the observer chose right and μ_{left} when the observer chose left. If the model predicts choices perfectly, then μ_{right} would be 1 and μ_{left} would be 0, giving a value for CD of 1. If the model is at chance at predicting choices than CD would be 0. CD therefore indexes the difference between the centers of the trial-by-trial prediction distributions and although not a true measure of variance explained it shares many of the properties of r^2 and can be interpreted in a similar manner (Tjur, 2009).

3.3 Results

3.3.1 Perceptual sensitivity to motion visibility

We characterized human perceptual sensitivity to the contrast and coherence of moving dots while observers had to report exclusively about one feature and ignore the other. We measured observers

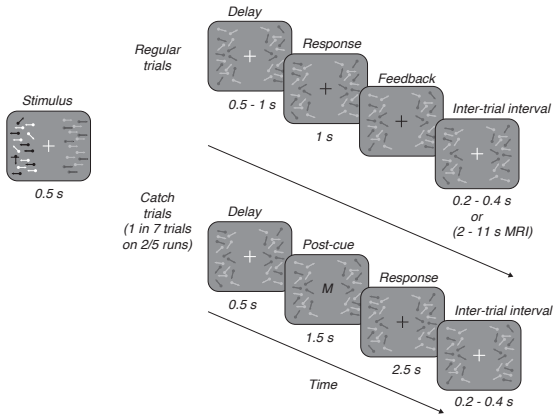

Figure 3.1: Behavioral task. Observers discriminated which of two random dot stimulus patches had higher contrast or coherence in different blocks of trials. Each block began with the word contrast or motion indicating that observers should report about contrast or coherence, respectively, and ignore the other feature. Between trials (Inter-trial interval) and during all but the Stimulus segment, the dot patches were presented at 25% contrast with incoherent motion. On each trial both dot patches increased by independent base increments of contrast and coherence (+7.5, +15, +30, or +60% contrast +15, +30, +45, or +60% coherence) for 0.5s (Stimulus). In addition, for each feature one side was chosen independently to have an additional threshold-level increment, determined by a staircasing procedure. For regular trials, after a 0.5 - 1s period (Delay), observers were asked to report which side contained the additional increment in contrast or coherence (Response) and were given feedback (Feedback). On a subset (Catch trials) of runs (2/5) on rare trials (1/7) the delay period was followed by a post-cue (Post-cue), the letter M or C, indicating that the observers should prepare a response about the un-cued feature. Additional time was given to observers to make these decisions (post-cue period of 1.5 s, response window of 2.5 s) and observers did not receive feedback on catch trials.

just-noticeable differences (JND) in image contrast or motion coherence between a pair of simultaneously presented random dot stimulus patches in a two-alternative forced choice task (Fig 3.1). Each block of trials began with either the word “contrast”, indicating that observers should report which dot patch had higher contrast while ignoring differences in coherence, or “motion”, indicating the opposite. Each trial consisted of a 0.5 s base increment in the contrast and coherence of both dot patches (at all other times the dot patches were kept visible at 25% contrast and 0% coherence). In addition to this base increment a small additional increment near perceptual threshold was added to one side independently for each feature. Therefore, for every trial regardless of cueing condition there was a difference in both features between the two dot patches and each patch was equally likely to contain the additional increment. After stimulus presentation and a brief delay, observers reported which side had the higher magnitude of the cued feature and received feedback.

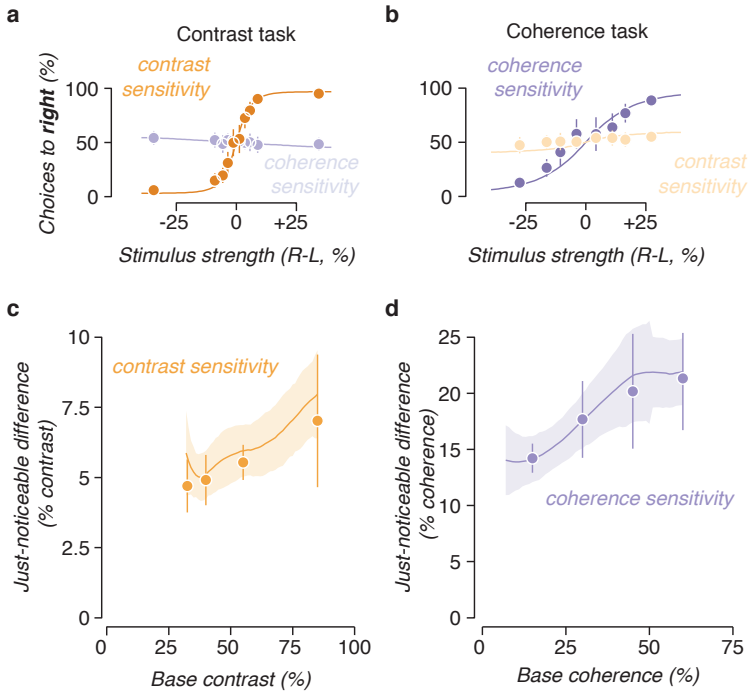

Figure 3.2: Perceptual sensitivity to contrast and motion coherence and fit of validation linking model. (a) Contrast task. The markers plot the average probability across observers and base stimulus strengths of indicating that the right dot patch had higher contrast or motion coherence while performing the contrast task, as a function of the difference in contrast (orange) or coherence (blue) between the two patches. Curves plot the predictions of the eight-area linking model using measurements made during passive viewing. These were fit to each individual observers behavioral data with a flexible readout, therefore fitting each task separately. (b) Coherence task, conventions same as (a). (c) Markers plot the just-noticeable difference for contrast during that task estimated from a Weibull function fit for each base stimulus strength, averaged across observers. Curves indicate the average prediction of the 8-area linking model across observers. (d) Same as (c) for the coherence task. All markers indicate the mean and error bars the 95% confidence interval across observers. Curves indicate the mean model prediction across observers and shaded areas the 95% confidence intervals. Some error bars are hidden by the markers.

3.3.2 Observers were able to report about each motion visibility feature independently.

Collapsing across observers and base stimulus strengths we found that observers were sensitive to the feature they were asked to report (dark orange, Fig. 3.2a, and dark purple, Fig. 3.2b), but insensitive to the features they were asked to ignore (light purple and light orange, Fig. 3.2a and 3.2b). The psychometric functions (circle markers in Fig. 3.2a) were well fit by cumulative normal distributions (not shown, average cross-validated $r^2_{pseudo} = 87.7\%$). This suggests that observers decisions were consistent with a signal detection process in which two sensory representations were

compared subject to Gaussian noise. Separating out sensitivity by base stimulus strength, we observed a proportional increase in just-noticeable differences (Fig. 3.2c-d) reminiscent of Weber's law. Webers law states that the slope of this relationship should be 1 on a log-log axis but we found slopes less than one for contrast, 0.44, 95% CI [0.41 0.50], consistent with previous studies (Gorea & Sagi, 2001; Pestilli et al., 2011), and 0.81, 95% CI [0.78 0.85] for coherence. Fitting a Weibull function on a subject-by-subject basis for base contrasts 32.5, 40, 55 and 85% we found just-noticeable differences in contrast (Fig. 2c) to be 4.6%, 95% CI [3.8, 5.5], 4.8%, 95% CI [3.9, 5.8], 5.5%, 95% CI [4.9, 6.2], and 7.5%, 95% CI [5.3, 9.8], respectively. For base coherences 15, 30, 45, and 60% we found just-noticeable differences in coherence (Fig. 2D) to be 14.2%, 95% CI [12.9, 15.5], 17.7%, 95% CI [14.3, 21.1], 20.2%, 95% CI [15.1, 25.3], and 21.3%, 95% CI [16.7, 25.9], respectively. Note that for contrast the base stimulus strengths are reported as the absolute value and not the relative increment from the 25% contrast and incoherent motion that was shown continuously throughout the experiment.

3.3.3 Changes in cortical representation of motion visibility during task performance

We measured BOLD signal in retinotopically defined visual areas and found small changes in sensory responses when observers switched between reporting contrast and coherence (Fig. 3). Ten of the observers who performed the behavioral experiments repeated the task in the magnet. We used these measurements to examine how the contrast and coherence responses changed, either by multiplicative gain or additive offset, in each visual area (see Methods). For a majority of subjects, we found that when reporting about contrast, compared to reporting about coherence, the response to contrast in cortex showed a multiplicative gain (Fig. 3.3a). The average increase in α_{con} (Eq. 3.4) over areas and observers was 0.13% signal change / unit contrast, 95% CI [0.07, 0.19]. The direction of this effect wasnt always consistent, in V1 8/10 observers showed an increase; for V2 6/10; V3 7/10; V4 7/10; V3a 7/10; V3b 7/10; V7 5/10; MT 6/10. For the coherence response, we found no consistent change in the slope of the response function when reporting about coherence 3.3b). The average over areas and observers was -0.02% signal change / unit coherence, 95% CI [-0.08, 0.04]), though some individual areas like MT showed an increase. These changes were inconsistent across observers, in V1 6/10 observers showed an increase in the linear slope of the coherence response; V2 6/10; V3 6/10; V4 6/10; V3a 4/10; V3b 5/10; V7 6/10; MT 6/10). In some linking models additive offsets have been shown to account for the perceptual benefits of selective attention (Pestilli et al., 2011). We found that reporting about the stimuli, rather than passively viewing them, led to an additive offset in most visual areas (Fig. 3.3c). Average increase in α_{task} (Eq. 3.6) over areas and observers compared to passive viewing was 0.36% signal change, 95% CI [0.30, 0.44]. Additive offsets were slightly larger during the contrast task than the coherence task (Fig. 3.3c). Averaged over areas and observers this effect was a modest 0.07% signal change, 95% CI [0.01, 0.14]. In summary,

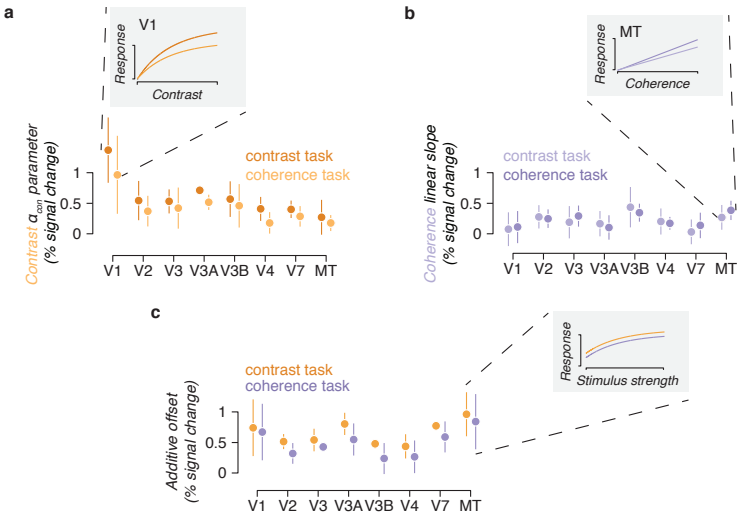

Figure 3.3: Cortical measurements during active viewing. Observers performed the behavioral task while hemodynamic responses in retinotopic visual cortex were measured. (a) The average across observers of the α_{con} parameter, a measure of contrast sensitivity, is shown for each task context (dark and light markers are the contrast and coherence task, respectively). Inset shows how the change in the parameter affects the change in the contrast response function for V1, ignoring any change in additive offset. (b) As in (a) for coherence sensitivity as measured by the linear slope of the coherence response function (dark and light markers are coherence and contrast task, respectively) and inset shows the relationship for MT. (c) As in (a-b) for the α_{task} parameter which absorbs additive offsets. Inset shows the additive offset shift for MT.

we measured small changes in sensory response between task conditions and found that in some cortical areas contrast sensitivity increases when subjects perform the contrast task and coherence sensitivity increases when subjects perform the coherence task. While these changes are in the right direction to underlie task performance, a formal linking model is required to determine if they are large enough to account for perceptual behavior.

3.3.4 Linking model between cortical representation and perception of motion visibility

We set out to build such a linking model (Fig. 3.4) that could quantitatively predict behavioral performance from measurements of cortical sensory representation. Once validated, such a model could then be used to assess whether the sensory changes we measured were large enough to explain behavioral performance in the task conditions. Linking models have been built for contrast discrimination tasks by assuming that higher contrast is detected by comparing the magnitude of cortical responses evoked by different stimuli, subject to some noise (Boynton et al., 1999; Foley & Legge, 1981; Gardner, 2015; Ling & Carrasco, 2006; Nachmias & Sansbury, 1974; Pestilli et al., 2009). Behavioral sensitivity is determined by the ratio of response difference to the standard deviation of

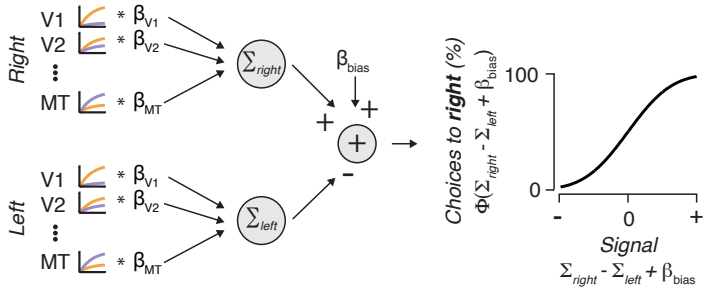

Figure 3.4: Readout linking model. The linking model simulates the cortical response evoked by each dot patch according to an existing framework (Birman & Gardner, 2018) which parameterized the contrast response function (orange curves) as a Naka-Rushton and the coherence response function (blue curves) as linear, or a saturating exponential function. The model weights the cortical responses from each visual area (β values) evoked by the stimulus (Right or Left) according to the current task. The model then takes the difference between the signals evoked by each stimulus, plus a bias term (β_{bias}) to account for any individual observers bias to choose one response over the other. To convert from this weighted signal to probability of choosing the patch on the right, the signal is passed through a cumulative normal distribution (curve on right). The linking model is analogous to probit regression with nonlinear input signals.

the noise, as in the classic signal detection measure d. In our task, cortical responses are the result of stimuli that differ both in contrast and coherence. The linking model therefore needed to be able to differentiate which feature caused a difference in response. We reasoned that this could be accomplished by properly weighting visual areas with different sensitivity to each stimulus feature. Our model took the form of a probit regression (Bliss, 1934) in which the difference in weighted response of visual areas to the two stimuli were computed and passed through the cumulative normal distribution to predict the probability of different choices (Fig. 3.4, see Methods: Linking model for full description).

Before evaluating such a model on the measurements of cortical activity during task performance (Fig. 3.3), we wanted to validate that such a linking model could in principle account for contrast and coherence discrimination. In previous work we published measurements of contrast and coherence response in cortex while observers passively viewed the same random dot stimuli used here (Birman & Gardner, 2018). These measurements were used to quantify the shape of contrast and coherence responses across retinotopically defined visual areas using functional forms (Naka-Rushton for contrast and a saturating exponential form for coherence, see Eq. 4 and 5). These passive-response data showed, for example, that V1-V4 are relatively more sensitive to changes in image contrast, whereas MT is more sensitive to changes in motion coherence. For reference, the parameters describing these differences in sensitivity are reported in the Methods (for additional details see: Birman and Gardner (2018)). Using the functional forms measured during passive viewing we simulated the trial-by-trial response of eight visual cortical areas, V1-V3, V4 (hV4), V3A, V3B, V7, and MT

(hMT+), and modeled sensory readout on each trial as a task-dependent linear weighting of the population responses (Fig. 3.4). This resulted in a scalar response for the left and right stimulus patches (Σ_{right}, Σ_{left}) on each trial. The observers decision about which side had the higher cued feature was modeled as a comparison between these two scalar responses ($\Sigma_{right} - \Sigma_{left}$) summed with a side bias (β_{bias}). This scalar decision variable was subject to Gaussian noise as implied by the cumulative normal of the probit link function. We fit the parameters of the linking model using maximum likelihood estimation for each observer (8 cortical area weights $\times$ 2 task conditions + 1 bias parameter = 17 total parameters) using the average population response functions from Birman and Gardner (2018).

We found that the linking model based on the passive viewing BOLD data was a good fit for the behavioral measurements (curves, Fig. 3.2), capturing both the shape of the psychometric functions and the increase in just-noticeable differences with increasing base stimulus strength. To evaluate models we examined Tjurs coefficient of determination (CD), a measure intended to be interpreted similarly to r^2 for models of binary decisions (Tjur, 2009). To compare models, we computed cross-validated log-likelihood ratios (see Methods: Model comparison). We found across observers an average CD of 0.44, 95% CI [0.42, 0.45] reflecting that the model captured the sensitivity of human observers to differences in visibility across both task conditions (curves, Fig. 3.2a-b) as well as the reduced sensitivity at increasing base stimulus strength (curves, Fig. 3.2c-d). The fits shown are for the 8-area model, but we also tested a model with only the two areas with the highest contrast and coherence sensitivity, V1 and MT (2 cortical area weights $\times$ 2 task conditions + 1 bias parameter = 5 total parameters). We found a similarly good fit, $\log(\frac{\mathcal{L}_2}{\mathcal{L}_8}) = 7.38$, 95% CI [-3.09, 32.78]. The average CD of the 2-area model was also 0.44, 95% CI [0.43, 0.45].

The linking model fit weights according to the relative sensitivity of each cortical area to contrast and coherence (Fig. 3.5). In the 8-area model the contrast task weights (x-axis, Fig. 3.5a) are proportional to how sensitive each area is to contrast relative to coherence: V1-V4 have positive weights, while only MT was given a negative weight. The negative weight on MT counteracts sensitivity to coherence in V1-V4 and ensures the linking model was insensitive to coherence when reading out contrast. The weights for the coherence task (y-axis, Fig. 3.5a) behaved similarly, with MT getting the largest positive weight and V1 a slight negative one. A similar pattern was observed for a model with only areas V1 and MT (Fig. 3.5b) but with less negative weighting in the coherence readout.

Using model comparison, we validated our linking model assumptions that noise is additive and that observers had no dependency on choice history. Models based on single-unit variability often assume a Poisson-like noise (Softky & Koch, 1993), but because our model is based on population activity for which independent single-unit variable would be expected to average out, we modeled an additive noise component. This choice of additive Gaussian noise was important. A model using Poisson noise which increased with stimulus strength did not fit the data (Fig. 3.6). On average

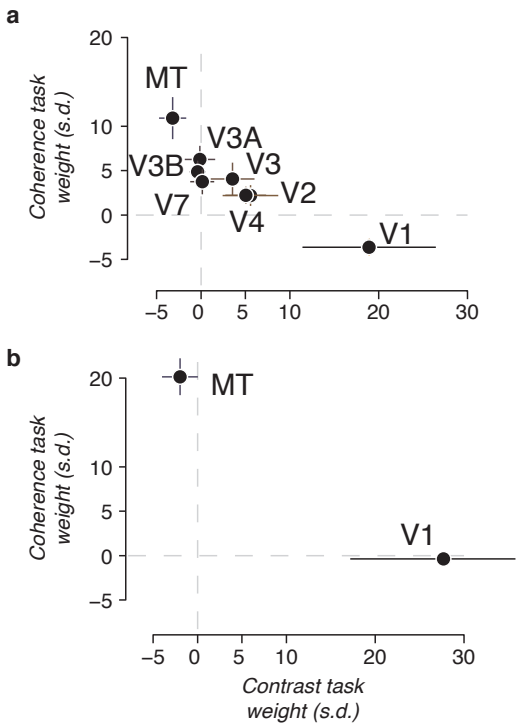

Figure 3.5: Cortical area weights. (a) The weights of the flexible-readout model fit to passive viewing data are shown for the contrast task (x-axis) and coherence task (y-axis) for the eight cortical areas we defined retinotopically: V1, V2, V3, V4 (hV4), V3A, V3B, V7, and MT (hMT+). (b) As in (a) but for the 2-area model with only V1 and MT. All markers indicate the mean across observers and error bars the 95% confidence intervals.

across observers the additive model was a better fit compared to the Poisson model, $\log(\frac{\mathcal{L}_{additive}}{\mathcal{L}_{Poisson}})$= 43.58, 95% CI [18.84, 77.93] (Fig. 6c) and improved CD by 0.01, 95% CI [0.00, 0.02] (Fig. 6d). A number of studies have found that observers performing psychophysical tasks are biased by previous choices even when those choices are uninformative for the current trial (Abrahamyan et al., 2016; Fründ, Wichmann, & Macke, 2014). We also tested for possible biases due to choice history (see Methods) but found that including these additional fit parameters caused the cross-validated log-likelihood to deteriorate, suggesting over-fitting, $\log(\frac{\mathcal{L}_{original}}{\mathcal{L}_{stay/switch}})$= 3.66, 95% CI [0.31, 9.08]. Thus, model comparison was able to validate that choice history effects were negligible, and that noise was best assumed to be additive rather than Poisson.

3.3.5 Using the linking model to test fixed vs flexible readout

Having verified the linking model based on passive viewing data, we now asked whether the small changes in sensory representation which we measured during task performance could account for how

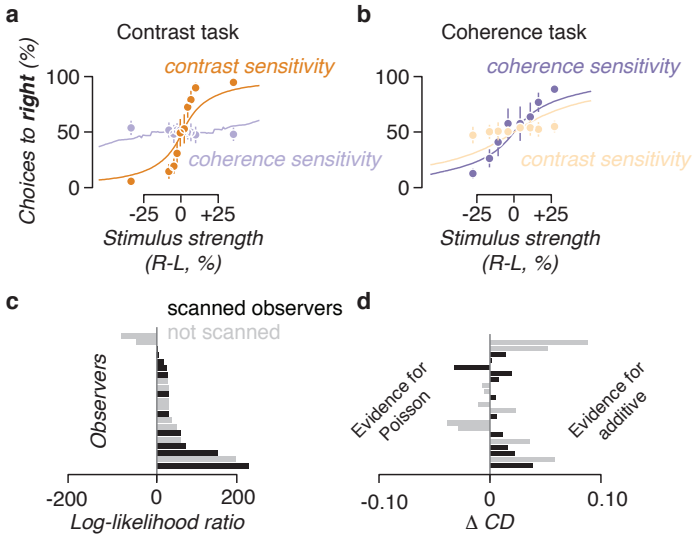

Figure 3.6: Poisson vs. additive noise. (a,b) Same conventions as Fig 3.2a and b, except curves indicate the prediction of a model incorporating Poisson-like noise in which variance scales equal to response strength. (c) The cross-validated likelihood ratio (difference in log-likelihood) between the additive and Poisson models is shown for each observer. (d) The change in Tjurs CD comparing the additive models against the Poisson models, a measure analogous to r^2. Observers are sorted as in (c). In panels (c) and (d) evidence for the Poisson model is plotted towards the left and evidence for the additive model to the right.

perceptual sensitivity changed when observers switched task. If sensory changes were sufficiently large, then the readout could be fixed between task conditions. Such a fixed readout model would only require a single set of cortical area weights with changes in perception accounted for only by changes in sensory responses. As a baseline for comparison, we first fit the fixed-readout model on the sensory responses measured during passive viewing where by definition there are no sensory changes between task conditions. This passive-response fixed-readout model can only produce behavior that is intermediate between the two tasks. That is, it is sensitive to both contrast and coherence (Fig 3.7a, orange/yellow contrast curves and blue/purple coherence curves are not flat) and cannot switch sensitivity between the two tasks (Fig 3.7a, curves for left and right panels are identical). The CD and likelihood of the passive-response fixed-readout model provide a lower bound on the possible explainable variance (Fig. 3.7d and e).

Fitting the fixed-readout model to sensory responses measured during task performance showed that while changes in sensory response could account for a substantial amount of the behavioral performance, the changes were insufficiently large to fully explain task performance. This task-response fixed-readout model achieved a better fit of the behavioral data than the passive-response fixed-readout model (Fig. 3.7d and e, compare magenta and blue points) thus quantifying how

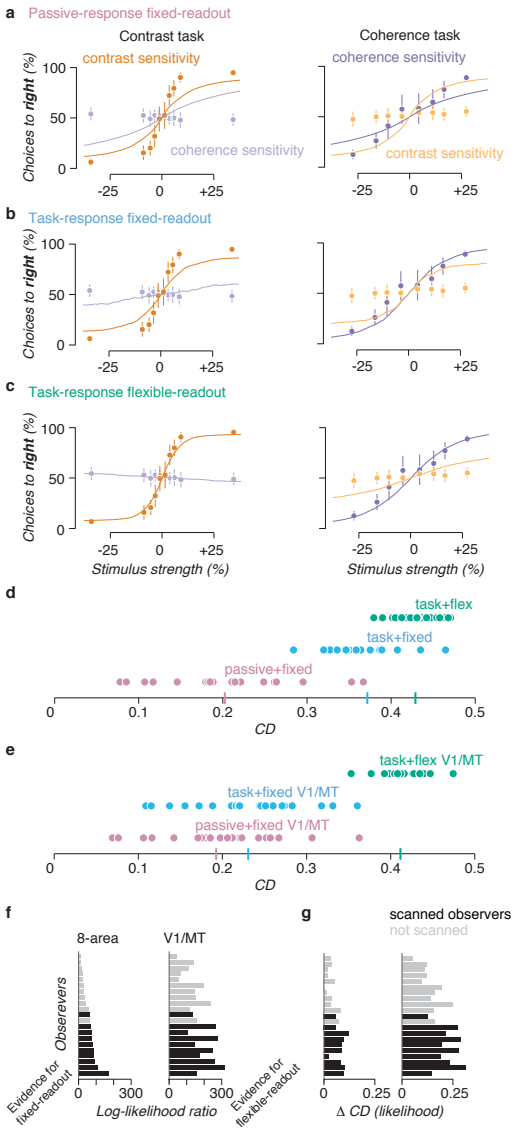

Figure 3.7: Comparing fixed- and flexible-readout linking models. (a-c) Same conventions as Fig. 3.2a and b, except curves plot: (a) the fit of the passive BOLD measurements with a fixed readout, (b) the BOLD measurements during task performance with a fixed readout, and (c) the BOLD measurements during task performance with a flexible readout. The fixed-readout model forces any change in perceptual sensitivity to be the result of differences in sensory response between tasks by using only a single set of cortical readout weights (eight weights and one bias term). The flexible-readout model allows a different set of weights for each task condition (sixteen weights and one bias term). (d) Tjurs CD for the models in (a-c). Averages are shown as a bar on the axis. (e) Conventions as in (d) for the 2-area models with only V1 and MT. (f) Model comparison of the cross-validated likelihood ratio (difference in log-likelihood) between the task-response fixed-readout and task-response flexible-readout models. Evidence for the fixed-readout model is plotted to the left and flexible-readout to the right (none of the fits are in favor of the fixed readout model) (g) As in (f) for Tjurs CD. Markers in panels a-c indicate the average across observers. Markers in panels d and e indicate individual observers. Error bars are the 95% confidence intervals. Some error bars are hidden by the markers.

much the sensory changes reported above can account for behavioral performance. Indeed, the task-response fixed-readout model was better able to capture differences in behavior between the contrast and coherence task (Fig 3.7b, compare curves for left and right columns). However, the linking model failed to completely capture the ability of subjects to change their perceptual sensitivity to contrast and coherence between the two tasks. In the contrast task, the contrast sensitivity curve (orange, Fig 3.7b) does not match the sensitivity of the observers and the model predicted a weak bias for coherence (light purple, left panel Fig 3.7b) that the subjects did not show. In the coherence task, the coherence performance was reasonably well-matched (purple curve, right panel), but the model predicted strong bias from contrast (orange curve).

Rather than rely only on changes in sensory representation between tasks, a linking model that could read out responses from visual areas differently between tasks was better able to fit the behavioral performance. We tested this task-response flexible-readout model by allowing the weights for different visual areas to change between tasks while still using the sensory responses measured during task performance. This model provided reasonable fits to the behavioral data (Fig. 3.7c), capturing the performance during the contrast task (left column), although it did predict more bias to contrast during the coherence task then the observers displayed (orange curve, right panel). Because the fixed-readout and flexible-readout models had different numbers of parameters (fixed-readout = 9 parameters, flexible-readout = 17) it was critical to evaluate the models with a cross-validated metric. We found that for the task-response measurements the flexible-readout model was a far better fit than the fixed-readout model (Fig. 3.6f and g), $\log(\frac{\mathcal{L}_{flexible}}{\mathcal{L}_{fixed}}))$= 60.16, 95% CI [44.90, 77.75], difference in CD, 0.06, 95% CI [0.04, 0.07]. Note that observers who we measured physiology for (black bars, Fig. 3.7f and g) show a larger improvement in model fit compared to the other observers, which we attribute to an effect of increased training.

Observers who we measured physiology for (black bars, Fig. 6C and D) show a larger improvement in model fit compared to the other observers (gray bars). One explanation for this effect is that the observers we measured in the scanner were better able to ignore the irrelevant feature due to having more practice. The fixed readout model, which predicts an inability to ignore the irrelevant feature, would then fail more dramatically for better-trained observers. Indeed, observers who were a part of the scanning were slightly better (n=11 observers, mean just-noticeable difference for contrast 5.62%, 95% CI [4.98, 6.50], and coherence 18.06%, 95% CI [16.32, 21.15]) compared to observers who did not participate in scanning (n=10, mean just-noticeable difference for contrast 10.02, 95% CI [6.23, 19.95], and coherence 21.40%, 95% CI [18.64, 25.03]).

As additive offsets have been used with a fixed readout to explain behavioral performance differences with spatial attention (Pestilli et al., 2011), we also tested an efficient selection model that weights responses according to their magnitude, but found that this model also could not explain the behavioral performance. An increase in additive offset during one task condition or the other

could be used by an efficient selection model that weighs signals by their magnitude (Hara & Gardner, 2014; Pestilli et al., 2011), e.g. selecting out V1 during the contrast task and MT during the coherence task. On average response magnitudes did increase a moderate amount when observers performed the task compared to the passive viewing condition, but these additive offsets were similar for both tasks (Fig. 3.5C). We found that the flexible model was a far better explanation than an efficient selection model (see Methods for implementation details), $\log(\frac{\mathcal{L}_{flexible}}{\mathcal{L}_{selection}})$= 130.39, 95% CI [109.66, 151.31], difference in CD, 0.30, 95% CI [0.28, 0.32].

3.3.6 Behavioral evidence for a flexible readout

One advantage to keeping sensory representations relatively stable is that observers can maintain information about unattended features. To measure whether observers could recall unattended information we included catch trials in the behavioral task. In catch trials, observers were shown a post-cue after stimulus presentation which indicated that they should report about the un-cued feature (bottom time line, Fig. 3.1). Observers made these reports despite the stimulus having already been presented and despite having already had 0.5 s to prepare their response for the main task. We were able to ensure observers did not split their attention by keeping the main task at perceptual threshold, making catch trials rare, and not providing feedback.

Because observers were told at the start of each block (65 trials or 4 minutes) whether or not catch trials would occur there is a concern that they could have split their attention, but we found no evidence for this. In other dual task settings there is a significant cost associated with performing two tasks at the same time (Sperling & Melchner, 1978), especially when one or both tasks are difficult (near perceptual threshold). Note that we designed the catch trials to minimize this effect by making them rare and not providing feedback. If observers split their attention, we would expect to detect an increased just-noticeable difference (JND) on the cued main task. Instead, we found that the just-noticeable differences were similar: on runs with catch trials the contrast task JND increased by only 0.19% contrast, 95% CI [-0.19, 0.78], and for the coherence task by 0.74% coherence, 95% CI [-0.76, 3.13].

During catch trials we found that observers were less sensitive to the un-cued motion visibility features compared to when they were cued, but nevertheless they maintained significant information about the unattended features. Observers just-noticeable differences were larger on the catch trials both for the contrast task (average Δ just-noticeable difference = +5.30% contrast, 95% CI [+3.83, +7.22]) and coherence task (Δ just-noticeable difference = +45.84%, 95% CI [+26.17, +98.23]) (Fig. 7). These averages (and subsequent analysis) exclude 4/21 and 1/21 observers for the coherence and contrast tasks, respectively, because they could not perform the task and their just-noticeable differences were not measurable (i.e. their JND was more than what could be displayed on the screen).

If observers had a fixed readout which could not switch to the ignored feature during catch trials,

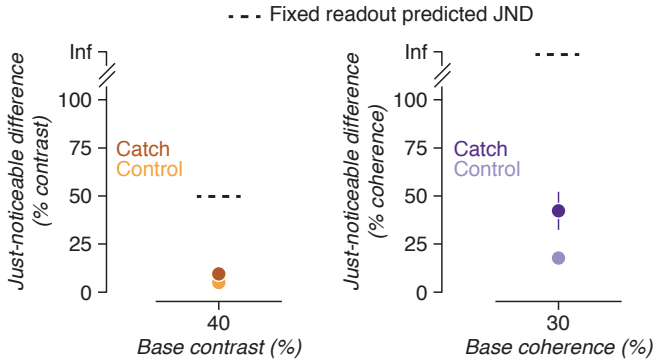

Figure 3.8: Perceptual sensitivity on catch trials. Just-noticeable differences (JND) for contrast (left) and coherence (right) are shown for the regular (control, light colors) and catch (dark colors) trials during runs that included catch trials. Predicted just-noticeable differences for catch trials are shown for the fixed readout model (dashed lines). Markers indicate the average across observers and error bars the 95% confidence intervals, some error bars are hidden by the markers.

then they would be forced to use the wrong readout and performance would be extremely poor. We found that this fixed readout model predicted much higher just-noticeable differences than measured and therefore could not account for catch trial behavior. That is, we used the task-response flexible-readout model to compute the expected JND on catch-trials assuming that observers were unable to switch the readout to the post-cued feature. For example, for the contrast-task in which the catch trials required making a coherence judgement, we used the cortical readout weights for the contrast-task (β_{V1},β_{V2},...), and vice-versa. This model underestimated human performance on catch trials (dashed lines, Fig. 3.8). On contrast catch trials (i.e. post-cued trials when observers reported about contrast, during a run where the main task was coherence) the model predicted just-noticeable differences of 56.9%, 95% CI [33.8, 80.1], but the average observer had a JND of only 10.1% contrast, 95% CI [7.5, 12.7]. On coherence catch trials the model predicted that observers would be incapable of performing the task but the average observer JND was 40.8% coherence, 95% CI [32.4, 49.2].

Instead, we found that a better explanation for catch trial behavior came from a readout which could dynamically change within trials but incurred an additional cost for maintaining sensory information in working memory. This cost could be due to a drop in the signal-to-noise of the sensory representation, perhaps due to responses degrading over time. We estimated the cost by dividing the thresholds measured during catch trials by the thresholds measured during regular trials. This approach suggests that on average σ increased (or responses degraded) for coherence by a factor of 3.44, 95% CI [2.46, 5.90] and for contrast by 2.21, 95% CI [1.87, 2.66]. The overlap in estimates suggests a single cost, i.e. the change from a discrimination task to a working memory task, might govern the change in performance for both tasks; averaging the increase in noise gives an estimated reduction in sensitivity of 2.83, 95% CI [2.31, 4.17]. Thus, a model which allows rapid re-weighting, combined with a fixed cost for using working memory, can explain behavioral

performance for both contrast and coherence catch trials.

3.4 Discussion

We found that observers were able to independently judge the visibility of patches of moving dots based on either their contrast or coherence. Concurrent measurements of BOLD activity showed that there were small changes in sensory representations during task performance. Cortical responses were somewhat more sensitive to contrast during contrast discrimination and, in some areas like MT, more sensitive to coherence during the coherence task. Our analysis with a fixed-readout linking model showed that these changes could account for some, but not all of the behavioral performance. Instead, the behavior was consistent with a flexible readout of sensory representations. Keeping representations relatively stable should allow observers to retain information about unattended features and we found that during catch-trials this was the case. Our results highlight the importance of using models that quantify the link between cortical representation and perception.

3.4.1 Linking models for human motion visibility perception

We manipulated the contrast and coherence of random dot motion stimuli because of the extensive existing knowledge of how neural representations of these features are related to visual perception (Gold & Shadlen, 2007) and because their similar representation in cortex suggests that changing the representation of one will necessarily affect the other. Contrast, the average difference between bright and dark (Bex & Makous, 2002), and coherence, the percentage of dots moving in the same direction, both control the visibility of motion. Human cortical visual areas are known to be sensitive to these properties such that an increase in visibility results in monotonically increasing responses throughout visual cortex (Avidan et al., 2002; Birman & Gardner, 2018; Britten et al., 1993; Gardner et al., 2005; Logothetis et al., 2001; Olman et al., 2004; Boynton et al., 1996; Olman et al., 2004; Rees et al., 2000; Tootell, Hadjikhani, Mendola, Marrett, & Dale, 1998b; Simoncelli & Heeger, 1998). For observers to judge these two features independently their sensory representations need to be separated according to context, a step which existing linking models built for single features have not had to contend with.

The computational steps from sensory representation to perception have been well characterized for contrast discrimination. In these linking models an observers choice is computed by comparing the evoked neuronal responses to different stimuli (Boynton et al., 1999; Foley & Legge, 1981; Ling & Carrasco, 2006; Nachmias & Sansbury, 1974; Pestilli et al., 2009). Individual neurons exhibit monotonically increasing responses to contrast (Albrecht & Hamilton, 1982), with different parameterizations (Tolhurst et al., 1983) that can be pooled into a population response (Shadlen et al., 1996). Such population responses to contrast are well-indexed by BOLD signal in human visual cortex (Avidan et al., 2002; Boynton et al., 1996; Boynton et al., 1999; Gardner et al., 2005; Heeger

et al., 2000; Logothetis et al., 2001). Linking models have been shown to account for BOLD signal measurements and perceptual responses during contrast discrimination tasks (Boynton et al., 1999), predict changes in these measures during surround masking (Zenger-Landolt & Heeger, 2003) and detection (Ress et al., 2000), and have been used to describe the selection of signals from attended locations (Hara & Gardner, 2014; Pestilli et al., 2011).

Our model extends a linking model of contrast discrimination (Boynton et al., 1999) to simultaneous judgments of contrast and coherence. To separate the intertwined sensory representations of these features we allowed a linear weighting of cortical areas. The weights fit by the model confirmed that the bulk of information for these simple perceptual decisions was available in V1 for contrast perception and MT for coherence. This matches with previous results implicating monkey MT in judgments about motion (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Katz et al., 2016; Newsome & Paré, 1988). But the weights also revealed that other areas could play an important role in perception by suppressing correlated signals about un-cued features in the readout. Our linking model is also specific to the random dot stimulus we chose. Changing the dot density (Smith et al., 2006) or aperture size (Ajina et al., 2015; Becker et al., 2008; Costagli et al., 2014) can result in decrements or zero response to increasing coherence, which would necessitate a linking model specific to those stimulus properties. We chose our stimulus size, dot density, and dot speed with these concerns in mind (for additional discussion of how stimulus properties affect the coherence response see Birman and Gardner (2018).

The linking model we developed held only if sensory noise was modeled as additive but not if variability increased in proportion with firing rate (Softky & Koch, 1993). Additive noise appears repeatedly in the literature using linking models (Boynton et al., 1999; Hara & Gardner, 2014; Pestilli et al., 2011; Sapir et al., 2005), in purely psychophysical approaches (Gorea & Sagi, 2001; Neri, 2010, 2018), and in measurements of population activity from voltage sensitive dyes (Chen et al., 2006). In our results, the Poisson noise model failed because it combined increasing noise with response functions that saturate (Birman & Gardner, 2018); either of which alone predicts the cumulative normal form of the psychometric functions and a Weber-law like effect at increasing base stimulus strengths. This result suggests that the noise that limits perceptual behavior is not the individual variability in firing rate of single neurons, which presumably is averaged out across a population, but a correlated source of variability which is not dependent on response amplitude.

3.4.2 Flexible readout of sensory representations

Our results demonstrate that sensory change due to attention does not transform the sensory representation directly into a form that can be used to drive motor responses. Instead, switching from reporting one stimulus property to another must change the readout (i.e. weighting of connections), which may begin to occur in sensory cortices (Ruff & Cohen, 2017) but must also extend beyond them. One possible role for the response gain is that it works together with changes in readout,

acting, as we calculated, as a weak form of sensory enhancement. Recent theoretical and experimental results suggest that such changes might improve the ability of a linear readout to differentiate between stimulus-driven and internal signals (Ecker et al., 2016; Rabinowitz et al., 2015; Snyder et al., 2018). These changes match with our finding that noise limiting perceptual sensitivity is due to correlated internal variability. Sensory changes might also drive responses to be more aligned with the readout dimension, effectively working together.

Although for our task the scale of sensory changes provided only a partial explanation for context-dependent behavioral reports, this need not always be the case. In the literature on visual attention there are many examples of changes in sensory representation as a result of task demands (Carrasco, 2011). We interpret these results and our own as falling within a continuum where task demands are implemented by complementary changes in sensory representation and sensory readout. Sensory effects that can alone account for behavioral changes would be at one end of this continuum. For example, measurements of changes in spatial tuning (Kay, Weiner, & Grill-Spector, 2015; Klein et al., 2014; Vo et al., 2017) may underlie bottom-up biases in spatial perception (Klein et al., 2016), additive shifts in response (Buracas & Boynton, 2007; Li et al., 2008; Murray, 2008) can be used by efficient selection mechanisms (Chen & Seidemann, 2012; Hara & Gardner, 2014; Pestilli et al., 2011) to account for perceptual threshold enhancement, and changes in correlation structure during focal spatial attention (Mitchell et al., 2009) can be sufficient to explain changes in perceptual sensitivity (Cohen & Maunsell, 2010, 2009). These spatial attentional effects may reflect the combination of a fixed sensory readout combined with changes in representation which select (Carrasco, 2011) and align (Ruff, Ni, & Cohen, 2018) relevant signals while suppressing others.

Our results suggest that judgments of motion visibility rely on both a context-dependent readout and changes in sensory representation, putting our task in a different part of the continuum described above. Relying on flexible readout could help maintain adaptability in the face of uncertain task demands. It is possible that given enough time and task-consistency observers could have shifted their cortical implementation to solve our task using a fixed readout. This could be done by learning to pre-select relevant sensory representations, saving computational cost and speeding decision making. Similarly, sensory representations may be kept stable for visual features that are relevant for a variety of behaviors. For example, scene gist is known to survive inattention, both perceptually (Li et al., 2002) and as information that can be decoded from BOLD signal measurements of visual cortex (Peelen et al., 2009). How the human brain implements task demands may depend not only on the form of sensory representation, the precise task demands, and the extent of learning, but also on the associated computational costs (Gardner, 2019). Flexible readout might be implemented by parts of prefrontal cortex which re-represent visual information in a context-dependent manner (Bugatus et al., 2017), using dynamical properties that can selectively integrate different features of sensory stimuli (Mante et al., 2013). Engaging these mechanisms requires resources to represent and process aspects of sensory stimuli that may not be behaviorally relevant. Changing sensory

representations and using a fixed readout may instead reflect a computationally efficient solution where the visual system no longer has to contend with representing irrelevant stimulus information. In general, the complimentary mechanics of sensory change and change in readout are both essential tools for the human brain, allowing us to meet the demands imposed by daily life where constant shifts in attention are necessary to achieve our goals.

Chapter 4

Aim 2: Comparing different forms of selective visual attention on a shared perceptual metric

4.1 Introduction

The demands of everyday life require us to flexibly shift our attention between many different aspects of the visual world. When researchers operationalize such behaviors they often ask observers to select information either from a specific location (spatial selection) or using a particular feature (feature-based selection). The difference between these forms of selection seems intuitive at first glance and may reflect a real physiological difference in how selection is implemented by the brain. It is also possible that space and feature are treated identically by the visual system and that selection of information is a generic computation.

Evidence exists to suggest that selection by location and feature differ in subtle ways. One large difference is that feature-based attention operates as a global (spatially non-specific) form of selection, e.g. for motion direction and color (Saenz et al., 2002). The implementation of this global vs. local selection may explain why selection by location operates at a slightly faster rate than selection by feature (Liu, Stevens, & Carrasco, 2007a; Hillyard & Münte, 1984; Harter, Aine, & Schroeder, 1982). Spatial selection is also primary in some ways (Soto & Blanco, 2004; Tsal & Lavie, 1988). This has been demonstrated by showing that subjects are more likely to recall letters near an attended location over letters of a similar color at distant locations (Tsal & Lavie, 1988) as well as by showing that errors in letter recall occur for spatial neighbors but not for color-matched distant neighbors (Snyder, 1972). It has also been proposed that to create objects out of visual properties they must be bound together by location (Treisman & Gelade, 1980), making location primary over

other features. Other comparisons have suggested that spatial and featural selection are impacted by perceptual noise in different ways (Ling, Liu, & Carrasco, 2009). All of these results indicate that under the right conditions small differences exist between these two basic forms of selection.

Are the small differences between spatial and feature-based selection a result of a difference in how selection is implemented in the brain? We sought to answer this question by building a stimulus in which selection by location, color, and motion direction can be compared on a shared perceptual metric. We use this stimulus in two tasks, a perceptual averaging task and a working memory estimation task. In both tasks observers are asked to select information either by spatial location or according to a stimulus property (either motion direction or color) while reporting about the other property. In both data sets we show that sensitivity is nearly identical between each form of selection and that any differences in performance are accounted for by changes in bias to the dot patches which were supposed to be ignored. We suggest possible implementations by which a common computation could select sensory representations and account for the behavioral observations.

4.2 Methods

4.2.1 Observers

In total 15 observers were subjects for the experiments (8 female, 7 male). All observers except one (who was an author) were naive to the intent of the experiments. Three observers were excluded during the initial training sessions due to an inability to maintain appropriate fixation (see eye-tracking below). Procedures were approved in advance by the Stanford Institutional Review Board on human participants research and all observers gave prior written informed consent before they participated in the experiment. When necessary, observers wore corrective lenses to correct their vision to normal. Observers were filtered prior to inclusion based on self-reported color vision and tested for colour vision deficits using the Ishihara test (Ishihara, 1987), one observer had to be excluded based on the test results.

Seven of the observers completed the averaging task, completing on average 988 trials (range 280 - 1475) over a series of ninety minute session. Five of the observers completed the estimation task, completing on average 2290 trials (range 1770 - 2613) over a series of sixty minute sessions.

4.2.2 Hardware setup for stimulus and task control

Visual stimuli were generated using MATLAB (The Mathworks, Inc.) and MGL (Gardner et al., 2018a). Stimuli were displayed on a 22.5 inch VIEWPixx LCD display (resolution of 1900x1200, refresh-rate of 120 Hz) and responses collected via keyboard. Output luminance and spectral luminance distributions were measured for the LCD display with a PR650 spectrometer (Photo Research, Inc.). The gamma table for each display was dynamically adjusted at the beginning of each trial to

linearize the luminance display such that the full resolution of the 8-bit table could be used to display the maximum contrast needed. The luminance spectra were used to compute a transformation matrix from the CIELAB color space to the RGB output of the screen, such that the a* and b* dimensions could be separately manipulated without changing the luminance (L*). Other sources of light were minimized during behavior. Observers used a circular volume controller to submit their responses in angle space (Powermate USB, Griffin Audio).

4.2.3 Eye tracking

Eye-tracking was performed using an infrared video-based eye-tracker at 500 Hz (Eyelink 1000; SR Research). Calibration was performed throughout each session to maintain a validation accuracy of less than 1 degree average offset from expected using a thirteen-point calibration procedure. Trials were initiated by fixating the central cross for 300 ms and canceled on-line when an observers eye position moved more than 1.5 degree away from the center of the fixation cross for more than 300 ms. During training and before data collection observers were excluded from further participation if we were unable to calibrate the eye tracker to an error of less than 1 degree of visual angle or if their canceled trial rate did not drop to near zero.

4.2.4 Experimental design

Averaging task

Stimuli consisted of two pairs of dot patches, to the left and right of a central fixation cross (0.5 x 0.5 deg). The dot patches were circular regions centered 8 degrees eccentric with a diameter of 10 deg, covering from 3 to 13 deg along the horizontal axis and -5 to +5 deg along the vertical axis. Each side had two dot patches was filled with moving dots (0.2 dots / deg^2, per set, 0.3 deg diameter). Dots within a patch were given an identical color and moved in the same direction at 3.5 deg / s. Dots were 'alive' for 0.25 s before vanishing and reappearing immediately at a new random location. One patch on each side was colored yellow and one blue (90 deg and 270 deg, in a* b* space).

On each trial in the averaging task observers were asked to report the average motion direction of two dot patches (Fig. 4.1. Before each set of 20 trials observers were told how they would select the two dot patches with the phrase "cue side" or "cue color" shown at fixation. Each trial was initiated by the observer fixating the central cross for 0.5 s. This was followed by a 0.75 s cue, either a line to the left or right or a miniature patch of colored dots. The feature instructed the observers about which two dot patches they would need to average: either the two on the left, on the right, or the two yellow or blue patches (one on the left and one on the right). A 0.75 s delay followed. During the stimulus period the dot patches began moving coherently in random directions. The target patches were constrained to be less than 135 degrees apart, to avoid confusion about the correct response (when 180 degrees apart, two possible answers are correct). Observers were shown the stimulus for

a variable duration of 0.25 to 0.75 s, then allowed unlimited time to rotate the response wheel and make a response. Feedback was given by showing the actual average motion direction (see Figure). Each trial was followed by a brief inter-trial interval (0 - 2 s, uniformly distributed).

Psychophysical distance

We report all of our results according to the normalized psychophysical distance between angles in motion direction and color space, rather than the physical units. This is based on a recent result showing that in working memory estimation tasks, correctly taking into account the psychophysical distance is critical to correctly interpreting data (Schurgin, Wixted, & Brady, 2018). In brief, the motivation for this scaling is that beyond a certain degree distance the "psychophysical" distance becomes compressed. If you are trying to compare North and Northeast to East, it's easy to tell that NE and E are closer. But if you are trying to quickly compare N and NE to S, the task becomes more difficult and there is little difference between that comparison and N and NE to SW. The re-scaling sets the 'distance' between two angles to the normalized sensitivity according to the comparison task just described. Without this re-scaling of distances it's easy to mistake poor sensitivity for a high lapse rate. The authors of Schurgin et al. (2018) as well as others (Bays, 2014) convincingly demonstrate that in fact lapse rates are consistently low in difficult working memory tasks. When observers appear to be guessing they are actually making low-probability choices with high-confidence. For our purposes we approximated the psychophysical scaling by fitting a sigmoidal function to data available in that paper:

$$d(x) = 1.1\frac{x^{1.5}}{x^{1.5} + 35^{1.5}} \tag{4.1}$$

This equation transforms the distance x between two motion directions or colors (in a* b* angle space) to the normalized psychophysical distance d. Unlike degree distance, the normalized psychophysical is set up such that an observer perceives the difference between 0 and 0.5 (0 and 31 deg, respectively) as equal to the difference between 0.5 and 1 (31 and 180 deg).

Estimation task

On each trial in the estimation task observers were asked to report about either the color or motion direction of a single dot patch. Before each block of 40 trials observers were told which feature would be reported with either the phrase "report color" or "report direction" appearing on the screen. Key to the task was that although observers ultimately reported about only one dot patch they could be cued to remember just that patch, or multiple patches, during a brief delay period. Each trial consisted of the following sequence (Fig. 4.4): a fixation period (0.5 s), a pre-cue indicating which patches needed to be memorized (0.75 s), an inter-stimulus interval (0.75 s), stimulus presentation (0.25 s), a delay (1 s), a post-cue resolving which dot patch should be reported (0.75 s) and then

unlimited time to report a response. The inter-trial interval was 0 - 2 s, sampled randomly from a uniform distribution. The stimulus duration (0.25 s) was chosen based on the averaging task to ensure that the task was difficult for participants but not impossible.

Estimation task data analysis

To analyze the results of the estimation task we fit a modified version of the "target confusability competition" model from Schurgin et al. (2018). The model is based on the idea that noisy internal channels are independently competing to represent a stimulus (Fig. 4.4a). On each trial the model proceeds in two steps. First, the stimulus (or stimuli) are encoded by the channels, setting their mean response. The tuning profile of each channel comes from the normalized psychophysical distance (Eq. 4.1). As an example, the encoding step the response of a small set of channels (Fig. 4.4a) to a stimulus with an angle of zero is shown (Fig. 4.4b). Each channels response is distributed as follows:

$$C_\theta(x) = \mathcal{N}(\mu = \alpha d(x - \theta), \sigma = 1) \tag{4.2}$$

Where θ is the preferred orientation for that channel and d is Eq. 4.1. α is an amplitude parameter which controls the scaling of the response and is the only free parameter in the model.

Once a stimulus is encoded by a set of responses the second step in the model is to find the observer's behavioral response by taking the maximum response over the channels. Because each channel has independent normally-distributed noise, the likelihood of each channel winning can be computed as the conditional probability of the channel exceeding all of the other channels. We approximate this likelihood by numerically integrating the likelihood over channel responses, as follows:

$$\mathcal{L}(\theta) = \int_0^\infty P(x_\theta = a) \prod_{j \neq \theta} P(x_j < a) \mathrm{d}a \tag{4.3}$$

Where $P(x)$ is normally distributed, according to (Eq. 4.2). To compute the full likelihood distribution we evaluate Eq. 4.3 at all values of θ.

Because the response calculation is analogous to signal detection the α parameter in Eq. 4.2 is actually the sensitivity of the channel (i.e. d'). We fit this free parameter to the responses of each observer by maximizing the likelihood of the observed data using Bayesian adaptive direct search (Acerbi & Ma, 2017) in MATLAB.

We performed the model fitting step in such a way as to separate an observer's bias (i.e. likelihood of responding about the incorrect dot patch) from their sensitivity (i.e. their variability in response quality, for a given dot patch). We modeled the observer's trial-by-trial response as a combination of a likelihood function for each stimulus patch (Eq. 4.3) with a set of bias parameters.

$$\mathcal{L}(\theta) = \beta_{target}\mathcal{L}_{target} + \beta_{side}\mathcal{L}_{side} + \beta_{feature}\mathcal{L}_{feature} + \beta_{distractor}\mathcal{L}_{distractor} \tag{4.4}$$

Where the terms *target*, *side*, *feature*, and *distractor* correspond to the dot patch that was reported on the trial, the patch on the same side, the patch on the opposite-side with matched-feature, and the patch on the opposite-side with mismatched-feature, respectively (Fig. 4.5). The actual β values were constrained so that $\beta_{target} + \beta_{side} + \beta_{feature} + \beta_{distractor} = 1$, by calculating them from three intermediate values:

$$\beta_{target} = \beta_s * \beta_f \tag{4.5}$$

$$\beta_{side} = \beta_s * (1 - \beta_f) \tag{4.6}$$

$$\beta_{feature} = (1 - \beta_s) * (1 - \beta_d) \tag{4.7}$$

$$\beta_{distractor} = (1 - \beta_s) * \beta_d \tag{4.8}$$

Where β_s, β_f, and β_d are each constrained to the range [0,1]. Setting $\beta_s = 1$ and $\beta_f = 1$ means that the observer is always choosing the target and never incorrectly being biased to respond about the other three dot patches (i.e. $\beta_{target} = 1$).

In sum, we fit four sensitivity parameters (α) and three bias parameters (β) for the data set in which observers selected by location or color (and reported motion direction) and separately for the data set in which they selected by location or motion direction (and reported color).

4.2.5 Implementing attention in a channel linking model

The channels in the behavioral model described above have tuning which, by definition, matches the behavior. In reality, the psychophysical scaling is a result of the readout process from neurons tuned with much sharper functions (Bays, 2014, 2019). To explore how attention might change the responses of neurons we explored how to connect sharp tuning functions, such as those neurons might have, to the psychophysical space described above.

In this simulation we assumed that channels had a Von Mises tuning with a relatively sharp profile (Fig. 4.7a). As before the response of each channel had independent Gaussian noise at the time of stimulus encoding. To read out from these channels we computed:

$$\hat{\theta} = \arg\max_\theta \sum_i r_i \overline{r}_i(\theta) \tag{4.9}$$

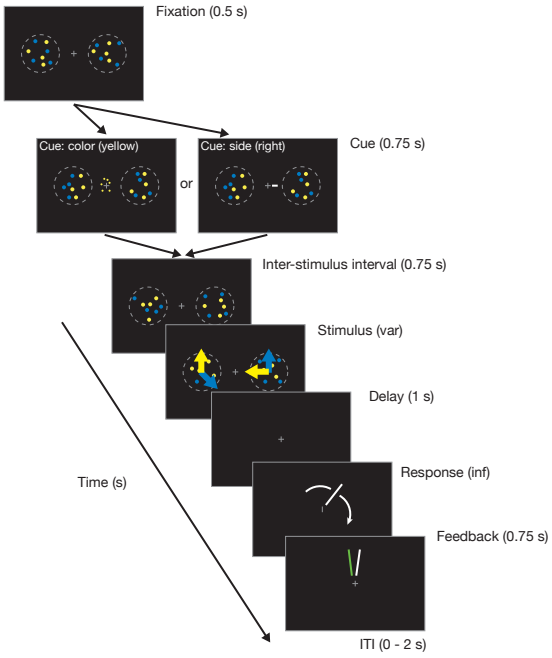

Figure 4.1: Motion direction averaging task. Observers were asked to select two out of four random dot patches and average their motion direction. Observers initiated trials by fixating a central cross, causing the two dot patches to appear with incoherent motion. A cue indicated whether they should select the left or right patches (spatial selection) or the yellow or blue ones (feature-based selection). After a brief delay the dot patches each began moving in random directions, before vanishing again for a second short delay. Observers used a rotating wheel to report the *average* direction of motion for the two dot patches they were asked to select. Feedback was given by indicating the true average motion direction.

Where r is the response of each channel to the stimulus and $\overline{r}$ is the response of that channel in the absence of noise.

To simulate different models of attention we either applied a gain to the noisy channel responses or changed the set of mean channel tuning values. When 'shifting' the tuning it is necessary to also shift the expected readout.

4.3 Results

We characterized human perceptual sensitivity to the average motion direction of two dot patches, while asking observers to select the two patches either based on their common location or a shared feature (Fig. 4.1. To measure perceptual sensitivity we recorded each observer's estimation error relative to the true average motion direction. We found that whether observers selected the two dot patches by spatial location (left or right) or by feature (yellow or blue), their estimation errors

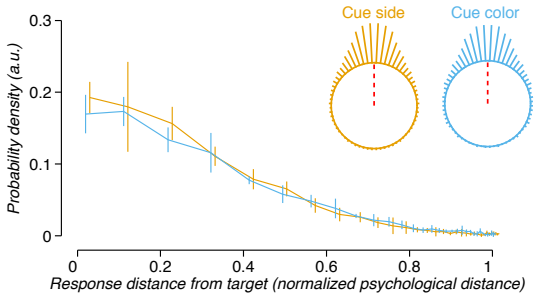

Figure 4.2: Estimation error during the averaging task. A histogram displaying the average proportion of responses at each distance from the true average motion direction (0) is shown, averaged across observers. Selection by spatial location (i.e. averaging the two patches on the right or left) is shown in yellow, and selection by color (i.e. averaging the two yellow or blue patches) is shown in blue. The two inset plots show the same histogram but in a circular space, with a red dashed line indicating the true average. Note that the x-axis has been re-scaled from degrees to psychophysical distance, see Methods.

remained nearly identical (Fig. 4.2). Consistent with the task design we found that giving observers a longer stimulus (Fig. 4.3a) or a smaller angle difference between the two dot patches (Fig. 4.3b) improved sensitivity slightly.

The averaging task demonstrates that if differences in selection exist they are small and may depend in specific ways on the context of particular tasks. For example, observers might be biased in different ways to the irrelevant dot patches according to how they selected from the stimulus. We refers to such errors as bias, while referring to the precision of reports as sensitivity. We next sought to design a task which could differentiate between changes in bias (which dot patch was reported) and sensitivity (how precise the reports were).

The estimation task uses the same stimulus as the averaging task, but we now asked observers to recall the properties of a single dot patch (rather than the average of two). Observers were cued in different ways to force them to select the stimulus according to different features. To set a baseline for performance we cued observers to the exact target they would later report (Cue 1, Fig. 4.4). In the most difficult case (Cue 4: Distributed, Fig. 4.4) observers memorized the directions of all four potential targets and were only post-cued after a brief delay about which target should be reported. In the two critical selection conditions observers were asked to memorize either the motion directions of the two patches on the left or right (Cue 2: Side, Fig. 4.4) or the two yellow or blue patches (Cue 2: color, Fig. 4.4), in the same manner as in the averaging task. In both of these conditions a Post-Cue was used to reveal which of the memorized dot patches had to ultimately be reported. Note that we had observers perform this task in two ways: once selecting by either location (left/right) or color (yellow/blue) and reporting motion direction, as shown in the figure, but also while selecting by either location or motion direction (up/down) and reporting color (see e.g. Fig. 4.6).

To understand the data we collected from the estimation task we needed to decompose bias

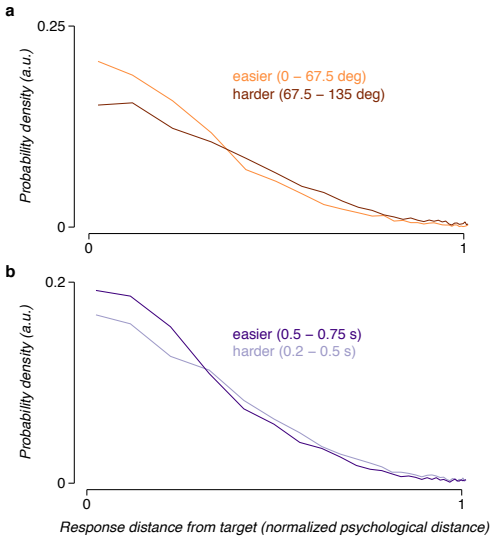

Figure 4.3: Averaging difficulty is controlled by stimulus duration and angle distance between patches. (a) The average estimation error across observers is shown for a median split of the angle difference between the two dot patches that were averaged. (b) As in (a) for a median split of stimulus duration.

from sensitivity. To do this we employed a simple model of perceptual sensitivity which fits two parameters for each condition (Fig. 4.5, see Methods for details). The model encodes the stimulus into a set of independent channels with Gaussian-distributed noise (Fig. 4.5a). A single parameter scales the responses of these channels to fit the sensitivity of an observer (Fig. 4.5b). To obtain the likelihood of an observer's responses the maximum is taken over the channel responses, resulting in a likelihood distribution (Fig. 4.5c). To decompose sensitivity from bias we allowed the channels to separately encode each of the dot patches with a separate sensitivity, then weighted those likelihood functions to create a mixed distribution from which actual trail-by-trial responses would be sampled (Fig. 4.5e). We fit all seven parameters of this model (four β and three d' parameters) to maximize the likelihood of predicting the responses of each observer.

We compared three conditions in the estimation task, using the Cue 1 condition as a baseline for performance (Fig. 4.6 and found that all of the variability in performance between conditions was accounted for by the bias parameters. Our main goal was to see whether during the two different forms of selection (Cue side and Cue feature) a difference in sensitivity to the target dot patch emerged. We did not find this to be the case, confirming the finding from the perceptual averaging task (left panels, orange curves are all identical, Fig. 4.6b-d). A direct comparison of these sensitivity parameters between conditions and against the same parameter in the Cue 4 condition showed no differences (Fig. 4.6e) between forms of selection, but a substantial advantage to selecting two patches compared to selecting only four.

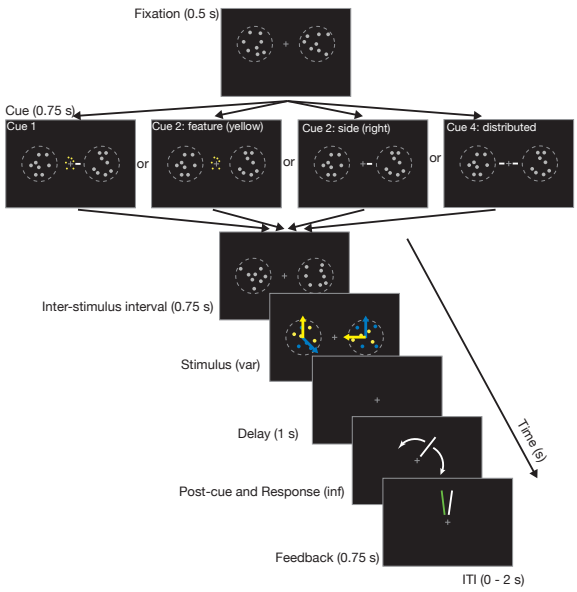

Figure 4.4: Estimation task. The task is shown where the cues were side (left or right) and colors (yellow or blue) and observers reported motion direction, but we also ran the reverse where the cues were side (left or right) and motion direction (up or down) and observers reported the color. Observers began each trial by fixating a central cross (Fixation). A pre-cue (Cue) was then shown to indicate to observers which of the four dot patches they should memorize. A brief delay (Inter-stimulus interval) gave observers time to prepare. The dots then became colored and coherent for a variable duration (Stimulus). Finally after another brief delay (Delay) observers were shown a second cue which was used to disambiguate the target stimulus (Post-cue). For example, if the observer was cued to remember the two stimuli on the right, the post-cue might be blue to indicate that of the two stimuli memorized only the motion direction of the blue dot patch on the right should be reported. Observers were given unlimited time to respond (Response) and received feedback before the next trial (Feedback).

While we found no differences in the sensitivity parameters we did find substantial differences in how biased observers were to report about the incorrect dot patches in different conditions (right panels, Fig. 4.6b-d). When remembering all four dot patches (Cue 4) we found that observers were only able to report the direction of the correct dot patch half the time. They were nearly equally likely to confuse the patch we asked them to report with the patch on the same side. A small percentage of the time (10.5%) observers reported about the feature matched stimulus on the wrong side. Performance improved substantially in both conditions where observers were cued to remember just two of the four dot patches. In the Cue side conditions we found that observers were only biased to report about the dot patch on the same side. In the Cue feature conditions we found that observers occasionally reported about all of the irrelevant dot patches with some frequency, indicating a difference in bias due to the form of selection.

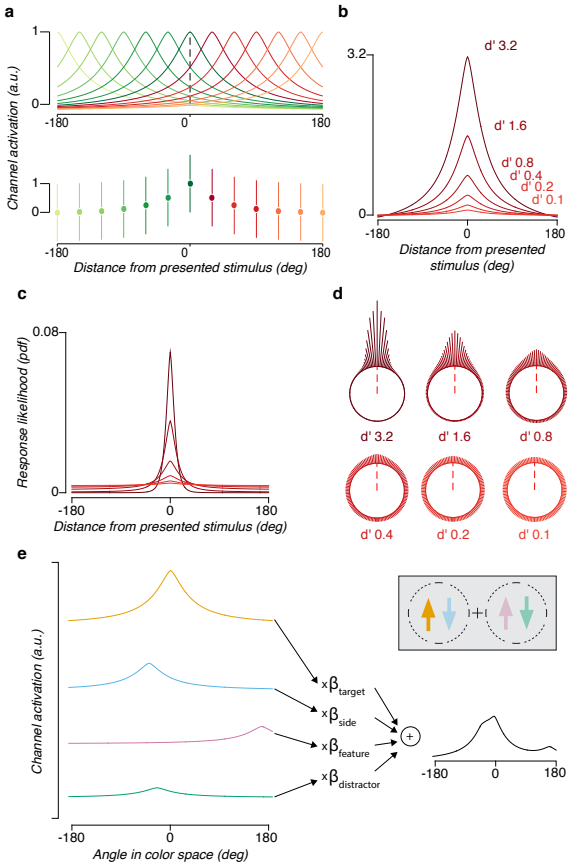

Figure 4.5: Estimation task model. The model of the estimation task is based on an existing model of working memory estimation by Schurgin, Wixted, and Brady (2018). (a) In the model independent channels encode the stimulus with a response profile defined by the psychophysical distance of each channel's preferred response and the stimulus. The channel responses are normally distributed with $\sigma = 1$. (b) A free parameter in the model controls the sensitivity of the channels (d'), which acts as a multiplicative gain on the amplitudes of each channel response. (c) To read out an estimate of the stimulus angle an observer takes the maximum response over the channels. We show here the full likelihood distribution over all angles, computed numerically (see Methods). (d) The same distributions in (c), the likelihood of response for different values of d', are shown in circular space. (e) To estimate the trial-by-trial likelihood of responses in the estimation task we fit a separate sensitivity parameter for each dot patch, indexed by its relative position to the reported target patch on that trial. The four patches are the reported patch (orange, *target*), the patch on the same side (blue, side), the patch on the opposite side with the same feature (pink, *feature*), and the patch on the opposite side with the mismatched feature (green, *distractor*). To decompose bias from sensitivity we weighted the likelihood distributions by β weights for each condition. These β weights were constrained to sum to 1 (see Methods). Once summed, these define a trial-by-trial likelihood distribution for each observer. Note that for clarity of presentation we are showing the task variant in which colors are reported, while Fig. 4.4 shows the variant in which motion directions are reported.

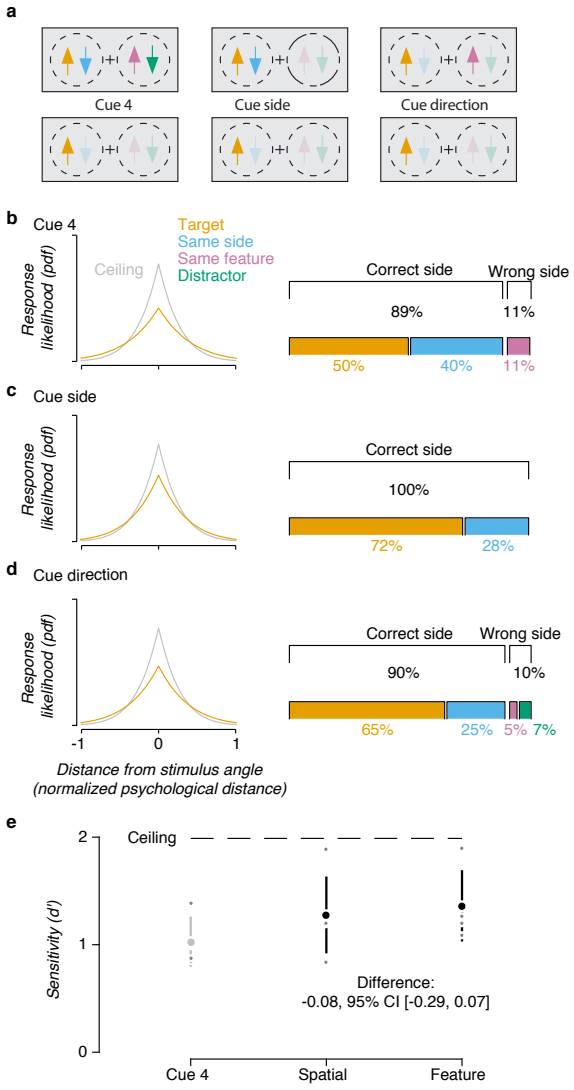

Figure 4.6: Performance in the estimation task. (a) Three of the conditions used in the experiment are shown for trials where selection was performed by the direction of motion and the report was the color. Opacity is used to indicate which dot patches are memorized in each condition and to emphasize that the response in all conditions is identical (reporting the color of a single dot patch). In Cue 4 trials an observer memorized all four colors shown and was then asked to report the color of a single dot patch, e.g. the dots moving upward on the left side (orange arrow, highlighted). In Cue 2 trials the observer either memorized the colors on one side (Cue side) and was post-cued about the direction, or memorized two dot patches moving in the same direction (Cue direction) and was post-cued about the side. (b) The model estimate of sensitivity for each of the four dot patches is shown separated from the probability of reporting about each dot patch. Observers reported about the distractor dot patch less than 2% of the time. (c-d) Conventions as in (b) for the two cue 2 conditions. (e) Confidence intervals for the *target* dot patch d' parameter are shown for each condition.

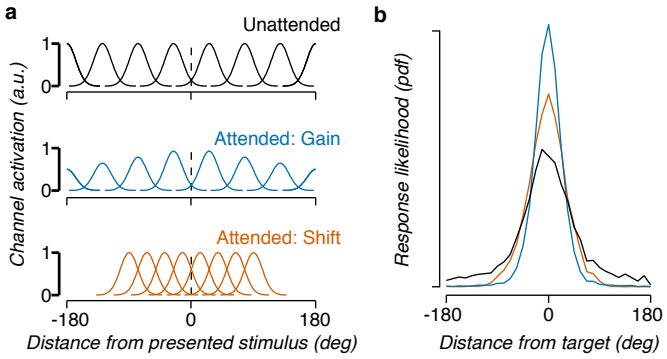

Figure 4.7: Implementations of attention in a hypothetical channel model. (a) Examples are shown of how neuron responses might change during attention. (b) Response likelihoods are shown for the different attention models, see Methods for model details.

In previous chapters we explored different possible models for how attention could be implemented in sensory representations or their readout. Here again we wanted to explore possible implementations for how neural selection might result in the behavioral changes we observed. An important step toward this goal is to describe a computational linking model which can connect the perceptual measurements described here with measurements of cortical activity. We describe here such a model and demonstrate that it can in theory account for the kinds of shifts in sensitivity and bias that could occur in our task (Fig. 4.7).

A plausible linking model has to have channels that are tuned according to functions with thinner response profiles than those described by the psychophysical scaling function (Eq. 4.1, see also Fig. 4.5a). Ultimately the size and number of these channels would need to be constrained by neural data. We modeled a small number of channels (128) with small variance (circular Von Mises distributions with $\kappa = 20$). To simulate read out from such a population during an estimation task we proceeded in two steps. First, we simulated the response to a stimulus by sampling from each channel's response to the stimulus angle, with additive Gaussian noise. Then, to decode the angle stored in the channels we compared the sampled responses to the ideal mean responses for every channel for every possible stimulus angle (Eq. 4.9, i.e. we computed the dot product of the sample vector and the mean channel response vector for each θ). The maximum of this readout step was chosen as the response for that simulated trial. We repeated this 1,000$\times$ and then plotted the distribution of response angles relative to the true stimulus angle (Fig. 4.7b).

Changes in sensitivity in such a model can be the result of different manipulations in the channels, such as multiplicative gain or shifts in tuning. Using the simulation we showed that a multiplicative gain or a shift in tuning can both result in changes to the distribution of estimated angles (blue and orange lines, Fig. 4.7a and b). Combined with the estimation task we therefore believe that this approach provides a computational linking model to connect measurements of cortical representation

to the behavioral measurements. When observers went from memorizing four dot patches to two, whether by location or feature, their perceptual sensitivity improved substantially (Fig. 4.6e). We expect that enhancement to be the result of both changes in sensory representation but also potentially changes in the readout (see Chapter 3). Combining this computational linking model with measurements of cortical activity will make it possible to evaluate whether the scale of observed changes in cortical representation are sufficient to account for the behavioral effects of selective visual selection.

4.4 Discussion

Using perceptual averaging and estimation we have demonstrated that selection by spatial location and by feature both enhance perceptual sensitivity by a similar amount. Although selection by different features does not manipulate sensitivity we did find that differences in task performance were well accounted for by changes bias. This showed that selection changes how likely observers were to report about the wrong dot patches, but did not change the strength of their encoding of the dot patches. These results make a compelling case for the hypothesis that selective visual attention has a shared neural implementation across different visual features. Our results also confirm that selection by spatial location may occur prior to selection by other features, paralleling our knowledge of the physiological structure of early visual cortex.

Many early experiments that compared different forms of visual selection came to the conclusion that spatial location is primary in some way (Liu et al., 2007a; Treisman & Gelade, 1980; Tsal & Lavie, 1988; Snyder, 1972; Hillyard & Münte, 1984; Harter et al., 1982; Soto & Blanco, 2004). In general, this idea that spatial selection precedes feature-based selection matches cortical physiology. Early visual cortex is organized retinotopically (Wandell et al., 2007) and the earliest visual areas are sensitive to specific visual features at particular retinotopic locations (Kuffler, 1953; Hubel & Wiesel, 1959, 1962). The progression from local simple features to more global complex features provides one explanation for the pattern of bias which we observed. We found that when selecting by spatial location observers were able to easily ignore the dot patches in the other visual field, consistent with an implementation preventing those dot patches from being fully processed. But to select dot patches by feature observers were required to not select by location, perhaps allowing processing to continue on those irrelevant dot patches. This possibility might explain why observers were sometimes biased to the feature mis-matched dot patches on the same and opposite side even when selecting by feature. Despite this bias, observers were nevertheless equally sensitive, i.e. their responses were just as variable, when selecting by location or by feature.

Chapter 5

Summary and conclusions

Together, these results demonstrate that human selective visual attention is made up of multiple computational components which are shared across different forms of sensory selection. These findings were possible through the use of *computational linking models* which make hypotheses about the connections between neural representation and perception explicit and testable (Barlow, 1972; Brindley, 1960; Cohen & Maunsell, 2010; Newsome et al., 1989; Pestilli et al., 2011; Hara & Gardner, 2014; Gardner, 2015).

In Aim 1 I sought to quantify the extent to which cortical changes during selective visual attention could account for perceptual changes. We chose to study the visibility of motion for this purpose. Motion visibility can be controlled by several different perceptual parameters: contrast, coherence, and duration. Having multiple stimulus properties which all manipulate the same perceptual property makes motion visibility an excellent tool to investigate how cortical changes might be connected to the perceptual enhancements during selective attention. Prior to this project nobody had laid out a full framework for how motion visibility is represented in human visual cortex. The first step was therefore to build a framework for this purpose.

I next measured how the sensory representation of motion visibility changed during directed attention and demonstrated that these changes were insufficient to account for perception. We validated that a linking model of motion visibility perception could be built, extending an existing linking model of contrast discrimination (Boynton et al., 1999). We then measured how sensory representations changed and, passing these through the linking model, showed that the scale of changes were too small to account for perception. Based on these observations we suggested that a *flexible readout* must change how signals are gated from sensory cortex into decision-related regions.

The findings in Aim 1 are consistent with the hypothesis that a substantial amount of the processing during sensory selection occurs outside of the areas thought to primarily represent sensory information. This, in turn, suggests that these computations might be largely invariant to the kind of information they receive. Put another way, selection that is implemented by flexible readout

should be similarly strong or efficient regardless of the feature selected for. In Aim 2 we sought to validate this prediction by directly comparing spatial and feature-based attention. I developed two variants of an estimation task for this purpose, one using perceptual averaging and a second working memory. Each task was designed to measure how the strength of sensory selection changes according to the feature being selected. I showed with these tasks that there are only subtle differences between spatial selection and feature-based selection whether by motion direction or by color. Importantly, all of these small differences were well accounted for by errors in bias and not changes in sensitivity. This suggests the fascinating possibility that different selection behaviors are all implemented by a common selection mechanism.

Bringing the findings in this dissertation together, these results hint at the possibility that selection is in large part implemented during the readout stage from sensory representation to a context-dependent one. What then is the role of small changes to the sensory representation, which clearly occur during attentional behaviors? I would suggest that these changes to sensory representations might play a role in selection, but they would be complimentary to this readout process. Much of the recent work on attention aligns with this idea that attention is a two-step process in which sensory changes work together with the readout process to improve perceptual abilities (Pestilli et al., 2011; Ruff & Cohen, 2018; Snyder et al., 2018; Rabinowitz et al., 2015).

This hypothesis appears to echo the ideas of a late selection account of selective attention (Deutsch & Deutsch, 1963), but they differ in crucial ways. In the late selection theory sensory processing runs to completion irregardless of an organism's behavioral goals. This matches with the results in Chapter 3, where I showed that observers retain information about unattended features. This would also be consistent with similar results for visual processing of faces and scenes (Li et al., 2002; Reddy, Reddy, & Koch, 2006). But in each of these cases the remaining perceptual representations are impoverished. Even if processing is going to 'completion', the sensory representation has degraded considerably by the time the readout process can be shifted to it. In contrast to a late selection account, what I have shown here is more in line with the idea of a continuous or graded selection process.

One explanation for why sensory selection might occur continuously during sensory processing and readout is that this balances the efficiency of processing against behavioral flexibility. Attention is often suggested to be part of a solution to the high cost of neural activity (Lennie, 2003). Shifts in tuning which allocate additional processing to attended features would seem to correspond to such a theory. But to effect such a change requires intervening on the sensory representation directly at the cost of lost selectivity for unattended stimuli (Mack & Rock, 1998). These feature-specific computations are also potentially complex: is the visual cortex structured in such a way that any arbitrary feature can be enhanced? One way to reconcile the need for efficiency against the complexity of implementation is to assume that attention is not a static computation, but one that is modified with experience. Give sufficient time and consistency in a task, it's possible that the human

observers in Chapter 3 might have been able to learn a sensory-change implementation to solve that task. In theory such an implementation should be more efficient compared to always representing the entire stimulus and then selecting out the important information at the last step. One potential way to study this further would be to compare attentional behaviors in animals, especially mice and non-human primates, with humans. We know that each of these model organisms learns in vastly different ways (Birman & Gardner, 2015). Finding differences in their sensory selection behaviors may clue us into the different ways in which selection can be implemented in the brain.

Bibliography

Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M., & Gardner, J. L. (2016). Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences*, *113*(25), E3548–E3557.

Acerbi, L. & Ma, W. J. (2017). Practical bayesian optimization for model fitting with bayesian adaptive direct search. In I Guyon, U. V. Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, & R Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 1836–1846). Curran Associates, Inc.

Ajina, S., Kennard, C., Rees, G., & Bridge, H. (2015). Motion area V5/MT+ response to global motion in the absence of V1 resembles early visual cortex. *Brain*, *138*(Pt 1), 164–178.

Albrecht, D. G. & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *J. Neurophysiol. 48*(1), 217–237.

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. *J. Neurophysiol. 102*(5), 2704–2718.

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, *20*(2), 870–888.

Aspell, J. E., Tanskanen, T, & Hurlbert, A. C. (2005). Neuromagnetic correlates of visual motion coherence. *Eur. J. Neurosci. 22*(11), 2937–2945.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci. 7*(5), 358–366.

Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., & Malach, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *J. Neurophysiol. 87*(6), 3102–3116.

Baker, D. H. & Wade, A. R. (2017). Evidence for an optimal algorithm underlying signal combination in human visual cortex. *Cereb. Cortex*, *27*(1), 254–264.

Baldauf, D. & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, *344*(6182), 424–427.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, *1*(4), 371–394.

Barlow, H. B., Fitzhugh, R, & Kuffler, S. W. (1957). Change of organization in the receptive fields of the cat's retina during dark adaptation. *J. Physiol. 137*(3), 338–354.

Baruch, O. & Yeshurun, Y. (2014). Attentional attraction of receptive fields can explain spatial and temporal effects of attention. *Vis. cogn. 22*(5), 704–736.

Bays, P. M. (2019). *Correspondence between population coding and psychophysical scaling models of working memory*.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *J. Neurosci. 34*(10), 3632–3645.

Beauchamp, M. S., Cox, R. W., & DeYoe, E. A. (1997). Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J. Neurophysiol. 78*(1), 516–520.

Becker, H. G. T., Erb, M., & Haarmeier, T. (2008). Differential dependency on motion coherence in subregions of the human MT+ complex. *Eur. J. Neurosci. 28*(8), 1674–1685.

Bex, P. J. & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis. 19*(6), 1096–1106.

Birman, D. & Gardner, J. L. (2019). A flexible readout mechanism of human sensory representations. *Nat. Commun. 10*(1), 3500.

Birman, D. & Gardner, J. L. (2018). A quantitative framework for motion visibility in human cortex. *J. Neurophysiol.*

Birman, D. & Gardner, J. L. (2015). Parietal and prefrontal: categorical differences? *Nat. Neurosci. 19*(1), 5–7.

Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. *Neuroimage*, *14*(4), 817–826.

Bliss, C. I. (1934). THE METHOD OF PROBITS—A CORRECTION. *Science*, *79*(2053), 409–410.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci. 16*(13), 4207–4221.

Boynton, G. M., Engel, S. A., & Heeger, D. J. (2012). Linear systems analysis of the fMRI signal. *Neuroimage*, *62*(2), 975–984.

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Res. 39*(2), 257–269.

Braddick, O. J., O'Brien, J. M., Wattam-Bell, J, Atkinson, J, Hartley, T, & Turner, R. (2001). Brain areas sensitive to coherent visual motion. *Perception*, *30*(1), 61–72.

Brindley, G. S. (1970). Central pathways of vision. *Annu. Rev. Physiol. 32*, 259–268.

Brindley, G. S. (1960). Physiology of the retina and the visual pathway.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S, & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci. 13*(1), 87–100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci. 10*(6), 1157–1169.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci. 12*(12), 4745–4765.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proc. Natl. Acad. Sci. U. S. A. 107*(1), 361–365.

Bugatus, L., Weiner, K. S., & Grill-Spector, K. (2017). Task alters category representations in prefrontal but not high-level visual cortex. *Neuroimage*, *155*, 437–449.

Buracas, G. T. & Boynton, G. M. (2007). The effect of spatial attention on contrast response functions in human visual cortex. *J. Neurosci. 27*(1), 93–97.

Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res. 33*(2), 261–304.

Busse, L., Wade, A. R., & Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, *64*(6), 931–942.

Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage*, *23 Suppl 1*, S220–33.

Carandini, M. & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nat. Rev. Neurosci. 13*(1), 51–62.

Cardoso, M. M. B., Sirotin, Y. B., Lima, B., Glushenkova, E., & Das, A. (2012). The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. *Nat. Neurosci. 15*(9), 1298–1306.

Carrasco, M, Penpeci-Talgar, C, & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Res. 40*(10-12), 1203–1215.

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Res. 51*(13), 1484–1525.

Carrasco, M. & Barbot, A. (2018). Spatial attention alters visual appearance. *Curr Opin Psychol*, *29*, 56–64.

Chelazzi, L, Duncan, J, Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol. 80*(6), 2918–2940.

Chen, Y. & Seidemann, E. (2012). Attentional modulations related to spatial gating but not to allocation of limited resources in primate V1. *Neuron*, *74*(3), 557–566.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat. Neurosci. 9*(11), 1412–1420.

Chen, Y., Palmer, C. R., & Seidemann, E. (2012). The relationship between voltage-sensitive dye imaging signals and spiking activity of neural populations in primate V1. *J. Neurophysiol. 107*(12), 3281–3295.

Cheng, K, Hasegawa, T, Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *J. Neurophysiol. 71*(6), 2269–2280.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am. 25*(5), 975–979.

Cohen, E. H. & Tong, F. (2013). Neural mechanisms of object-based attention. *Cereb. Cortex*, *25*(4), 1080–1092.

Cohen, M. R. & Maunsell, J. H. R. (2010). A neuronal population measure of attention predicts behavioral performance on individual trials. *J. Neurosci. 30*(45), 15241–15253.

Cohen, M. R. & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci. 12*(12), 1594–1600.

Cohen, M. R. & Maunsell, J. H. R. (2011). Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron*, *70*(6), 1192–1204.

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *J. Neurophysiol. 75*(3), 1306–1308.

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. *J. Neurosci. 22*(5), 1994–2004.

Costagli, M., Ueno, K., Sun, P., Gardner, J. L., Wan, X., Ricciardi, E., ... Cheng, K. (2014). Functional signalers of changes in visual stimuli: cortical responses to increments and decrements in motion coherence. *Cereb. Cortex*, *24*(1), 110–118.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci. 16*(6), 763–770.

Dale, A. M. & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain Mapp. 5*(5), 329–340.

Dale, A. M., Fischl, B, & Sereno, M. I. (1999). Cortical surface-based analysis. i. segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179–194.

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

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

Deutsch, J. A. & Deutsch, D. (1963). Attention: some theoretical considerations. *Psychol. Rev. 70*(1), 80.

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

Dupont, P, Orban, G. A., De Bruyn, B, Verbruggen, A, & Mortelmans, L. (1994). Many areas in the human brain respond to visual motion. *J. Neurophysiol. 72*(3), 1420–1424.

Ecker, A. S., Denfield, G. H., Bethge, M., & Tolias, A. S. (2016). On the structure of neuronal population activity under fluctuations in attentional state. *J. Neurosci. 36*(5), 1775–1789.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *J. Exp. Psychol. Hum. Percept. Perform. 10*(1), 32–39.

Eriksen, C. W. & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Percept. Psychophys. 12*(2), 201–204.

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-Dependent representation of a size illusion in human V1. *Curr. Biol. 18*(21), 1707–1712.

Field, G. D. & Chichilnisky, E. J. (2007). Information processing in the primate retina: circuitry and coding. *Annu. Rev. Neurosci. 30*, 1–30.

Foley, J. M. & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Res. 21*(7), 1041–1053.

Friston, K. J., Josephs, O, Rees, G, & Turner, R. (1998). Nonlinear event-related responses in fMRI. *Magn. Reson. Med. 39*(1), 41–52.

Fründ, I., Wichmann, F. A., & Macke, J. H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. *J. Vis. 14*(7).

Fründ, J., McCann, K. S., & Williams, N. M. (2016). Sampling bias is a challenge for quantifying specialization and network structure: lessons from a quantitative niche model. *Oikos*, *125*(4), 502–513.

Gardner, J. L. (2015). A case for human systems neuroscience. *Neuroscience*, *296*, 130–137.

Gardner, J. L. (2019). Optimality and heuristics in perceptual neuroscience. *Nat. Neurosci. 22*(4), 514–523.

Gardner, J. L., Sun, P., Waggoner, R. A., Ueno, K., Tanaka, K., & Cheng, K. (2005). Contrast adaptation and representation in human early visual cortex. *Neuron*, *47*(4), 607–620.

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *J. Neurosci. 28*(15), 3988–3999.

Gardner, J. L., Merriam, E. P., Schluppeck, D., & Larsson, J. (2018a). MGL: visual psychophysics stimuli and experimental design package. *Zenodo*.

Gardner, J. L., Merriam, E. P., Schluppeck, D., Besle, J., & Heeger, D. J. (2018b). Mrtools: analysis and visualization package for functional magnetic resonance imaging data. *Zenodo*.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *J. Cogn. Neurosci. 17*(3), 507–517.

Gold, J. I. & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci. 30*, 535–574.

Gorea, A & Sagi, D. (2001). Disentangling signal from noise in visual contrast discrimination. *Nat. Neurosci. 4*(11), 1146–1150.

Haines, R. F. (1991). A breakdown in simultaneous information processing. In G. Obrecht & L. W. Stark (Eds.), *Presbyopia research: from molecular biology to visual adaptation* (pp. 171–175). Boston, MA: Springer US.

Hammett, S. T., Smith, A. T., Wall, M. B., & Larsson, J. (2013). Implicit representations of luminance and the temporal structure of moving stimuli in multiple regions of human visual cortex revealed by multivariate pattern classification analysis. *J. Neurophysiol. 110*(3), 688–699.

Händel, B., Lutzenberger, W., Thier, P., & Haarmeier, T. (2007). Opposite dependencies on visual motion coherence in human area MT+ and early visual cortex. *Cereb. Cortex*, *17*(7), 1542–1549.

Hara, Y & Gardner, J. L. (2014). Encoding of graded changes in spatial specificity of prior cues in human visual cortex. *J. Neurophysiol. 112*(11), 2834–2849.

Hara, Y., Pestilli, F., & Gardner, J. L. (2014). Differing effects of attention in single-units and populations are well predicted by heterogeneous tuning and the normalization model of attention. *Front. Comput. Neurosci. 8*(February), 12.

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proc. Natl. Acad. Sci. U. S. A. 111*(10), E962–E971.

Harter, M. R., Aine, C, & Schroeder, C. (1982). Hemispheric differences in the neural processing of stimulus location and type: effects of selective attention on visual evoked potentials. *Neuropsychologia*, *20*(4), 421–438.

Heeger, D. J., Boynton, G. M., Demb, J. B., Seidemann, E, & Newsome, W. T. (1999). Motion opponency in visual cortex. *J. Neurosci. 19*(16), 7162–7174.

Heeger, D. J., Huk, A. C., Geisler, W. S., & Albrecht, D. G. (2000). Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nat. Neurosci. 3*(7), 631–633.

Helmholtz, H. (1924). *Treatise on physiological optics* (J. P. C. Southall, Ed.). The Optical Society of America.

Hillyard, S. A. & Münte, T. F. (1984). Selective attention to color and location: an analysis with event-related brain potentials. *Percept. Psychophys. 36*(2), 185–198.

Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. 195*(1), 215–243.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. 160*, 106–154.

Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. 148*, 574–591.

Huettel, S. A. & McCarthy, G. (2000). Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. *Neuroimage*, *11*(5 Pt 1), 547–553.

Huk, A. C. & Heeger, D. J. (2000). Task-related modulation of visual cortex. *J. Neurophysiol. 83*(6), 3525–3536.

Huk, A. C. & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci. 25*(45), 10420–10436.

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

Ishihara, S. (1987). *Test for colour-blindness*. Kanehara Tokyo, Japan.

James, W. (1981). *The principles of psychology* (F. H. Burkhardt, Ed.). Harvard University Press.

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. *J. Neurosci. 31*(22), 8210–8219.

Kahneman, D. (1973). Attention and effort.

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, De Weerd, P, Desimone, R, & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*(5386), 108–111.

Katz, L. N., Yates, J. L., Pillow, J. W., & Huk, A. C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature*, *535*(7611), 285–288.

Kay, K. N. & Yeatman, J. D. (2017). Bottom-up and top-down computations in word- and face-selective cortex. *Elife*, *6*.

Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. *Curr. Biol. 25*(5), 595–600.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355.

Klein, B. P., Harvey, B. M., & Dumoulin, S. O. (2014). Attraction of position preference by spatial attention throughout human visual cortex. *Neuron*, *84*(1), 227–237.

Klein, B. P., Paffen, C. L. E., Pas, S. F. T., & Dumoulin, S. O. (2016). Predicting bias in perceived position using attention field models. *J. Vis. 16*(7), 15.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *J. Neurosci. 33*(41), 16275–16284.

Kuelpe, O. (1902). *The problem of attention*. books.google.com.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol. 16*(1), 37–68.

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn. Sci. 9*(2), 75–82.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen. 133*(3), 339–354.

Lennie, P. (2003). The cost of cortical computation. *Curr. Biol. 13*(6), 493–497.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U. S. A. 99*(14), 9596–9601.

Li, X., Lu, Z.-L., Tjan, B. S., Dosher, B. A., & Chu, W. (2008). Blood oxygenation level-dependent contrast response functions identify mechanisms of covert attention in early visual areas. *Proc. Natl. Acad. Sci. U. S. A. 105*(16), 6202–6207.

Ling, S. & Carrasco, M. (2006). When sustained attention impairs perception. *Nat. Neurosci. 9*(10), 1243–1245.

Ling, S., Pratte, M. S., & Tong, F. (2015). Attention alters orientation processing in the human lateral geniculate nucleus. *Nat. Neurosci. 18*(4), 496–498.

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Res. 49*(10), 1194–1204.

Liu, J. & Newsome, W. T. (2002). Functional organization of speed tuned neurons in visual area MT. *J. Neurophysiol. 89*(1), 246–256.

Liu, T., Stevens, S. T., & Carrasco, M. (2007a). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Res. 47*(1), 108–113.

Liu, T., Larsson, J., & Carrasco, M. (2007b). Feature-based attention modulates orientation-selective responses in human visual cortex. *Neuron*, *55*(2), 313–323.

Liu, T. T. & Frank, L. R. (2004). Efficiency, power, and entropy in event-related FMRI with multiple trial types. part i: theory. *Neuroimage*, *21*(1), 387–400.

Logothetis, N. K., Pauls, J, Augath, M, Trinath, T, & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157.

Luck, S. J., Chelazzi, L, Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas v1, v2, and V4 of macaque visual cortex. *J. Neurophysiol. 77*(1), 24–42.

Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*(5371), 1930–1934.

Mack, A. & Rock, I. (1998). *Inattentional blindness*. MIT press Cambridge, MA.

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

Marr, D. & Vision, A. (1982). A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, *1*(2).

McAdams, C. J. & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci. 19*(1), 431–441.

McBride, E. G., Lee, S.-Y. J., & Callaway, E. M. (2019). Local and global influences of visual spatial selection and locomotion in mouse primary visual cortex. *Curr. Biol.*

McLachlan, G. & Peel, D. (2000). *Finite mixture models (wiley series in probability and statistics)*. Wiley-Interscience.

Melcher, D. & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nat. Neurosci. 6*(8), 877–881.

Mesgarani, N. & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233–236.

Miconi, T. & VanRullen, R. (2016). A feedback model of attention explains the diverse effects of attention on neural firing rates and receptive field structure. *PLoS Comput. Biol. 12*(2), e1004770.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci. 6*, 414–417.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron*, *63*(6), 879–888.

Moore, T & Fallah, M. (2001). Control of eye movements and spatial attention. *Proc. Natl. Acad. Sci. U. S. A. 98*(3), 1273–1276.

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

Moore, T., Armstrong, K. M., & Fallah, M. (2003). Visuomotor origins of covert spatial attention. *Neuron*, *40*(4), 671–683.

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

Moray, N. (1959). Attention in dichotic listening: affective cues and the influence of instructions. *Q. J. Exp. Psychol. 11*(1), 56–60.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas v1, v2, and V4 in the presence of competing stimuli. *J. Neurophysiol. 70*(3), 909–919.

Murray, S. O. (2008). The effects of spatial attention in early human visual cortex are stimulus independent. *J. Vis. 8*(10), 2.1–11.

Nachmias, J & Sansbury, R. V. (1974). Letter: grating contrast: discrimination may be better than detection. *Vision Res. 14*(10), 1039–1042.

Naka, K. I. & Rushton, W. A. H. (1966). S-potentials from colour units in the retina of fish (cyprinidae). *J. Physiol. 185*(3), 536–555.

Nakajima, M., Schmitt, L. I., & Halassa, M. M. (2019). Prefrontal cortex regulates sensory filtering through a basal Ganglia-to-Thalamus pathway. *Neuron*.

Neisser, U. (1979). The control of information pickup in selective looking. *Perception and its development: A tribute to Eleanor J. Gibson*, 201–219.

Neri, P. (2010). How inherently noisy is human sensory processing? *Psychon. Bull. Rev. 17*(6), 802–808.

Neri, P. (2018). The empirical characteristics of human pattern vision defy theoretically-driven expectations. *PLoS Comput. Biol. 14*(12), e1006585.

Nestares, O. & Heeger, D. J. (2000). Robust multiresolution alignment of MRI brain volumes. *Magn. Reson. Med. 43*(5), 705–715.

Newsome, W. T. & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci. 8*(6), 2201–2211.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52–54.

Nothdurft, H. C. (1993). The role of features in preattentive vision: comparison of orientation, motion and color cues. *Vision Res. 33*(14), 1937–1958.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci. 5*(11), 1203–1209.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT–MST. *Neuron*, *18*(4), 591–598.

Ogawa, S, Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U. S. A. 87*(24), 9868–9872.

Ohzawa, I, Sclar, G, & Freeman, R. D. (1982). Contrast gain control in the cat visual cortex. *Nature*, *298*(5871), 266–268.

Ohzawa, I, Sclar, G, & Freeman, R. D. (1985). Contrast gain control in the cat's visual system. *J. Neurophysiol. 54*(3), 651–667.

Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc. Natl. Acad. Sci. U. S. A. 112*(4), E351–60.

Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Res. 44*(7), 669–683.

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97.

Pestilli, F., Ling, S., & Carrasco, M. (2009). A population-coding model of attention's influence on contrast response: estimating neural effects from psychophysical data. *Vision Res. 49*(10), 1144–1153.

Pestilli, F., Carrasco, M., Heeger, D. J., & Gardner, J. L. (2011). Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. *Neuron*, *72*(5), 832–846.

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., & Churchland, A. K. (2019). *Lapses in perceptual judgments reflect exploration.*

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol. 32*(1), 3–25.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *J. Exp. Psychol. 109*(2), 160–174.

Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The neural representation of speed in macaque area MT/V5. *J. Neurosci. 23*(13), 5650–5661.

Priebe, N. J., Lisberger, S. G., & Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *J. Neurosci. 26*(11), 2941–2950.

Prins, N. (2012). The psychometric function: the lapse rate revisited. *J. Vis. 12*(6).

Rabinowitz, N. C., Goris, R. L., Cohen, M., & Simoncelli, E. P. (2015). Attention stabilizes the shared gain of V4 populations. *Elife*, *4*, e08998.

Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Res. 46*(15), 2336–2343.

Rees, G, Friston, K, & Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nat. Neurosci. 3*(7), 716–723.

Rees, G, Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, *278*(5343), 1616–1619.

Ress, D & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nat. Neurosci. 6*(4), 414–420.

Ress, D, Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nat. Neurosci. 3*(9), 940–945.

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

Reynolds, J. H. & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185.

Roitman, J. D. & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci. 22*(21), 9475–9489.

Rolls, E. T. & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp. Brain Res. 65*(1), 38–48.

Ruff, D. A. & Cohen, M. R. (2017). A normalization model suggests that attention changes the weighting of inputs between visual areas. *Proc. Natl. Acad. Sci. U. S. A. 114*(20), E4085–E4094.

Ruff, D. A. & Cohen, M. R. (2016). Attention increases spike count correlations between visual cortical areas. *J. Neurosci. 36*(28), 7523–7534.

Ruff, D. A. & Cohen, M. R. (2018). *Simultaneous multi-area recordings suggest a novel hypothesis about how attention improves performance.*

Ruff, D. A., Ni, A. M., & Cohen, M. R. (2018). Cognition as a window into neuronal population space. *Annu. Rev. Neurosci. 41*, 77–97.

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

Sàenz, M., Buraĉas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Res. 43*(6), 629–637.

Sapir, A., d'Avossa, G., McAvoy, M., Shulman, G. L., & Corbetta, M. (2005). Brain signals for spatial attention predict performance in a motion discrimination task. *Proc. Natl. Acad. Sci. U. S. A. 102*(49), 17810–17815.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2018). *Psychophysical scaling reveals a unified theory of visual memory strength.*

Sclar, G, Maunsell, J. H., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res. 30*(1), 1–10.

Sclar, G, Lennie, P, & DePriest, D. D. (1989). Contrast adaptation in striate cortex of macaque. *Vision Res. 29*(7), 747–755.

Sclar, G, Ohzawa, I, & Freeman, R. D. (1985). Contrast gain control in the kitten's visual system. *J. Neurophysiol. 54*(3), 668–675.

Seidemann, E, Poirson, A. B., Wandell, B. A., & Newsome, W. T. (1999). Color signals in area MT of the macaque monkey. *Neuron*, *24*(4), 911–917.

Serences, J. T. & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301–312.

Shadlen, M. N. & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol. 86*(4), 1916–1936.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci. 16*(4), 1486–1510.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *J. Neurophysiol. 97*(1), 229–237.

Simoncelli, E. P. & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Res. 38*(5), 743–761.

Simons, D. J. & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Singh, K. D., Smith, A. T., & Greenlee, M. W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage*, *12*(5), 550–564.

Sirotin, Y. B. & Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature*, *457*(7228), 475–479.

Smith, A. T., Wall, M. B., Williams, A. L., & Singh, K. D. (2006). Sensitivity to optic flow in human cortical areas MT and MST. *Eur. J. Neurosci. 23*(2), 561–569.

Snyder, A. C., Yu, B. M., & Smith, M. A. (2018). Distinct population codes for attention in the absence and presence of visual stimulation. *Nat. Commun. 9*(1), 4382.

Snyder, C. R. (1972). Selection, inspection, and naming in visual search. *J. Exp. Psychol. 92*(3), 428–431.

Softky, W. R. & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci. 13*(1), 334–350.

Soto, D. & Blanco, M. J. (2004). Spatial attention and object-based attention: a comparison within a single task. *Vision Res. 44*(1), 69–81.

Sperling, G & Melchner, M. J. (1978). The attention operating characteristic: examples from visual search. *Science*, *202*(4365), 315–318.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29.

Spitzer, H, Desimone, R, & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*(4850), 338–340.

Stigliani, A., Jeska, B., & Grill-Spector, K. (2017). Encoding model of temporal processing in human visual cortex. *Proc. Natl. Acad. Sci. U. S. A. 114*(51), E11047–E11056.

Stocker, A. A. & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci. 9*(4), 578–585.

Sun, P., Ueno, K., Waggoner, R. A., Gardner, J. L., Tanaka, K., & Cheng, K. (2007). A temporal frequency–dependent functional architecture in human V1 revealed by high-resolution fMRI. *Nat. Neurosci. 10*, 1404.

Taylor, M. & Creelman, C. D. (1967). PEST: efficient estimates on probability functions. *J. Acoust. Soc. Am. 41*(4A), 782–787.

Teller, D. Y. (1984). Linking propositions. *Vision Res. 24*(10), 1233–1246.

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention*. Macmillan.

Tjur, T. (2009). Coefficients of determination in logistic regression Models—A new proposal: the coefficient of discrimination. *Am. Stat. 63*(4), 366–372.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res. 23*(8), 775–785.

Tolias, A. S., Moore, T, Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, *29*(3), 757–767.

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R, Born, R. T., Brady, T. J., ... Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci. 15*(4), 3215–3230.

Tootell, R. B., Hadjikhani, N. K., Vanduffel, W, Liu, A. K., Mendola, J. D., Sereno, M. I., & Dale, A. M. (1998a). Functional analysis of primary visual cortex (v1) in humans. *Proc. Natl. Acad. Sci. U. S. A. 95*(3), 811–817.

Tootell, R. B. H., Hadjikhani, N. K., Mendola, J. D., Marrett, S, & Dale, A. M. (1998b). From retinotopy to recognition. *Trends Cogn. Sci. 2*(5), 174–183.

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

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156–177.

Treisman, A. M. (1960). Contextual cues in selective listening. *Q. J. Exp. Psychol. 12*(4), 242–248.

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

Treue, S & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, *382*(6591), 539–541.

Tsal, Y & Lavie, N. (1988). Attending to color and shape: the special role of location in selective visual processing. *Percept. Psychophys. 44*(1), 15–21.

Verhoef, B.-E. & Maunsell, J. H. R. (2017). Attention-related changes in correlated neuronal activity arise from normalization mechanisms. *Nat. Neurosci. 20*(7), 969–977.

Vintch, B. & Gardner, J. L. (2014). Cortical correlates of human motion perception biases. *J. Neurosci. 34*(7), 2592–2604.

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

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

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

Wang, L. & Krauzlis, R. J. (2018). Visual selective attention in mice. *Curr. Biol. 28*(5), 676–685.e4.

Watson, J. D., Myers, R, Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., ... Zeki, S. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb. Cortex*, *3*(2), 79–94.

Wichmann, F. A. & Hill, N. J. (2001). The psychometric function: i. fitting, sampling, and goodness of fit. *Perception and Psychophysics*, *63*(8), 1293–1313.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychon. Bull. Rev. 1*(2), 202–238.

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., & Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat. Neurosci. 9*(9), 1156–1160.

Womelsdorf, T., Anton-Erxleben, K., & Treue, S. (2008). Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. *J. Neurosci. 28*(36), 8934–8944.

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat. Neurosci. 8*(11), 1595–1602.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A. 111*(23), 8619–8624.

Zarahn, E, Aguirre, G, & D'Esposito, M. (1997). A trial-based experimental design for fMRI. *Neuroimage*, *6*(2), 122–138.

Zeki, S, Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C, & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci. 11*(3), 641–649.

Zenger-Landolt, B. & Heeger, D. J. (2003). Response suppression in v1 agrees with psychophysics of surround masking. *J. Neurosci. 23*(17), 6884–6893.

Zohary, E, Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*(6485), 140–143.