UNIVERSITY OF CALIFORNIA, SAN DIEGO

Dynamic Neural Coding: An Information-theoretic Account of Attention

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

in

Psychology

by

Sameer Saproo

Committee in charge:

Professor John T. Serences, Chair
Professor Virginia R. de Sa
Professor David E. Huber
Professor Donald I. A. Macleod
Professor John H. Reynolds
Professor Tatyana O. Sharpee

2012

The Dissertation of Sameer Saproo is approved, and it is acceptable in quality and form or publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

EPIGRAPH

Throughout the centuries there were men who took first steps, down new roads, armed with nothing but their own vision.

Ayn Rand

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Every dissertation has a background story. This one has too, and it would be incomplete without a brief exposition of that story.

For the longest time, I had wanted to discover new ideas and invent new techniques, to try to give back to the world in some measure of what I had received from it. Initially, I thought that the way to go would be to become a tech entrepreneur. But, after joining the computer science department at UC Irvine as graduate student, I decided that I wanted to stay in academia. I worked briefly with some professors in the CS department before I met John. It was pure accident that my random web browsing in the science library led me to John's webpage. I saw that he did research in multivariate classifiers and Bayesian models for brain data, something that I had recently learnt in a machine learning course. Impulsive as I am, I sent him an email. We met. After a couple of weeks, he asked me if I wanted to do a PhD with him. Now, this was an unexpected and unconventional offer; he was in the cognitive science department, and even though his research involved computational elements, I had had no training in the brain sciences. In fact, I had not studied anything remotely biological since early high school. My friends and family were bemused by the idea that I would even try to get a PhD in psychology. John asked me to think carefully before I made the leap, and everyone else told me not to do it. What did I think? Well, I don't recollect thinking hard about the consequences of my decision. It seemed like a big adventure, and the intense curiosity to see how it would turn out superseded any caution that I could have exercised. Besides, I told myself that if John thinks that I can do it then I probably can. Here was a guy who was probably as crazy as I am, and I wanted to work with him. I took the plunge. So, how did it turn out?

It was every bit an adventure as I had hoped, and like every adventure there was joy and sorrow, victories and defeats. It was a joy to be exposed to psychology, neuroscience, cognitive science, and philosophy among other disciplines that I had known little about. I felt as if I had been only partly educated during my engineering/CS years, and now I was receiving the other half of my education. I got to do original research and got to publish scientific papers. The moments of pride and joy were intense; I vividly remember the times I solved a particularly gnarly technical problem, or when I had a deep insight into the mechanics of brain function, or when I had an inspiration for an experiment or a theoretical investigation. But all of this was facilitated by John. He provided the runway for my flights of imagination to take off, channeled them in the right direction, and consoled me when they crashed. I also found out that most interesting research is somewhat accidental, including mine. A technical problem posed by John to the entire lab led me to work with information theory. Then, John introduced me to the world of population codes (by assigning a long reading list on my first vacation). It led to chapter 2 of this dissertation. That in turn led to my qualifying paper, where I tried to theoretically assimilate the field of information theory, population codes and attention. Novel insights while writing the qualifying paper led to chapters 1 and 3 of this dissertation. During the winter-break of 2009, when I was working feverishly on what seemed like a chore (yes, I hated the qualifying paper), little did I know that I was laying the foundations of this dissertation and research work that would one day win me acclaim. Looking back, that was probably the best chore I ever did. During my PhD, I also learned to love. I found out that there probably isn't an emotion more powerful than love, and it is what makes life worth living. It is good for science too. This dissertation is a testament to that.

The unpleasant parts of the adventure were primarily due to my teething problems with learning this brand new field of neuroscience. For a mind shaped by the perfection and concreteness of engineering and computer science, the messiness of the brain was bewildering

and overwhelming at times. The jargon, the verbosity of papers, the avalanche of experimental details, and the nuances of collected data, nauseated me. I had jumped into the deep end of a piranha infested pool, and now had to learn how to swim while being bitten. In fact, I now believe that the personality required to be a good computer scientist or an engineer doesn't overlap much with the personality expected from a good experimental neuroscientist. Thankfully, I was neither a good computer scientist nor a good neuroscientist. I am still learning how to think like an experimentalist. The break-neck pace and indefatigable nature of John's working style posed another challenge. I felt that I had to run just to be able to keep up with what looked like a leisurely walk for John. Sometimes I had to remind myself that my advisor was a human and not an android (although I have heard rumors to the contrary). To John's credit, although he pushed me to excel, he never pushed me beyond the breaking point. Had he been a less patient and understanding man, I probably would have not been writing this preface. We bickered, argued, and sometimes outright fought over ideas and decisions, but I never felt that he was beating me down using his position of authority. He made me feel like an equal in his presence. Although, both John and I know that I could have achieved much more during my time with him, I hope he realizes, as I have, that given some of the circumstances whatever I *did* achieve – including this dissertation – is a monumental accomplishment and an ode to his worth as an advisor and a human being. I would also like to confess that there were more than a few times during the past 4 years when I was on the verge of quitting. It was only due to the unwavering support of my family, friends, and most of all John, that I did not. They made me realize that quitting is not the answer to tough problems, and that persisting in the face of overwhelming odds is the *only* path to glory. I am immeasurably indebted to all of them.

They say that one does not finish one's dissertation, one abandons it. Therefore, I abandon mine. Although it is an infinitesimally small contribution to human knowledge, it is still a contribution. I have succeeded in the aim that I had set out to accomplish 4 years ago. I don't

know what the future holds for me, whether I will fail or succeed at life, or how I will cope with upcoming challenges, but what I know is that I am better prepared than I was 4 years ago and I look forward to what's ahead. The end of one adventure can only be the beginning of another.

ACKNOWLEDGEMENTS

I would like to acknowledge Professor John Serences for his invaluable support, patience and guidance, without which this dissertation would not have been possible. Thank you for believing in me.

Chapter 1, in part, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material, while Reynolds J., Sharpee T., and Serences J. were co-authors.

Chapter 2, in full, is a reprint of the material as it appears in "Spatial attention improves the quality of population codes in human visual cortex", Saproo S., Serences J., Journal of Neurophysiology, 2010 Aug;104(2):885-95. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material, while Serences J. was co-author.

Finally, I would like to acknowledge my family, friends and well-wishers, who were emotionally a part of my eventful and enlightening journey to this point.

This work was completed with support given to Serences J. by National Institute of Mental Health grants R21-083902 and R01-092345 as well as UCSD interdisciplinary research fellowship given to the dissertation author.

CURRICULUM VITAE

Education

University of California, San Diego — 2008 – 2012
Ph.D. in Psychology

University of California, Irvine — 2006-2007
M.S. in Computer Science

University of Mumbai, India — 1999-2003
Bachelors in Engineering

Cold Spring Harbor Laboratory, NY — 2010
Summer Camp for Computational Neuroscience of Vision

Institute for Neural Computation, UCSD — 2009
Summer Camp for Temporal Dynamics of Learning

Awards

Dean's TA Fellowship, UC Irvine — 2007
Distinguished Student Award, UC Irvine — 2007
Anderson Travel Award, UCSD — 2008, 2012
CSHL Summer Course Fellowship, NY — 2010
COSYNE Travel Award, Salt Lake City — 2011
Chancellor's Interdisciplinary Fellowship, UCSD — 2011
VSS Travel Award, Naples — 2012

Professional Experience

Tutorz.com, Port Hueneme, CA — 2007
Co-Founder
Winner of hiTec Octane Business Plan Competition, UC Irvine

Design Science, Long Beach, CA — 2007
Summer Research Intern

SAP Labs, Bangalore, India — 2004-2006
Software Development Specialist

Caritor Inc., Bangalore, India — 2003-2004
Software Engineer

Journal Publications

Serences, **Saproo**, Scolari, Ho & Muftuler (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions, *NeuroImage*.
Best paper award in System Neuroscience section.

Serences & **Saproo** (2010). Population response profiles in early visual cortex are biased in favor of more valuable stimuli, *Journal of Neurophysiology*.

Saproo & Serences (2010). Spatial attention increases the information content of population codes in human visual cortex, *Journal of Neurophysiology*.
Recommended by Faculty of 1000 Biology.

Serences and **Saproo** (2011). Computational advances in linking BOLD and behavior, *Neuropsychologi*a.

Saproo, Reynolds, Sharpee and Serences. Dynamic Neural Coding: A functional model of visual attention. (in prep)

Saproo and Serences. Attention improves information transmission between V1 and MT. (in prep)

Presentations

Saproo & Serences (2008). Spatial attention increases the information content of population codes in human visual cortex. Poster at Society of Neuroscience Annual Meeting, Washington D.C.

Saproo & Serences (2009). Top-down allocation of spatial attention to competing objects of equal behavioral relevance but unequal physical salience. Poster at Society of Neuroscience Annual Meeting, Chicago.

Peterson, Mullane, Saproo, Tran, Yazdani, Lee, Sejnowski, Poizner (2010). Tough decisions, rough moves: the kinematics of reaching for choices during rewarded learning. Poster at Society of Neuroscience Annual Meeting, San Diego.

Saproo & Serences (2011). Attention improves information transmission between V1 and MT. **Talk** at Computational and Systems Neuroscience (Cosyne) Annual Meeting, Salt Lake City.

Saproo & Serences (2011). Attention improves information transmission between V1 and MT. **Talk** at Cognitive Neural Systems Colloquia, UCSD.

Saproo & Serences (2011). Attention improves information transmission between V1 and MT. **Talk** at Brain Corporation, San Diego.

Saproo & Serences (2012). Attention improves communication between V1 and MT. **Talk** at Vision Science Society Annual Meeting, Naples, FL.

ABSTRACT OF THE DISSERTATION

Dynamic Neural Coding: An Information-theoretic Account of Attention

By

Sameer Saproo

Doctor of Philosophy in Psychology

University of California, San Diego, 2012

Professor John T. Serences, Chair

Visual attention facilitates faster and more accurate decision-making for behaviorally relevant stimuli. To understand the computational mechanism underlying attention, I investigated the impact of attention on population codes in early visual cortex. First, I lay the theoretical foundation of this investigation, whereby the known neuromodulatory activity of attention is examined within the context of information theory to propose a framework termed 'Dynamic Neural Coding'. The framework suggests that attention dynamically alters neural codes used to represent basic stimulus features so that behaviorally relevant stimuli are represented and communicated by population responses with higher fidelity than irrelevant stimuli. The framework also

suggests that attention and adaptation might jointly mediate neural codes to achieve metabolically efficient sensory information processing. The framework is supported by two experimental studies that use a combination of visual psychophysics, functional magnetic resonance imaging (fMRI), computational modeling, and information-theoretic data analysis, to show how attention modulates population codes to impact information processing. The first experiment reveals that attention increases the quality of sensory representations – measured by an increase in mutual information between population response and stimuli – as early as primary visual cortex. This increase in sensory representation is largely driven by multiplicative scaling of tuning function that forms the population code. According to dynamic neural coding framework, attention-induced improvement in population codes in upstream areas should also improve the transmission of encoded information to downstream areas. The second experiment tested this hypothesis, and found that attention improved the efficacy of communication between two cortical areas – V1 and MT – that interact with during motion processing. Furthermore, a simulation of the inter-cortical interaction between V1 and MT using a computational model reveals that attention-modulations in V1 dictate the fidelity of representations in MT as well as the synchrony between the two areas.

GENERAL INTRODUCTION

Living organisms react to a dynamically changing environment by continuously gathering information from different sensory organs and then combining this information to produce a unified perceptual representation of the environment. Sensory information is represented in the cortex by patterns of neural activity in domain specific areas, and these representations require a protocol or a *code* for correct interpretation and transmission within and between brain regions. Current theories hold that populations of inherently noisy neurons encode information about basic stimulus features in early visual cortex using sophisticated *population codes* (Rieke et al., 1999; Pouget et al., 2003). These accounts implicitly assume that coding protocols remain static for a given perceptual domain; the quality of encoded information can only change with changes in the quality of incoming signals. However, neural signals are relayed over an essentially noisy channel during sensory processing due to stochastic synaptic activity, which can degrade the accuracy of neural computations (Manwani and Koch, 2001). Some codes are going to be more suitable for communication over a particular noisy channel over other codes, necessitating their tuning in order to optimize computations. Furthermore, the 'data processing inequality' – which states that data processing can only degrade or at best sustain the information contained in the raw sensory data (Cover and Thomas, 1991) – dictates that given fixed processing capacity, information about behaviorally-relevant inputs must be preferentially preserved during sensory processing, while information about behaviorally irrelevant inputs can be allowed to degrade as a consequence. Cognitive factors such as attention are known to prioritize information processing of behaviorally relevant sensory objects so that decisions regarding them are made more quickly and with better accuracy. Some models suggest that neurobiological modulations associated with attention can influence information processing by normalizing stimulus-driven neural response (Reynolds and Heeger, 2009; Carandini and Heeger, 2011). However, the key functional role that attention plays in perception requires more

theoretical exploration. For instance, merely saying that attention leads to a 20-40% increase in neural gain does not tell us *how* attention affects the information contained in population responses or *how* attention facilitates the transmission of this information from one brain area to another, both of which are vital to understanding how attention affects neural computations. Although, a few attempts have been made to understand attention in the context of population codes and information processing (e.g. Pouget et al., 2001), this largely remains an unexplored territory.

Here I develop a novel account of how attentional modulations in neural gain and (co)variance gives rise to functionally meaningful changes in the quality of population codes based on the tenants of information theory. First, the quality of any code depends on the fidelity with which the building blocks of the code, or *codewords*, represent inputs (MacKay, 2004). For example: Gaussian tuning functions are the codewords used in V1 to communicate information about stimulus orientation. Second, information theory provides a principled way of evaluating changes in the representational capacity of any code, including population codes, using statistical properties of the codewords (Doya et al., 2007). Thirdly, different distributions of codewords are optimal for different stimulus statistics and noise statistics for efficient neural communication of sensory information (Latham and Roudi, 2009). These principles can be used to provide an information-theoretic interpretation of existing empirical data that describe the impact of attention, such as gain modulation of feature-tuning functions (McAdams and Maunsell, 1999; Martinez-Trujillo and Treue, 2004), noise reduction (Mitchell et al., 2007) and noise de-correlation in population responses (Cohen and Maunsell, 2009; Mitchell et al., 2009). I hypothesize that attention can be thought as operating to dynamically tune population codes such that population responses carry relatively more information about behaviorally relevant stimuli as compared to behaviorally irrelevant stimuli, which leads to more accurate transfer of information regarding attended stimuli within the cortical processing hierarchy, ultimately facilitating more

accurate computations associated with attended objects. This hypothesis led to the development of a broad theoretical framework – termed 'Dynamic Neural Coding' - that attempts to formally link attention modulations directly with changes in the information content of perceptual representations. The framework is outlined in Chapter 1. Dynamic neural coding framework can also be used to explain interactions between attention and adaptation towards facilitating information transmission between neural populations (Sharpee et al., 2006; Pestilli et al., 2007; Gutnisky and Dragoi, 2008) as well as how the dynamic capacity of codes might change to optimize perception given fundamental neurobiological constraints such as metabolic consumption and neural refractory periods (Levy and Baxter, 1996; Balasubramanian et al., 2001; Laughlin and Sejnowski, 2003). Although, I provide a brief exposition of the rather complex interaction between attention and adaptation in Chapter 1, future theoretical and empirical investigation to required to fully understand it.

There are two important experimental investigations that are used to verify the framework. The first experiment shows that spatial attention increases quality of population codes encoding behaviorally relevant features – as measured by an increase in the mutual information between population responses and stimuli - through multiplicative scaling of voxel tuning functions (VTFs) in visual areas as early as primary visual cortex. The experimental design and results are presented in Chapter 2. In the course of this investigation, we realized that BOLD signal from all voxels in a region of interest (ROI) could not be compared on an equal footing. Some voxels – due to their more homogenous composition – were more reflective of underlying neural activity than others, and would therefore see more pronounced and accurate attention modulations in their VTFs. One manner of sifting such 'pure' voxels is to find voxels that carry more information about the stimulus and have sharper tuning, a methodological innovation presented in Chapter 2, along with some rudimentary concepts behind linking voxel tuning functions (VTFs) and neural tuning functions (NTFs).

Within the context of the theoretical framework, showing that neural codes increase their encoding capacity due to attention – such as found in the first experiment – has several implications, including one for inter-cortical communication. For instance, dynamically changing neural codes that represent basic features in upstream areas would theoretically make the transmission of this information to downstream areas more robust to intrinsic noise, thereby ensuring that relevant objects are represented with higher fidelity at each point along the sensory processing chain. The second experiment - detailed in chapter 3 - tested this prediction in the context of motion computation since it is well established that MT direction-selective cells derive their complex direction selectivity by pooling signals from V1 direction selective cells, thus forming a communication network. Since attention has been shown to improve encoding characteristics of population codes in V1 (as shown in Chapter 2), the transmission of signals from V1 and the resultant pooling by MT neurons should also improve as a result. Data from the experiment indeed show that attention to direction of motion increases the fidelity of communication between direction-selective neurons in V1 and MT. On methodological front, this experiment saw the first use of forward-modeling approach for decomposing voxel activation in response to motion (from a random dot kinetogram) into estimated response for underlying direction-selective neural populations. Forward-modeling approach has been previously used to model BOLD activation in response to color (Brouwer and Heeger, 2009) and orientation (Brouwer and Heeger, 2011).

The experiment, however, threw up a paradoxical observation; we found that the modulation of cross-area correlation between neural populations was inversely proportional to the modulation of their mean response i.e. the neural populations that showed an increase in mean response in the two areas actually had a decrease in cross-area noise correlation. I simulated inter-cortical interactions within a known model of MT computation (Rust et al., 2006), incorporating recent experimental observations regarding the relationship between neural gain and neural

correlations (Cohen and Maunsell, 2010), to show that the paradoxical results are derived from the inheritance of attentional modulations in V1 populations by MT population during motion computation. This simulation and modeling work, and the results thereof, is also presented in Chapter 3.

Overall, the goal of this research work has been to move beyond the established concepts, such as how attention modulations can increase neural signal-to-noise (SNR), and take tentative next step towards developing a rigorous and formal treatment that accounts for *how* these neural changes due to attention might produce population codes capable of more efficiently processing sensory input, enhancing the quality of perception, and ultimately guiding more efficient behavioral interactions with the environment.

CHAPTER 1

EFFICIENT SENSORY INFORMATION PROCESSING THROUGH DYNAMIC NEURAL CODING

Selective attention modulates the gain and the variability of sensory neurons in response to changing behavioral goals. These operations have a functional role: changing neural gain fundamentally alters the codes that are used to represent sensory information, which along with changes in variability regulates the transfer of sensory information from node to node in the cortical hierarchy. This preferential modulation of neural communication could underlie the broad based selective information processing that leads to performance improvements with attention. Furthermore, sensory adaptation might work in conjunction with attention to facilitate efficient information processing by dynamically adapting neural codes to changing stimulus statistics.

1.1 INTRODUCTION

Perception relies on a continuous filtering process in which neural signals representing sensory stimuli undergo a series of transformations (Marcelja 1980; Olshausen and Field 1996a; b; van Hateren and van der Schaaf 1998). At each stage of filtering, the total information available in the neural signals is reduced due to stochastic synaptic activity. Therefore, perceptual acuity ultimately depends on the extent to which task-relevant information is preserved during processing. Selective attention refers to the collection of neural mechanisms that modulate neural activity to decrease the loss of information related to a behaviorally relevant stimuli, thereby ensuring that decisions regarding attended features, locations and objects can be made more rapidly and accurately.

Deploying spatial attention to a particular location in the visual field increases the firing rate of neurons irrespective of their feature-tuning (Connor et al. 1997; Reynolds et al. 2000), reduces the trial-by-trial variability in the firing rate of neurons (Bressler and Silver 2010;

Mitchell et al. 2007), and changes the co-variability across neurons that are responding simultaneously to the same sensory input (Cohen and Maunsell 2009; 2011a; Mitchell et al. 2009). On the other hand, attending to a target-defining feature increases the response of neurons tuned to the attended feature while suppressing the response of neurons tuned to other features (Martinez-Trujillo and Treue 2004); these modulations occur even beyond the current focus of spatial attention (Cohen and Maunsell 2011a; Liu and Mance 2011; Martinez-Trujillo and Treue 2004; Saenz et al. 2002; Serences and Boynton 2007; Treue and Maunsell 1999; White and Carrasco 2011). As in the case of spatial-attention, feature-based attention also leads to a reduction in trial-by-trial variability (Cohen and Maunsell 2011b) and a changes in the correlation structure of neural responses (Cohen and Maunsell 2011a). Therefore, selective attention operates via the action of three distinct mechanisms; enhancing neural gain, reducing neural variability, and changing the correlational relationship between neurons in a population (see Figure 1.1 for a graphic representation). These modulations are commonly thought to improve the signal to noise ratio of population-level neural responses (Reynolds and Heeger, 2009), yet we lack a complete understanding of the algorithmic mechanism through which attention impacts neural computations.

Here, we focus on understanding how attentional modulations dynamically alter firing statistics so that responses evoked by relevant stimuli are more efficiently represented and communicated within sensory systems. First, we use an information theoretic framework to understand how attention-related changes in the dynamic range of neural responses can influence the encoding capacity of population codes. Next, we discuss how attention-related changes in the magnitude and in the correlational structure of neural variability supplement the impact of gain changes in ensuring that signals related to relevant stimuli are more effectively relayed from one processing node to another. Finally, we discuss how neural modulations associated with selective

attention and bottom-up adaptation can jointly mediate gain and variability to achieve an optimal balance between the quality of stimulus encoding and the expenditure of energy.

1.2 SENSORY CODING AND NEURAL COMMUNICATION

Neural codes are widely accepted as the building blocks of information representation and computation in the brain (see Appendix A). Although neural codes – especially population codes – have been well-described in the context of representing a specific stimulus feature, they are less appreciated in the role that they are intended for; neural communication. This disconnect is ironic since the two aspects of neural codes - representation and communication - are inextricably linked. If neurons communicate using spiking activity then neural codes are how neural responses can be interpreted, either at the source of spikes by the experimenter or at intended destination of these signals by receiving neurons. Therefore, any attention-induced modulation of neural codes is ultimately going to affect the communication of information between neural ensembles. Moreover, as with any code, there are two ways to make neural codes easier to decipher: the basic units of the code (codewords) can move farther apart so that there is less potential for confusability given noisy encoding, or the stimulus can be more reliability encoded. These two routes to improving a code amount to changes in the dynamic range and the (co)variability of a population code, respectively. We first focus on the role of gain changes since research has almost exclusively examined this domain until very recently. We then discuss how and why changes in neural variability interact with gain changes to synergistically increase in the quality of population codes.

1.2.1 DYNAMIC RANGE AND NEURAL CODES

To understand how changes in the dynamic range of neural responses can change the neural code, first consider a generic schema of communication over any noisy channel (Figure

1.2) (MacKay 2004) that consists of 3 steps: encoding, transmission and decoding. This schema is also applicable to neural communication given the unavoidable loss of information during spiking and synaptic transmission (Conti and Wanke 1975; Manwani and Koch 2001; Schneidman et al. 1998; White et al. 2000). Since a noisy channel invariably limits the speed and accuracy of communication (Cover and Thomas 1991; MacKay 2004; Rieke et al. 1999), sensory systems should mitigate the deleterious effects of channel noise by adapting the neural encoding scheme to channel's noise characteristics by a technique known as the *system solution* (MacKay 2004). The system solution involves making structural changes to the communication protocol to overcome noise, and although this solution can be implemented in various ways depending on reliability, speed, and energy constraints, it invariably involves using more bits per message for communication than is necessary to communicate over a noiseless channel.

One system solution would be to recruit additional sensory neurons in the communication network. For instance, an ensemble of similarly tuned neurons could encode and transmit the same information in parallel, which could then be averaged in real-time by receiving neuron(s) through dendritic summation before decoding. Any uncorrelated noise induced either at the source (while encoding) or during transit (through channel noise) could be effectively removed through averaging at the decoder. A recent study has suggested that sensory systems might employ efficient pooling of sensory neurons in addition to gain and noise reduction to improve behavioral performance (Pestilli et al, 2011; Serences, 2011). However, since each additional spike in the brain consumes ATP/O2, recruiting larger neural ensembles to improve communication efficacy may not be an optimal solution from a metabolic point of view, moreover such pooling does not mitigate any noise correlations.

An arguably more efficient method to improve the error tolerance of the neural code via gain changes is by increasing the average statistical distance between individual codewords (the

basic units of a code). This method would entail changing the input-output relationship of the encoder and hence change the nature of the code itself. For instance, if a visual feature α is encoded with mean firing rate of 10 Hz and feature β is encoded as 12Hz, then a misfiring of 1 Hz when either α or β was encoded would yield 11Hz as the message that is received by the decoder. Since this message is equidistant from both features, unambiguous decoding is not possible. However, if the neural code is changed such that β is now mapped to 15Hz, then a misfiring of 1Hz would not greatly influence the probability of correct decoding. Instead, it would take a misfiring in the rage of 3Hz to again cause decoding problems for the new code, but statistically speaking it is less probable than a misfiring of 1Hz given that neural noise generally follows a near-Poisson distribution (~7% vs ~11% chance). Furthermore, this method of dynamically changing the code to improve communication efficacy is more metabolically efficient solution to noisy neural communication compared to recruiting additional sensory neurons.

Variability in neural signals (or noise) can be introduced during synaptic activity, biophysical activity in the cell body, or during the transit of an action potential over the axon. Therefore, at the level of a decoder that receives a set of signals, it is impossible to differentiate between variability introduced by the encoder from variability introduced by a noisy channel. Moreover, from the decoder's perspective it is immaterial at what stage variability was introduced in the received message as long as the message is unambiguous. Thus, changing the code by increasing the distance between codewords (as in Figure 2) should generally improve decoding accuracy irrespective of how noise was introduced in the received message. Therefore for the sake of simplicity, we assume perfect encoding but assume the presence of a noisy channel that corrupts all neural signals.

1.2.2 SHANNON'S CHANNEL CODING THEOREM

Formally, Shannon's channel coding theorem shows how codes can be made more tolerant to channel noise in the pursuit of improving decoding accuracy (Latham and Roudi 2009; MacKay 2004). As we discussed above, increasing the dynamic range of neural responses effectively increases separability of adjacent codewords that are used to represent different exemplars of an input set. This increase in separability is achieved because increasing the bit-depth of each codeword increases the number of *unique* codewords that are available to encode a set of stimulus features. Thus, if each codeword is an *N* bit binary string, then there are 2^N possible codewords that can be used to represent an input set *M*. The ratio of the size of the feature set to be encoded (*M*) and the number of available codewords dictates the efficacy of communication across a noisy channel (i.e. the ratio $M/2^N, M \ll 2^N$). Higher values will correspond to more efficient but error prone codes, as the number of codewords will be close to the number of features to be encoded. In contrast, lower values will correspond to less efficient but more error-tolerant codes, as the number of codewords will far exceed the number of possible input features. An optimal value of this ratio given a fixed input set *M* can be achieved if the distance between codewords is large enough to ensure correct decoding, without making the bit-depth of each codeword so large as to hamper either communication rate or the energy overhead. Note that the entropy of responses across a population of neurons is equivalent to the average bit depth of neural codewords, and therefore sets the upper limit on the amount of information that can be encoded using a code, which we term encoding capacity. Since spike trains can be modeled as a binary symmetric channel (Brenner et al. 2000b; Nemenman et al. 2004; Strong et al. 1998), this theorem is also applicable for spiking networks in the brain.

Attention related changes in the dynamic range of neural responses will increase total entropy, and thus the encoding capacity, of a population code. This increase in encoding capacity

ultimately is used towards improving the error tolerance of the neural code. However, this increase in encoding capacity comes at a cost: metabolic consumption increases approximately linearly with the frequency of spiking (Laughlin, 2001), thereby placing biophysical constraints on how much gain can be used to improve the quality of neural codes (Balasubramanian et al. 2001; Dayan and Abbott 2001; Laughlin and Sejnowski 2003; Levy and Baxter 1996). Therefore, the entropy of neural responses (or the coding quantity *N*) cannot be increased indiscriminately or sustained indefinitely. Instead, a flexible coding scheme could switch between 'costly but accurate communication' and 'cheap but inaccurate communication' by simply changing the entropy of neural response.

The coding quantity *M* is functionally equivalent to the size of the feature set encoded by a sensory neuron. Thus, in addition to changing encoding capacity via dynamic gain, the reliability of neural communication can also be increased by decreasing the selectivity of neural responses (i.e. decreasing the bandwidth of neural tuning functions). However, there is a non-monotonic relationship between the feature-selectivity of individual sensory neurons and the efficacy of population codes (Series et al. 2004) because a large decrease in feature selectivity will also dramatically reduce the range of inputs that can be encoded. Therefore, unless a subset of the full feature set is selected in advance based on behavioral goals or perceptual learning, changing *M* may not be a good general solution for improving encoding capacity.

In sum, improving the efficacy of a neural code – and concurrently improving neural communication – can be achieved by dynamically changing either the dynamic range of responses (which affect *N*), or by changing the selectivity of neural responses (which affects *M*). An optimal choice of these two parameters is dictated by 3 independent factors: 1) the characteristics of neural noise, 2) the statistics of sensory input, and 3) the behavioral importance

of input being encoded. We posit that the operation of attention (and adaptation) can be formulated in these terms, i.e. they operate by dynamically changing *M* and *N* of the neural code.

1.3 IMPACT OF ATTENTION ON NEURAL CODES: GAIN

Viewed through the lens of Shannon's theorem, we can re-evaluate the functional consequences of attentional gain for neural coding and communication. Spatial attention modulates the tuning curves of orientation-selective neurons in mid-level visual areas like V4 by inducing multiplicative scaling (McAdams and Maunsell 1999), and similar effects have been found as early as primary visual cortex (or *V1*) using BOLD fMRI (Saproo and Serences, 2010). Multiplicative scaling increases the dynamic range of neural responses, which in turn should increase the entropy of the neural code or quantity *N*. According to Shannon's theorem, this increase in *N* should increase the separability of neural codewords and the error tolerance of the neural code. To experimentally determine if the statistical distance between neural codewords measurably increases as a result of attentional modulation, an information-theoretic quantity - mutual information – can be computed between neural responses and stimulus inputs. Mutual Information assesses the amount of response variability (or entropy) that is associated with the stimulus rather than with other unknown factors that are not correlated with the stimulus (termed *noise entropy*), thereby providing a metric of the separability of neural codewords; for a review see (Quian Quiroga and Panzeri 2009). Consistent with the observation of multiplicative gain across many regions of visual cortex, attention leads to an increase in the mutual information of population responses measured using BOLD fMRI as early as area V1 (Saproo and Serences, 2010). Figure 1.3 shows intuitively how attention causes an increase the distance between encoded signals by increasing the dynamic range of neural responses. Given its importance as an early cortical relay station, an increase in mutual information in V1 should improve the transmission of incoming sensory information to all downstream visual and higher-order areas.

Therefore, spatial attention appears to fundamentally change the neural code so that information about behaviorally relevant objects or spatial locations can be relayed across the cortical hierarchy with greater fidelity.

Sensory codes can also be improved by reducing the effective range of the encoded feature set M, which is achieved by reducing the bandwidth of neural tuning function (see Fig 3B). For instance, Spitzer, Moran and Desimone (1988) demonstrated a reduction in the bandwidth of single-unit tuning functions in V4 when monkeys performed a matching task on two oriented gratings. Extending these results, Martinez-Trujillo and Treue (2004) demonstrated that feature-based attention simultaneously increases the gain of cells that were tuned to an attended direction of motion and suppresses cells tuned away from the attended direction of motion. This combination of response enhancement and suppression at the single unit level gives rise to a corresponding increase in amplitude and reduction in bandwidth of the population-level response profile. The increase in amplitude mirrors similar changes in the dynamic range of responses that are observed with spatial attention, and the selective reduction in response bandwidth (decrease in *M*) may further enhance the efficiency of the code by down-regulating neurons that are tuned far from the attended feature and thus have little meaningful information to contribute. However, as mentioned earlier, a sustained reduction in the bandwidth of a population response profile would decrease sensitivity to stimuli that are being suppressed. Thus, there is some tradeoff between a reduction in the range of features that are encoded and the flexibility of sensory systems to react to unexpected stimuli that fall far from the attended stimulus in feature space.

1.4 IMPACT OF ATTENTION ON NEURAL CODES: VARIABILITY

Although changing the neural code by increasing the dynamic range of responses is an efficient way of increasing the fidelity of information transmission, the upper limit on this

increase in fidelity will be determined by the variability and correlational structure of the data (Shadlen and Newsome 1994; 1998; Sompolinsky et al. 2001; White and Carrasco 2011). Instantaneous variability in spiking activity that is unrelated to the stimulus will almost always degrade encoding capacity because it will increase the probability of confusing adjacent neural codewords. Recently, many empirical studies have turned to measuring the variability of single unit and the correlations structure of population-level responses and found that attention reduces the deleterious effects of instantaneous variability and correlated variability in a highly optimal manner that may increase the fidelity of neural codes more vigorously than changes in gain alone. For example, attention reduces trial-by-trial variability in firing rates as quantified by the Fano factor (Mitchell et al. 2007). This reduction in variability increases the signal-to-noise ratio (SNR) of individual neural responses, and suggests that more of the overall entropy in a neural code is dedicated to representing task-relevant information as opposed to irrelevant factors, rendering neural responses more efficient from both a coding and a metabolic standpoint. Ultimately, a reduction in instantaneous variability over a population of redundant neurons will improve the precision of any message transmitted over a noisy neural channel because the individual codewords will be more easily separable when they are decoded. Furthermore, reducing the variability across many neurons may also reduce the amount of noise that is added to signal during the actual transmission process. For instance, if channel noise is positively correlated with the signal strength of transmitted messages as it is in many physical systems (Conti and Wanke 1975; Manwani and Koch 1999; 2001), then any noise in the encoded message would get amplified during transmission, thereby further decreasing the probability of correct decoding by the receiver.

Ideally, a decoder could mitigate any neural variability by averaging over a set of similarly tuned neurons. However, correlated variability between neurons cannot be removed through simple averaging, and thus will limit the precision of point estimates based on a

population response profile (Abbott and Dayan 1999; Averbeck and Lee 2006; Panzeri et al. 1999; Shadlen and Newsome 1994; 1998). Therefore, within decoding schemes based on response pooling, reducing correlated variability should improve the decodability of neural signals. Indeed, recent evidence suggests that attention-mediated decorrelation can predict behavioral performance on a trial-by-trial basis, and may lead to approximately 80% increase in the signal conveyed by the averaged population response, compared to approximately 20% increase in signal conferred by changes in gain (Cohen and Maunsell 2009; Mitchell et al. 2009).

While decreasing noise correlations is intuitively productive in many situations, the relationship between correlated variability and the precision of population codes is highly dependent on the nature of the stimulus, neuron's tuning and the behavioral task. In a now classic theoretical study, Abbott and Dayan (1999) showed that when performing a fine discrimination between two similar but spatially separated features, correlations between dissimilarly tuned neurons should not hamper task performance, and might even be beneficial to the information encoded by a neural population since correlations reduce noise entropy (Abbott and Dayan 1999; Ecker et al. 2011). A recent study (Cohen and Maunsell 2011a) found that while feature attention caused neural populations tuned to the attended feature to become decorrelated, those populations that were tuned away from attended feature became more correlated with attention. One can speculate the functional benefit of this perplexing observation; neural populations tuned away from attended feature also had a reduction in mean response and variability, thus decreasing the overall entropy of their response. Given that correlations lower noise entropy (Abbott and Dayan 1999; Ecker et al. 2011), increasing correlations for neurons with lowered entropy, might conserve their capacity to encode task-relevant information.

While not well understood, correlations between heterogeneous neural populations are thought to be introduced either through common inputs or through shared feedback connections

(Nienborg and Cumming 2010). Therefore, attention might attenuate shared afferents which results in the modulation of correlations between the single units that form a population code. When considering neurons that encode more than one feature, suppression of the irrelevant feature during a task would decorrelate neurons that were jointly encoding the suppressed feature. Such suppression of irrelevant feature would improve the quality of encoding for the relevant feature, since the removal of some information from a sensory signal improves detection of the information that is left behind (Barlow 2001). In support of this notion, Ghose and Maunsell (2008) found that attention either suppressed or enhanced inputs to neurons but did not affect how the inputs were integrated.

An outstanding question concerns the degree to which gain modulations, noise reduction, and noise decorrelation are independent mechanisms of attention. Cohen and Maunsell's (2011a) recently demonstration that noise correlations across a population of neurons are inversely proportional to their mean response, with and without attention, suggests that modulation of correlational structure might simply be a byproduct of gain modulation; since a reduction in mean response would also reduce overall variability for a population of similarly tuned neurons (in spite of changes in Fano factor), an increase in correlations with attention is clearly expected but whether it is independent of gain modulation is not obvious.

1.5 IMPACT OF ADAPTATION ON NEURAL CODES

Thus far we discussed how attention related changes in the dynamic range and variability can make neural codes more robust to intrinsic noise, and thereby improve neural communication. However, the benefit derived from increasing dynamic range is offset by the increased metabolic cost of additional spikes. Furthermore, the increased dynamic range (which also increases encoding capacity of the neural code) is not used towards encoding additional information about the stimuli, and therefore represents wasted representational energy (also called

redundancy). Accordingly, some have argued that redundancy reduction is an important principle of information encoding by neural populations (Attneave 1954; Barlow 1961). Redundancy in neural codes is quantifiable; if the entropy of neural response *H(R)* is greater than the entropy of the encoded stimulus set *H(S)*, the code is said have redundancy amounting to $\frac{H(R)}{H(S)} - 1$. An efficient neural code should also be sensitive to input statistics, such that a change in input entropy dynamically leads to a corresponding change in code entropy, keeping the metabolic demands associated with increased firing rates (or redundancy in the code) to a minimum. A neural encoder that adapts its code based on the statistical nature of the inputs represents a regulatory mechanism that balances demands on communication efficacy and metabolic expenditure, known as adaptation (Fairhall et al. 2001; Rieke et al. 1999; Wark et al. 2007).

Neurons in different modalities - vision, audition, and somatosensation - adapt their response characteristics to reflect the statistical changes in the input, and this adaptation has been shown to maximize rate of information transfer while minimizing metabolic costs (Brenner et al. 2000a; Fairhall et al. 2001; Gutnisky and Dragoi 2008; Laughlin 1981; Maravall et al. 2007; Muller et al. 1999; Sharpee et al. 2006; Wainwright 1999). For example, adaptation to high contrast stimuli produces a graded depression of the contrast response curve in the visual cortex of cats (Albrecht et al. 1984), monkeys (Sclar et al. 1989) and humans (Gardner et al. 2005). This reduction in the dynamic range of the contrast response function can be explained in terms code optimization. Repeated exposure to a stimulus should gradually change internal estimates of relative stimulus probabilities, which translates into a reduction in the entropy of the input set *H(S)*. Therefore, the entropy of the evoked neural response *H(R)* should be lowered in order to minimize excessive firing rates, which can be achieved by reducing the dynamic range of neural responses. Neurons have also been found to optimize information coding by adapting to higher order statistics such as the power spectrum of natural stimuli by adapting their spatiotemporal

filtering properties (Sharpee et al. 2006). The improvement in coding efficiency is even more significant when considering information coded in population responses as compared to individual neurons, since adaptation also impact variability and correlations in neural populations (Cortes et al. 2011; Gutnisky and Dragoi 2008; Muller et al. 1999). This reduction in correlations should theoretically reduce the amount of redundancy in the population code, thereby improving its efficiency (Reich et al. 2001; Schneidman et al. 2003). Accordingly, this improvement in coding efficiency due to decorrelation was observed empirically in V1 by Gutnisky and Dragoi (2008), however they found that decorrelation was highly feature-specific i.e. orientation-selective neurons undergoing the largest change in correlation coefficient preferred orientation that was quite far from the adaptor. Although, both feature-specific decorrelations with adaptation and with those with attention generally improve the efficiency of the population code, it is not clear whether they tap into a common mechanism.

There is an evolutionary basis for adaptive coding as organisms that adapt their neural codes to optimally represent the information in the external world will have a competitive advantage (Geisler and Diehl 2002; Niven and Laughlin 2008). Therefore, an optimal neural code should always try to balance two competing forces: robustness and efficiency. Robustness – due to added redundancy in the neural code – would lead to more accurate information processing, while efficiency – due to minimized redundancy in the neural code – would lead to less metabolically demanding information processing. Based on this principle, we argue that attention and adaptation are closely linked, and in fact arise from the interplay of these two competing forces whose ultimate aim is to process sensory information as reliably and cheaply as possible (See Appendix B for a rudimentary analysis).

1.6 DYNAMIC NEURAL CODING AND NEURAL COMPUTATION

The framework of dynamic neural coding suggests that attention – by means of gain modulation and noise modulation – should increase the efficacy of neural communication, especially long range communication of sensory information. In particular, dynamically improving the neural codes that represent basic features of attended objects in early sensory areas should improve the transmission of sensory information to later sensory areas. For instance, there is a direct flow of motion information between visual areas V1 and MT; therefore, it is conceivable that this flow should be impacted by attention to motion. This is an important yet untested prediction, especially given robust connectivity between the two areas (Born and Bradley 2005) as well as a well established model of how MT derives its complex direction selectivity by collating information across tuned V1 populations (Rust et al. 2006; Simoncelli and Heeger 1998). Furthermore, since V1 and MT are physically separated by a significant cortical distance, the conduit of information between them can be modeled as a noisy channel. Shannon's channel coding theorem provides a framework for empirically evaluating this hypothesis. The theorem states that when communicating over a noisy channel, if T is the set of transmitted codewords and R is the corresponding set of received codewords, then mutual information $I(T,R)$ provides a measure of how well that specific neural code is suited to the channel (Latham and Roudi 2009). Moreover, every channel has a fixed maximum rate of communication – called channel capacity $C = \max_{(\forall T)} I(T,R)$ - that cannot be exceeded no matter what the coding scheme is, but more efficient coding schemes can bring the communication efficacy closer to the channel capacity. Therefore, an increase in the mutual information between two feature-selective neuronal populations in two distinct brain regions that are known to communicate directly, such as V1 and MT, would be direct empirical evidence that gain modulation of neural codes actually leads to more robust communication. Ideally, one would also find that the mutual information

between the response of feature-selective neuronal populations and stimuli increases for each brain region forming such a communication network, implying an improvement in sensory representations within each area.

Theoretically, a "computation" is defined as the repeated transmission of information through space and time, guided by some algorithm (Hilbert and Lopez, 2011; Turing, 1937). Therefore, information loss during communication can limit the depth and accuracy of any computation, but especially for computations performed by a connectionist system such as neural networks. Furthermore, *data-processing inequality* (Cover and Thomas, 1991) ensures that information lost during feed-forward computations in early visual pathways would not be recovered in higher cortical areas involved in decision-making. Therefore, the functional importance of an improvement in neural communication in early sensory areas would be to effectively preserve more sensory information related to attended objects at each processing stage, leading to more accurate computations, while information related to unattended objects is allowed to decay. This improvement in accuracy of computations would also allow decision-making areas such as LIP to accumulate evidence in favor of right decision, more rapidly, leading to faster behavioral responses. Thus, selective processing of information related to relevant stimuli can be understood as attention-induced improvement in neural communication that facilitates the filtering and preservation of relevant information contained in sensory input to more efficiently guide decision making and behavior.

Summary: Neural codes are the protocol used for representing and communicating information in neural systems. Here, we attempted to understand how attention and adaptation dynamically change neural codes in response to rapidly changing behavioral demands and stimulus statistics so that information is more efficiently represented and transferred within sensory systems. This framework of *dynamic neural coding* can provide an algorithmic

understanding of how attention induced neural modulations leads to selective information processing of behaviorally relevant stimuli.

Chapter 1, in part, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material, while Reynolds J., Sharpee T., and Serences J. were co-authors.

1.7 APPENDIX A

Codes: Neural computation requires transferring information across a hierarchy of cortical areas, and like every communication system, a neural code is required that determines how information is physically represented and transferred between a sources and receivers (Huffman 1952). For instance, in a generic communication system (such as in Fig. 2), an encoder generates an output for every input it receives according to a *code*, such that the output is a systematic function of the input (Hamming 1950). The same code must then be used at the receiving end to decode the received information. Codes are ubiquitous, e.g. written and spoken languages (Pinker and Bloom 1990), road signs, and even physical gestures form codes. All of these codes consist of symbols and codewords that are used to convey information from a source (e.g. the speaker of a sentence) to a receiver (e.g. the person listening to the spoken sentence). Similarly, sensory neurons transform input from the external environment into patterns of neural activity in the central nervous system using neural codes. For instance, an orientation-selective neuron in primary visual cortex will fire at a rate that is monotonically related to the orientation of a stimulus in its receptive field (Hubel and Wiesel 1968; 1962), forming a code that can relay information about a low-level stimulus feature (Bialek and Rieke 1992). Neurons with feature-selective tuning functions have been observed in many cortical areas suggesting common coding principles (Hubel and Wiesel 1968) (Albright et al. 1984; Maunsell and Van Essen 1983) (Perrett et al. 1985) (Galambos and Davis 1943; Recanzone et al. 2000) (Ivey et al. 1998). Furthemore, neural codes can be made more robust by using a distributed representation of the stimulus feature via the combined activity of neurons tuned to different feature values, known as a *population code* (Pouget et al. 2000; Seung and Sompolinsky 1993). However, in all cases, the ultimate use of codes is to facilitate the communication of information from a source to a destination.

Ideally, an efficient code should be communicated rapidly, accurately, and with the lowest possible energy overhead. However, since codes are used to communicate over a medium or *channel*, the properties of the channel play a vital role in determining the success of the information transfer. Any noise in the channel will corrupt a transmitted code and will contribute to ambiguity at the receiving end. This noise can have various sources such as the physical properties of the channel, the presence of unpredictable input from other channels (crosstalk), or the use of a channel by multiple senders. Moreover, the noise in a channel can take on many different forms (additive, Gaussian, Poisson, correlated, uncorrelated, etc.), with each having different effect on the efficacy of different codes. Therefore, the construction of a robust code has to account not only for statistical properties of the input being transmitted, but also for the form and distribution of noise over the communication channel.

1.8 APPENDIX B

Joint mechanism of attention and adaptation: Adaptation and attention can interact to improve the efficiency of sensory encoding by dynamically modulating sensory gain in opposing directions. Assume an objective function **G** that the sensory system is trying to optimize using the operations of attention and adaptation. Let **S** be the set of all stimuli competing for attention, out of which stimulus **s** is being attended. **G** will depend on following 3 factors: 1) the behavioral relevance of stimuli in each spatial location (**P)**, where **P** is a probability distribution over all stimuli **S** available in the visual field, with higher **P(s)** representing higher behavioral relevance. 2) The entropy of the neural code used at each spatial location (**I)**. 3) The entropy of the sensory input being encoded (**J)**.

Whenever **J(s) ≤ I(s)**, the neural code used for **s** has redundancy amounting to **I(s) /J(s) - 1**. For behaviorally relevant stimuli, this redundancy is useful since it would facilitate robust neural communication. We assume that the introduction of redundancy would be the handiwork of selective attention via a modulation of dynamic range of neural responses e.g. multiplicative scaling of feature tuning functions. However, for behaviorally irrelevant objects, any attention-induced redundancy would be un-necessary. Since added redundancy is metabolically costly, it should be matched across different stimulus statistics. For instance, a reduction in stimulus entropy **J(s)** – such as with adaptation – should necessitate a reduction in code entropy I(s) in order to maintain an adequate level of redundancy without being incurring unnecessary metabolic cost. Therefore, the reduction in redundancy according to changing input statistics would be the handiwork of adaptation processes, since adaptation is known to reduce the dynamic range of neural responses. Therefore, attention and adaptation are affecting the dynamic range of neural responses in an opposing push-pull fashion, and it is this interaction that we suggest is critical to optimal information processing of incoming sensory data.

Ideally, the sensory system should try to maximize the quantity **P*I**, i.e. maximize the representational capacity (e.g. bits) of neural codes devoted to behaviorally relevant objects. Suppose, the peak of **P** changes from **s1** to **s2** due to re-assignment of behavioral needs, then **I** should also change its distribution such that **I(s2)** > **I(s1)**, since the product of **P** and **I** should be maximal at all times. Furthermore, the quantity **I/J** should be minimized since it reflects unwarranted redundancy over the ensemble of input stimuli. Suppose, due to repeated presentation **J(s1)** becomes lower than **I(s1)**, then the neural code should be modified such that **I(s1)** is proportionately reduced given its behavioral importance **P(s1)**. We suppose that the operation of attention is to maximize **P*I**, and adaptation to minimize **I/J**, with higher brain areas trying to optimize these two operations, given some desired error tolerance of the code ***L***, thereby producing an optimal neural code $I^* \sim opt_I(\max(P * I), \min\left(\frac{I}{J}\right), L)$

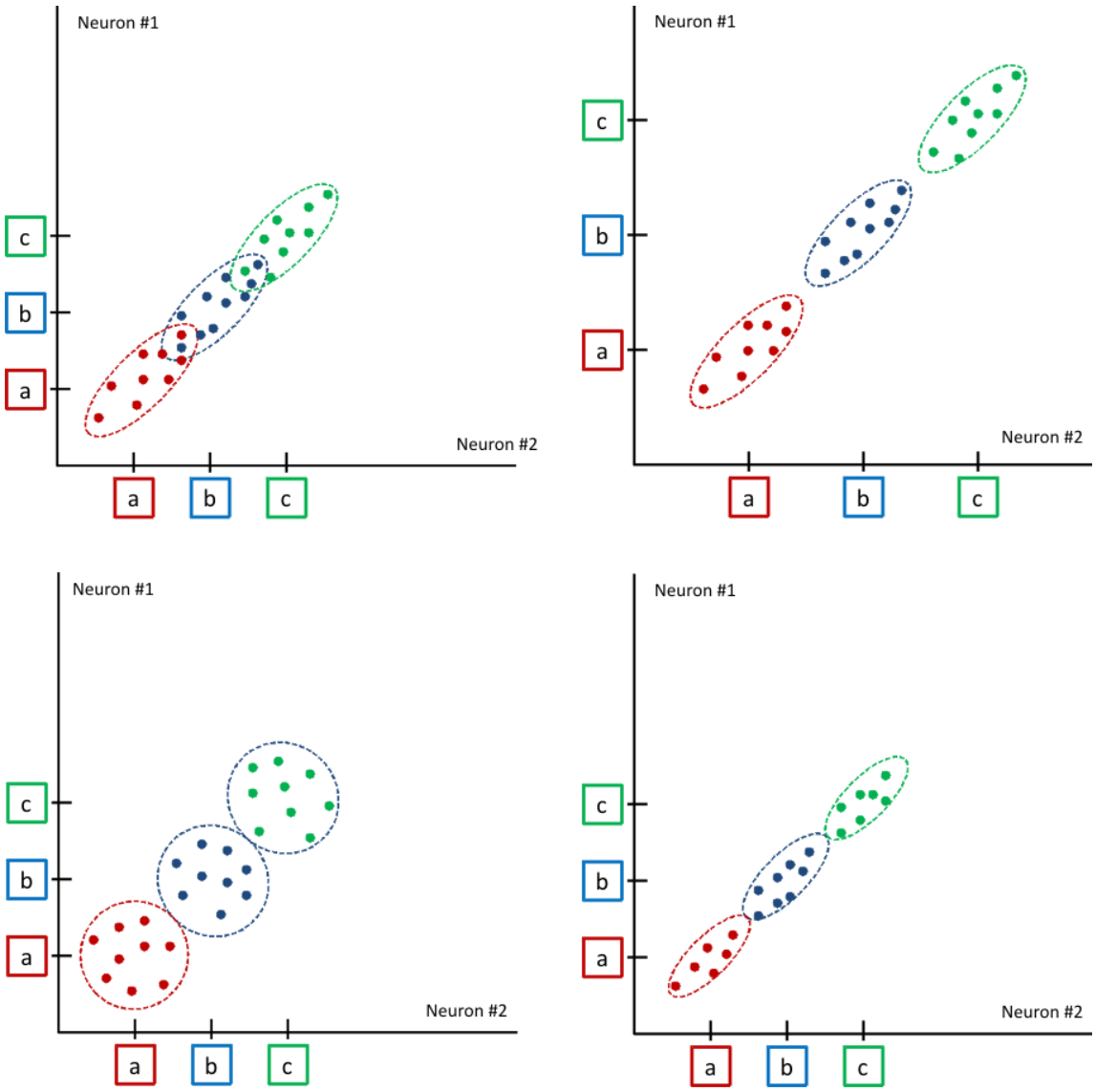


Figure 1.1 A graphical description of the three putative mechanisms of attention viz. gain modulation, noise reduction and decorrelation. Top left panel shows the joint encoding 3 distinct features (a, b and c) by 2 sensory neurons with shared tuning. The degree of overlap between response clouds provides a measure of decoding error. Top right panel depicts attention induced gain, which separates the response clouds and increases the accuracy of the population code. Bottom left panel depicts attention induced decorrelation of neurons; this also decreases the response cloud overlap and increases decoding accuracy. Other panels graphically show the 3 putative mechanisms for attention. Bottom right panel shows a reduction in neural variability in each neuron, even without affecting the degree of direction of correlation, which still decreases the response cloud overlap and thus improves decoding error.

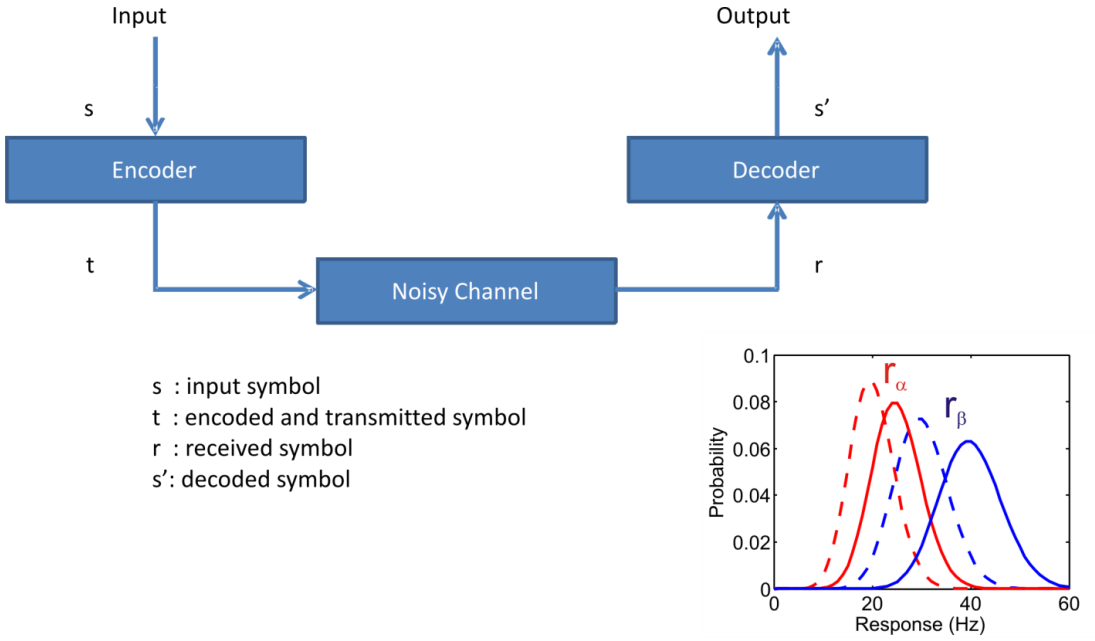


Figure 1.2 Schematic of the communication of codes over a noisy channel. The source input *s* is encoded by the encoder into codeword *t* using code *t=g(s)*, which usually involves adding redundancy to make the coding scheme noise-resistant. Due to channel noise, the transmitted codeword *t* may be received by the decoder as codeword *r* such that $t \neq r$. However, if the code is noise-resistant, such that mutual information *I(t,r)* is high, the decoded output *s'* would be equal to the original input *s*. We assume that decoding is done using the inverse mapping function $s' = g'(r)$ by an optimal decoder. For example, consider a hypothetical noisy channel where the input α is encoded as mean firing rate of 20Hz and the input β is encoded as 30Hz, by a neural encoder. The dotted lines give the probability distribution of signal received (r_α and r_β) at the decoder, when α and β were input respectively. Here, the distribution of r_α and r_β follow Poisson noise model. The area under the intersection of the two curves, gives the probability of incorrect decoding. If the code is changed such that the two inputs (α and β) are remapped to 25Hz and 40Hz respectively, the probability of incorrect decoding, given by the area under the intersection of solid red and blue curves, is lower than before. This remapping of the code increased the distance between codewords (mean firing rate) for the two inputs, and led to more error tolerance during communication.

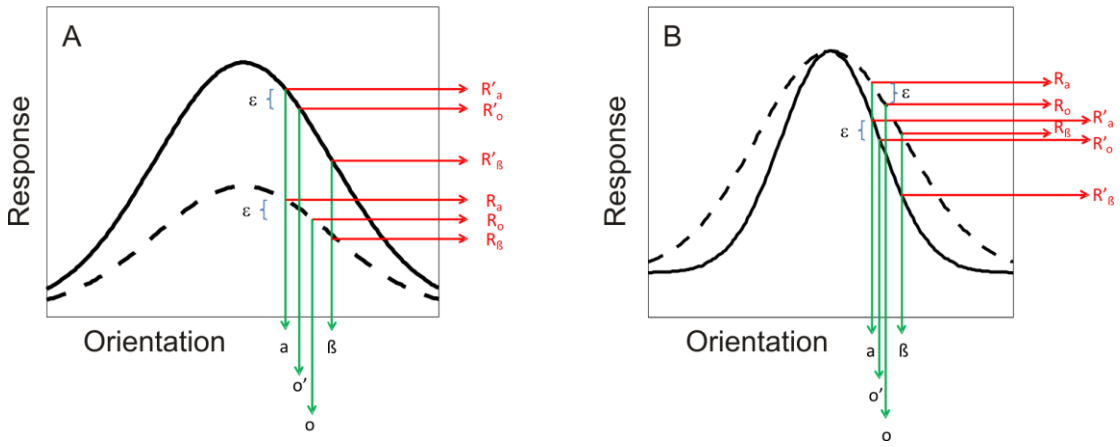


Figure 1.3 A) Hypothetical depiction of how multiplicative scaling of the tuning curve can increase robustness to noise by increasing the distance between codewords. Suppose the tuning curve (dotted line) encodes orientation and must distinguish between two orientations, α and β. Let R'_α be the mean response (codeword) encoded for orientation α. Furthermore, due to channel noise ε, the receiver receives codeword R_o when codeword R_α is transmitted. In this case, the receiver cannot decode it correctly as α, since R_o is equidistant from the encoded responses for α and β. However, when the same tuning function undergoes multiplicative scaling (solid line), the distance between the encoded response of α and β is increased. For the same channel error ε, the received codeword R'_o can now be correctly decoded as the input α, since the received codeword is closer to R'_α (the encoded response for α) than R'_β (the encoded response for β).
B) Similar to panel A but shows how bandwidth reduction of the tuning curve can in increase robustness to noise by increasing the distance between codewords.

CHAPTER 2

SPATIAL ATTENTION IMPROVES THE QUALITY OF POPULATION CODES IN HUMAN VISUAL CORTEX

Selective attention enables sensory input from behaviorally relevant stimuli to be processed in greater detail, so that these stimuli can more accurately influence thoughts, actions and future goals. Attention has been shown to modulate the spiking activity of single feature-selective neurons that encode basic stimulus properties (color, orientation, etc.). However, the combined output from many such neurons is required to form stable representations of relevant objects, and little empirical work has formally investigated the relationship between attentional modulations on population responses and improvements in encoding precision. Here, we used fMRI and voxel-based feature tuning functions to show that spatial attention induces a multiplicative scaling in orientation-selective population response profiles in early visual cortex. In turn, this multiplicative scaling correlates with an improvement in encoding precision, as evidenced by a concurrent increase in the mutual information between population responses and the orientation of attended stimuli. These data therefore demonstrate how multiplicative scaling of neural responses provides at least one mechanism by which spatial attention may improve the encoding precision of population codes. Increased encoding precision in early visual areas may then enhance the speed and accuracy of perceptual decisions computed by higher-order neural mechanisms.

2.1 INTRODUCTION

In many situations an observer must selectively enhance the perceived visual detail in a behaviorally relevant portion of the periphery at the expense of detail in other peripheral locations. The perceptual benefits obtained by covertly monitoring relevant para-foveal objects

depends on modulating the neural responses that encode basic visual features such as edge contours and colors, a phenomenon referred to as *selective attention* (Moran and Desimone, 1985; Desimone and Duncan, 1995; Connor et al., 1997). The spatial component of selective attention (*spatial attention*) increases neural activity in retinotopically corresponding cortical regions for all features in that spatial location, much like a spotlight illuminating only a part of a theater stage (Posner et al., 1980; Koch and Ullman, 1985; Tsotsos et al., 1995; Boynton, 2005). But how exactly is the encoding of relevant features impacted by spatial attention and how might it enhance perception? Single-unit recording data collected in non-human primates suggests that spatial attention primarily acts to multiply feature tuning functions by a constant factor (referred to here as *multiplicative scaling;* see McAdams and Maunsell, 1999; Treue and Martinez Trujillo, 1999; Fig. 1A). In addition to multiplicative scaling, spatial attention is also thought to induce a feature non-specific increase in response amplitudes, which can occur even in the absence of a stimulus (referred to here as *additive scaling*, also known as a *baseline shift;* Luck et al., 1997; Kastner et al., 1999; see Fig. 1B).

Functionally, spatial attention is thought to improve the quality of perceptual representations in part by selectively increasing the firing rates of neurons tuned to the attended stimulus (assuming approximately Poisson neurophysiological noise; Shadlen and Newsome, 1994; McAdams and Maunsell, 1999; Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004). Even though single-unit recording studies have demonstrated such attentional modulations, coherent perceptual representations are thought to be based on populations of sensory neurons working in concert (Pouget et al., 2003; Averbeck et al., 2006; Butts and Goldman, 2006; Kang et al., 2004; Jazayeri and Movshon, 2006). Therefore, the goal of the present study was to explicitly examine the theoretical and empirical link between spatial attention and the encoding precision of population responses in early areas of human visual cortex.

We used fMRI to define feature-selective 'voxel tuning functions, and then used information theoretic measures (Nevado et al., 2004; Panzeri et al., 2008) to infer changes in the encoding precision of population responses due to spatial attention. We find that spatial attention increases the amount of information encoded in early visual cortices about the features of an attended stimulus, primarily by modulating the multiplicative scaling of population response profiles. Since the quality of a population code can be assessed using information theory (Doya, 2007; Mackay 2003), the observed multiplicative scaling provides a mechanism by which spatial attention improves encoding precision of population codes. In turn, the improved quality of stimulus representations in early visual areas should facilitate the read-out of information by higher brain areas during perceptual decision making (Ditterich et al., 2003; Jazayeri and Movshon, 2006; Newsome et al., 1989; Shadlen and Newsome, 2001).

2.2 METHODS

Subjects: Eight neurologically healthy subjects between the age of 18 and 30 were recruited from the University of California, Irvine (UCI) community to participate in the experiment. Data from one subject were subsequently discarded because of an inability to obtain robust retinotopic maps. Each subject gave written informed consent as per Institutional Review Board requirements at UCI, and completed 1hr of training outside the scanner before completing two 1.5 hr scanning sessions held on separate days. Thus, the data presented here represent a total of 14 scanning sessions.

Selective attention scans: Subjects were instructed to maintain fixation on a central fixation point (white in color, subtending 0.5° of visual angle) that persisted on the screen for the duration of each scan (where a 'scan' refers to a single fMRI data collection run lasting 410s). Each scan consisted of several trials; on each 10s trial, a full contrast grayscale sinusoidal annular grating (0.5 cycles/degree) was flickered at 2 Hz (250ms on, 250ms off) in one of 8 possible

orientations (0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, and 157.5°, where 0° is horizontal, see Fig. 2A), which defined a single stimulus. The spatial phase of the stimulus grating was randomly shifted every 250ms within each trial to attenuate adaption and apparent motion. The entire stimulus (left-to-right) subtended 22.5° of visual angle with a central circular aperture (11.2° diameter) removed around fixation; the annulus was disconnected at top and bottom so that each semi-circular stimulus grating only occupied one visual hemi-field (each stimulus 'wedge' occupied 40% of the full annulus). The order of stimulus orientations was randomized on each scan with the constraint that the same orientation could not be presented on successive trials. Stimulus gratings in both hemi-fields had the same orientation to ensure that any effects of global-feature-based attention remained constant throughout the task (Treue and Martinez Trujillo, 1999; Saenz et al., 2002; Serences and Boynton, 2007). Subjects attended to the stimulus grating presented in either the left or the right hemi-field in response to a horizontal cue extending 1° of visual angle from the central fixation point (Fig. 2B). This cue remained on the screen for the duration of each trial. Subjects were instructed to respond when the contrast of the stimulus grating in the attended hemi-field decreased slightly, which is henceforth referred to as a target event (subjects were to ignore equally frequent contrast changes that occurred in the unattended hemifield, which was done to equate sensory factors between the attended and unattended stimuli). The contrast reduction that defined a target was titrated on an individual basis so that the hit rate remained at approximately 75-80% over the course of all scanning sessions (a 'hit' was defined as a response that occurred up to 1s after the presentation of a target; subjects false alarmed infrequently on only 1.9%±0.7% S.E.M. of contrast changes in the distractor stimulus). Each target was presented for a single 250ms frame, and there were 4 targets in each trial. The timing of each target was pseudo-randomly determined with the following constraints: each target was separated from the previous one by at least 1.5s, and targets were restricted to a temporal window of 1-9s following the onset of the trial. Each trial was separated from the next by a blank

500ms inter-trial-interval of passive fixation. Observers completed five or six attention scans per 1.5 hour scanning session; each scan lasted 410s and contained 4 presentations of each stimulus orientation (32 in total) along with 6 pseudo-randomly interleaved 'null' trials in which only the central fixation point was visible for the entire 10s trial interval. These null trials were presented to provide a common baseline for comparison of responses evoked by attended and unattended stimuli.

Independent functional localizer scans: Two independent functional localizer scans were run during each scanning session to identify voxels within each retinotopically-organized visual area that responded to the spatial position occupied by the stimuli in the left and right hemi-fields during the attention scans. The stimuli used during localizer scans were identical to the stimuli used during attention scans; however, only one hemi-field was stimulated on each trial (Fig. 2C). Subjects were asked to fixate on the central fixation point and direct their attention to the presented stimulus. There was a blank inter-trial-interval of 10s following each stimulus presentation and each of the 8 orientations was presented twice during each localizer scan (total scan duration: 320s).

Retinotopic mapping procedures: Retinotopic mapping data were obtained in one to two scans per subject, using a checkerboard stimulus and standard presentation parameters (stimulus flickering at 8 Hz and subtending 60° of polar angle; Engel et al., 1994; Sereno et al., 1995). This procedure was used to identify ventral visual areas V1, V2v, V3v, and hV4. Our high-resolution scanning protocol did not provide sufficient coverage to acquire data from dorsal occipital areas V2d, V3d and V3a or visual areas in parietal and frontal cortex. To aid in the visualization of early visual cortical areas, we projected the retinotopic mapping data onto a computationally inflated representation of each subject's gray/ white matter boundary.

fMRI data acquisition and analysis: MRI scanning was carried out on a Philips Achieva 3-Tesla scanner equipped with an 8-channel SENSE head coil at the John Tu and Thomas Yuen Center for Functional Onco-Imaging, University of California, Irvine. Anatomical images were acquired using a MPRAGE T1-weighted sequence that yielded images with a $1mm^3$ resolution (TR/TE=11/3.3 ms, TI=1100ms, 150 slices, flip angle=18^0 with no SENSE acceleration). Functional images were acquired using a gradient echo planar imaging (EPI) pulse sequence, which covered the occipital lobe with 25 oblique transverse slices. Slices were acquired in sequential order with 1.5mm thickness and 0.5mm gap to avoid slice crosstalk; thus a 50mm thick slab was acquired (TR = 2500 ms, TE = 30ms, flip angle = 70°, image matrix = 120 (AP) × 92 (RL), with FOV = 240mm (AP) × 180mm (RL), SENSE factor = 2, voxel size = 2mm × 2mm × 1.5mm).

Data analysis was performed using BrainVoyager QX (v 1.86; Brain Innovation, Maastricht, The Netherlands) and custom time-series analysis routines written in Matlab (version 7.1; The Math Works, Natick, Massachusetts). All EPI images were slice-time corrected, motion-corrected (both within and between scans) and high-pass filtered (3 cycles/run) to remove low frequency temporal components from the time-series.

Region of interest selection procedure: To identify voxels that responded to the retinotopic position of the stimulus aperture, data from the functional localizer scans were analyzed using a GLM that contained two regressors marking each 10s epoch of stimulation in the left and right hemifields (a boxcar convolved with a standard double-gamma function as implemented in Brain Voyager: time to peak 5s, undershoot ratio 6, time to undershoot peak 15s; Boynton et al., 1996). Voxels within each visual area that responded more to one epoch of stimulation compared to the other were retained for further analysis if they passed a threshold of $p<.01$, corrected for multiple comparisons using the False Discovery Rate (FDR) algorithm

implemented in BrainVoyager QX (see Genovese et al., 2002). See Table 1 for the number of voxels within each visual area that passed this threshold.

Voxel tuning functions: We used voxel tuning functions (VTFs) to evaluate feature-selective BOLD responses in early areas of visual cortex. The feature selectivity is thought to be indirectly determined by biases in the distribution of orientation-selective neurons (e.g. columns in V1) that are idiosyncratically sampled within each voxel (Haynes and Rees, 2005; Kamitani and Tong, 2005; Sasaki et al., 2006; Kay et al., 2008; Serences et al., 2009; Swisher et al., 2010). In turn, if a voxel has a bias in the number of neurons that prefer a particular feature, then that voxel should exhibit a weak but detectable feature-selective response bias.

Estimating orientation-specific responses: First, the time series from each voxel was normalized by subtracting the mean activation level for that voxel on a scan-by-scan basis. Next, the magnitude of the BOLD response in each voxel was estimated using a GLM that contained regressors for attended and unattended stimuli rendered in each of the 8 possible orientations (boxcar model of stimulus sequence convolved with standard difference-of-two gamma functions). For example, on a trial where the subject was attending to the stimulus in the left hemi-field, an 'attended' response would be measured from all voxels in the contralateral right hemisphere, and an 'unattended' response would be measured in the ipsilateral left hemisphere.

Assigning orientation preference to each voxel: The orientation preference of each voxel was assigned using a leave-one-out procedure to ensure that the resulting tuning functions reflect reliable changes in signal as opposed to idiosyncratic noise. First, each voxel from a visual area was assigned to one of eight orientation-preference bins based on data from all scans except one; orientation preference was heuristically established by determining the orientation that evoked the largest response after removing the mean response across all voxels in response to each orientation (see Supplemental Figure 7 for an alternate heuristic). This mean-subtraction was

performed to correct for main effects of stimulus orientation that had a common influence on the response of every voxel, thereby emphasizing the *differential* pattern of responses across voxels (see e.g., Haxby et al., 2001). Supplemental Figure 11 shows the mean response amplitude to each orientation in each visual area, as well as the distribution of orientation preferences with and without this correction. Supplemental Figure 8 shows the fully analyzed data without removing the main effect of orientation: qualitatively similar results are observed.

After determining the orientation preference of each voxel, we used data from the held-out scan to compute the response of voxels in each bin to both attended and unattended stimuli rendered in all eight possible orientations (thus producing a VTF). This hold-one-out procedure was then repeated using all unique permutations of holding one scan out, and the final VTF for a given visual area in a given subject reflects the average across all permutations. Finally, VTFs from each visual area were averaged across left and right hemispheres as no systematic asymmetries were observed.

Characterizing VTFs using a circular Gaussian: Voxel tuning functions were mathematically characterized as

$$B(\Theta) = \alpha + \beta * f(\kappa,\Theta) \quad \text{1a.}$$

$$f(\kappa,\Theta) = \exp(-([(\pi/2 + \Theta - \Theta_{pref}) \bmod \pi] - \pi/2)^2 / \kappa^2) \quad \text{1b.}$$

where $f(.)$ is a circular Gaussian function (Dayan and Abbott, 2001; Pouget et al., 2003; Nevado et al.,2004), α is the additive scaling parameter, β is the multiplicative scaling parameter, and k is the concentration parameter or bandwidth. Once the distribution for $B(\Theta)$ had been estimated using above-mentioned procedure, an iterative curve-fitting algorithm was used to derive best-fit values for α, β and k. Thus, voxel tuning curves were constructed reflecting BOLD activation under attention $B_a(\Theta)$ and without attention $B_u(\Theta)$ such that

$$B_a(\Theta) = \alpha_a + \beta_a * f(\kappa_a, \Theta) \quad \text{2a.}$$

$$B_u(\Theta) = \alpha_u + \beta_u * f(\kappa_u, \Theta) \quad \text{2b.}$$

The effect of spatial attention was then characterized as

$$\text{Additive scaling } (\alpha') = \alpha_a - \alpha_u \quad \text{3a.}$$

$$\text{Multiplicative scaling } (\beta') = \beta_a - \beta_u \quad \text{3b.}$$

$$\text{Bandwidth scaling } (\kappa') = \kappa_a - \kappa_u \quad \text{3c.}$$

The model was fit to data from each visual area in each subject using a gradient descent algorithm implemented in the Matlab Optimization Toolbox (built around the "fminsearch" function that employs Nelder-Mead Simplex Direct Search; Lagarias, 1998). Because the starting value assigned to each parameter can influence the final outcome of gradient descent algorithms (in case of many local minima), each fit was performed 60 times with a new seed for each of the three free parameters drawn randomly from a normal distribution in an attempt to find the globally optimal parameter values. The mean of each normal distribution used to generate the seeds was set as follows: (1) the minimum of the data for the additive scaling parameter, (2) the maximum response minus the minimum response for the multiplicative scaling parameter, and (3) .5 for the bandwidth parameter. The standard deviation of the seed distribution was set to the square root of the respective mean for that distribution. The iteration with the lowest overall root mean square error (RMSE) across all 60 iterations in a given subject was then used in the final analysis (mean RMSE: 0.02157, ±SEM: 0.002162, indicating that the fits were excellent overall). A validation procedure in which 500 iterations of the analysis were run with randomized labels was also carried out: this resulted in a more than 2-fold increase in RMSE, further supporting the significance of the observed Gaussian shape of the VTFs (mean RMSE with randomized labels: 0.049).

Since all stimuli had the same luminance contrast, we could not model the tuning function with contrast as a free parameter. Thus, we cannot rule out the possibility that the values of α, β and k have some dependence on stimulus contrast, and hence these parameters could be a function of contrast.

We also investigated the appropriateness of an alternate model $B(\Theta) = \beta*(f(\kappa, \Theta)+\alpha)$ where the multiplicative scaling parameter also operates on the additive scaling parameter. When reasonable constraints on allowable parameter values are imposed (e.g. non-negative additive scaling factor), then the model described in 2a and 2b proved to be a better fit to the data (Supplemental Fig. 4).

Mutual information measures: In the present paper, we use an information-theoretic quantity Mutual Information (MI) to both rank order voxels based on the theoretical homogeneity of comprising neural populations, as well as to evaluate the influence of attention on the quality of neural codes (the methods described here follow those described in Serences et al., 2009). MI has the advantages that it indexes the information conveyed by a voxel about the stimulus, and makes no a priori assumptions about the precise shape of the response distribution (Fuhrmann Alpert et al., 2007; Panzeri et al., 2008; Serences et al., 2009). MI is based on Shannon's entropy of the BOLD responses in each voxel, which is a measure of response uncertainty across all stimulus orientations.

$$H(B) = -\sum_{b \in B} p(b) \log_2 p(b) \qquad 4.$$

To compute Shannon's entropy, we converted the continuous BOLD response into a discrete variable (B) by dividing the range of responses into a set of equidistant bins (b) of sufficiently small size (Cover and Thomas, 1991). In this formulation, $p(b)$ is the frequency with which a response falls into bin b divided by the total number of responses measured from a given voxel.

The bins were statically defined i.e. fixed for all data sets, and spanned the entire range of values encountered in the data. Before discretizing BOLD responses for a given voxel, we collapsed data collected in response to both attended and unattended stimuli after subtracting out the mean activation levels of 'attended' and 'unattended' conditions. This subtraction was done because additive shifts due to attention would have induced error in the process of binning BOLD responses during the computation of *p*(b). Next, we computed conditional entropy *p(b|θ)*, which yields a measure of response uncertainty given the knowledge of stimulus orientation (*θ*). If there is a dependence between stimulus orientation and observed BOLD responses, the introduction of *Θ* as a conditional random variable should reduce the uncertainty and hence the entropy of the random variable *B*

$$H(B \mid \Theta) = -\sum_{\theta \in \Theta} p(\theta) \sum_{b \in B} p(b \mid \theta) \log_2 p(b \mid \theta) \qquad 5.$$

The information content carried by each voxel can then be defined as the reduction in uncertainty for each voxel's BOLD response given the stimulus orientation, or mutual information (MI). Since we are using logarithm to the base 2 in our calculations, the unit of measure is *bit*.

$$I(B;\Theta) = H(B) - H(B \mid \Theta) \qquad 6.$$

Bayes' rule ensures that the equation is symmetric, so that *I(B|Θ)* is equal to *I(Θ|B)*, or the reduction in uncertainty about stimulus orientation given a distribution of BOLD responses. Therefore, if there is a strong dependence of the distribution of BOLD responses (*B*) and stimulus orientation (*θ*) for a particular voxel, then that voxel will yield a high MI value. If they are completely dependent – such that knowing one gives complete information about the other – then MI would be equal to the entropy of either one of them. On the other hand, if the distribution of BOLD responses and stimulus orientation are completely independent, then that voxel will yield a

MI value of zero. In the present experiment, where 8 different stimulus orientations were employed, the maximum value of MI is 3 ($\log_2 8$).

Normalized MI: For the analysis of modulation in MI with attention (Fig. 9), we used the normalized version of mutual information (Kojadinovic, 2005) to normalize the values of voxel MI derived from different subjects to range between 0 and 1 before the data were averaged.

$$I^*(B;\Theta) = I(B;\Theta)/H(B) \qquad 7.$$

A value of 1 for I^* means that knowing B gives complete information about the value of Θ, or vice versa. Conversely, a value of 0 implies that knowing B provides no information about Θ, *or* complete independence between the distributions of B and Θ.

Relationship between the shape of a VTF and M: MI depends on the difference between overall response variability (total entropy or $H(B)$) and response variability that is attributable to factors uncorrelated with the stimulus (noise entropy or $H(B|\Theta)$). The difference between total entropy and noise entropy thus represents how much variability in the BOLD responses is attributable to changes in the stimulus, which indicates how much signal, or useful information, is present in the data. For an informative voxel, entropy $H(B)$ should be high, implying that the magnitude of the BOLD response changes substantially as a function of stimulus orientation. Thus a voxel that has a peaked tuning function would be more informative than a voxel that has a flat tuning function, all else being equal (although the relationship between MI and tuning function bandwidth is non-monotonic, such that flat or overly peaked voxels both have a lower MI than voxels with some optimal intermediate bandwidth). Conditional entropy $H(B|\Theta)$ for an informative voxel should be relatively low (with respect to the entropy $H(B)$) because that voxel should yield similar BOLD responses on each successive presentation of the same orientation, implying low noise.

Using MI as a criterion to select representative voxels: We hypothesized that the magnitude of MI in a voxel should be directly proportional to the underlying distribution of feature-selectivity at the neural level, making only the assumption that there is a monotonic relationship between neural activity and the magnitude of the BOLD response (Logothetis et al., 2001; Logothetis, 2003). Consider two hypothetical and extreme types of voxel: (1) a voxel with very weak orientation selectivity, and (2) a voxel with strong selectivity and a robust VTF. In the first case, weak selectivity might be caused either by poor SNR (e.g. if the voxel partially samples white matter), or by a relatively heterogeneous distribution of orientation-selective neurons within the voxel, which would result in a null orientation preference because the voxel would respond equally well to all orientations. In turn, this non-selective response profile would have low entropy and correspondingly low MI. Therefore, we argue that voxels with weak VTFs (characterized by low MI) are difficult to interpret as multiple factors may contribute to poor orientation tuning and consequently denature the observable effects of attention. The opposite is true in a relatively homogenous voxel containing many neurons tuned to a specific orientation. Such a voxel should have a highly selective and peaked VTF, high total entropy, and a correspondingly high MI (assuming near-Poisson or additive noise). Therefore, voxels with high MI should contain a *relatively* homogeneous sample of orientation-selective neurons and hence provide the best insight into the operating characteristics of attention on specific neural populations (see Supplemental Figures 1 and 2 for a more formal description of this model, as well as Nevado et al., 2004).

Based on this logic, voxels were rank ordered based on their MI score and orientation preference within all scans save one, and then the VTFs were generated based on data from the remaining scan (rank ordering always done separately for each visual area, session, and subject). Different quartiles out of this rank-ordered list were then selectively analyzed because quartiles

were large enough intervals to provide an adequate number of voxels in each bin to compute reliable tuning curves (see e.g. Figs. 6 and 7, Supplemental Fig. 3). Note that the mutual information measure suffers from the problem of bias in estimated values when data are limited (Panzeri et al., 2007). However, since we used MI values only to rank-order voxels derived from the same dataset, the exact MI value of each voxel becomes immaterial as long as it gives an accurate representation of the relative information content of that voxel vis-à-vis other voxels in its orientation-preference bin (and visual area, session, and subject).

Relationship between attentional gain and MI: Spatial attention can impact a Gaussian tuning function in 3 ways: additive scaling, multiplicative scaling and bandwidth scaling. Multiplicative scaling should increase MI since it increases the dynamic range of mean responses, which in turn increases the entropy of responses *H(B)*. Even under the assumption of Poisson noise, multiplicative scaling will cause the entropy of responses to increase faster than the corresponding increase in noise entropy, leading to an overall increase in MI. This also holds for additive Gaussian noise, as long as the variance increases by a factor smaller than the increase in the mean BOLD response. In contrast, additive scaling of a tuning function will not increase entropy because it simply translates the tuning curve up or down without increasing the dynamic range of mean responses. If the noise scales with mean response, e.g. Poisson noise, additive scaling might actually lower MI because the noise entropy will increase while the overall entropy will remain constant. Finally, the relationship between MI and the bandwidth of a tuning function is non-monotonic. A decrease in bandwidth would increase MI if the bandwidth was non-optimally wide (e.g. a flat tuning function). However a decrease in bandwidth beyond an optimal intermediate point would also lower MI because the tuning function would only convey information about a highly restricted range of stimulus values. Thus, an increase in bandwidth – to a point – would help if the bandwidth was overly narrow.

Using MI of BOLD responses to evaluate the quality of population codes: If BOLD responses are approximately linearly related to changes in underlying neuronal firing (Heeger et al. 2000; Logothetis et al., 2001; Logothetis, 2003), then an increase in a voxel's MI with attention implies an increase in information conveyed by the population of neurons that are contained within that voxel. This relationship holds even if we allow a non-linear relationship between neural activity and the BOLD response, as long as the relationship remains monotonic (i.e., if an increase in neural spiking activity leads to some increase in BOLD amplitude; Nevado et al., 2004). Under this relaxed assumption of monotonicity, the BOLD response profile can be thought of as a filtered version of the neural population response profile, filtered through some monotonic mapping function, forming a Markov Chain. The data inequality theorem (Cover and Thomas, 1991) mandates that for such a Markov chain, the MI of a voxel must be equal or less than the MI of the neural population. Therefore, even the most informative voxel can never have higher information in its BOLD responses than the MI that exists in the underlying neural population activity, and thus provides an upper bound on any estimates of the information content of population responses. As a result, the absolute value of the increase in MI with attention exhibited at the voxel level is not expected to be equal to the actual increase in MI at the neural level; however, the qualitative relationship between changes in MI at the neuronal and voxel level should be preserved.

Eye tracking: Eye tracking data were acquired with an ASL-LRO MR compatible system that tracked the right pupil of three subjects during scanning (one subject in both sessions, and the others in a single session each) at 60Hz. Data were analyzed offline using the ILAB analysis toolbox implemented in Matlab (Gitelman, 2002). First, the raw data were binned into epochs corresponding to the 10s stimulus presentation interval from each trial. Blinks (periods when the pupil disappeared), as well as 5 samples on either side of each blink, were then marked and removed from the epoched data. Saccadic eye movements were then identified when the velocity

of the eye exceeded 30°/s for at least 2 samples; a minimum saccade distance of 1° was also imposed, as this is the manufacturer-stated resolution of the eye-tracker. Supplemental Figure 5 shows the mean X,Y position of the saccade endpoints for each of the eight stimulus orientations; no systematic differences as a function of the attended location were observed.

We could acquire eye-tracking data from only 3 subjects because of various technical impediments that made it difficult to acquire steady measurements from several subjects. In such cases where data could not be acquired consistently, subject fixation was monitored manually by both authors (using the eye-tracker camera) to ensure that subjects did not make systematic saccades with the attentional cues and were fixating consistently on the central fixation spot.

2.3 RESULTS

To characterize the impact of spatial attention on population response profiles, we used the spatial attention task shown in Fig. 2. We first generated voxel tuning functions (VTFs) with and without spatial attention within each visual area (Fig. 3, see Materials and Methods – *Voxel-based tuning functions*). One-way repeated measures ANOVAs revealed significant orientation selectivity of the VTFs in all areas (after collapsing across attended and unattended trials, V1: $F(7,42)=18.3$, $p<.001$; V2v: $F(7,42)=19.4$, $p<.001$; V3v: $F(7,42)=5.9$, $p<.001$; hV4: $F(7,42)=3.7$, $p<.005$).

When considering all spatially selective voxels identified using the independent functional localizer scans, we found that spatial attention primarily produced an additive scaling of VTFs that was highly significant in all four visual areas. This magnitude of the additive scaling also increased modestly from V1 to hV4 (Fig. 4, $F(3,18)=3.13$, $p=.051$, marginally significant one-way repeated measures ANOVA). This monotonic rise in additive scaling along the ventral pathway is consistent with previous reports in human visual cortex (Kastner et al., 1998; Ress et

al., 2000; O'Connor et al., 2002) and might be driven by increasing receptive field size and a corresponding increase in the amount of within-receptive field competition, which is thought to play a critical role in determining the magnitude of attentional modulations (Desimone and Duncan, 1995; Kastner et al., 1998; Kastner and Ungerleider, 2001).

We next used an information-theoretic measure *mutual information (MI)* to identify those voxels that are theoretically most representative of changes within specific subsets of orientation-selective neurons (see Methods: *Using MI as a criterion to select representative voxels*; Cover and Thomas, 1991; Borst and Theunissen, 1999; Fuhrmann Alpert et al., 2007; Panzeri et al., 2008; Serences et al., 2009). A three-way repeated measures ANOVA with visual area (V1-hV4), attention (attended, unattended) and MI bin (1st-4th quartile) revealed an overall increase in the additive scaling parameter with attention ($F(1,6)=28.5$, $p<0.002$), and an increase in the additive scaling parameter with increasing MI ($F(3,18)=16.9$, $p<0.001$); this later modulatory effect of MI on additive scaling was more pronounced in later visual areas (e.g. V3v, hV4) compared to earlier areas (three-way interaction between attention, MI bin, and visual area: $F(9,54)=3.01$, $p<0.01$, see Fig. 6). In addition, attention led to significant multiplicative scaling in high-MI voxels (two-way interaction between attention and MI bin: $F(3,18)=7.00$, $p<0.005$, see Fig. 7); however, the three-way interaction between attention, MI bin and visual area did not approach significance ($F(9,54)=1.14$, $p=0.35$), suggesting a qualitatively similar pattern in all regions. Finally, no significant changes in bandwidth scaling was observed with attention (Supplemental Fig. 3, attention by MI interaction, collapsed across visual areas $F(3,18)=1.45$, $p=0.26$), which is perhaps not surprising given similar results in previous single unit recording studies (e.g. McAdams and Maunsell, 1999) and the notion that a decrease in bandwidth does not always benefit the quality of a population code (Series et al., 2004).

When considering high-MI voxels, which we hypothesize are the most representative of

underlying neural activity, the effect of multiplicative scaling becomes more pronounced as compared to additive scaling (~34% vs. ~21% change in respective parameter estimates with attention, averaged across all visual areas). This change in the multiplicative scaling parameter is roughly comparable to that found by a classic single-unit study of spatial attention (McAdams and Maunsell, 1999a,b) where a 23% average increase in multiplicative scaling was observed in V4 neurons.

Although we sorted the data into four equinumerous MI bins for the main analysis – so that enough voxels would be in each bin to yield reliable VTF estimates – the magnitude of additive and multiplicative scaling with attention seems to approximately increase monotonically with increasing MI (Fig. 8 and Supplemental Figs. 6, 12), which we posit is due to the fact that more homogeneous voxels (indexed by higher MI values) should more accurately reflect attention-induced scaling in underlying neuronal tuning functions (see simulations in Supplemental Fig. 2).

The observed multiplicative scaling of responses with attention should *theoretically* increase MI by increasing the dynamic range of the BOLD response (see Methods: *Relationship between attentional gain and MI*). However, it is important to directly evaluate whether or not MI actually increases with attention, because an increase in MI is not only influenced by multiplicative scaling but also by the noise characteristics of the responses. Thus, we next used MI to evaluate the amount of information conveyed by the population response profiles. First, MI was estimated in each voxel using only responses evoked by *unattended* stimuli, and then the voxels were rank-ordered by their respective value of MI. As outlined previously, voxels with high MI are postulated to be more homogenous and hence representative of underlying neural populations (see Methods: *Using MI as a criterion to select representative voxels*). When we computed the ratio between each voxel's normalized MI for attended stimuli and unattended

stimuli (see Methods: *Normalized MI*); this ratio was greater than 1 for the most representative voxels, indicating that spatial-attention increases the amount of information carried by BOLD responses (~25% increase, see Fig. 9 and Supplemental Fig. 13 for a more conservative cross-validated version of the same analysis). Supplemental Figure 9 shows where in cortex the most informative voxels were found in a set of representative subjects.

2.4 DISCUSSION

In the present study, we used fMRI and feature-selective VTFs to show that spatial attention increases the MI of population response profiles within regions of early visual cortex, primarily by inducing a multiplicative scaling of feature-selective neural tuning functions. If early visual cortex can be thought of as a representational screen on which external visual stimuli get projected and encoded by millions of tiny sensors, then the implication of our finding is that attention effectively increases the bit-depth of the information encoded about the stimulus. This increase in bit-depth with attention in early visual areas should enable later areas that integrate this information to operate more efficiently, thereby supporting faster and more accurate perceptual decisions (Ditterich et al., 2003; Jazayeri and Movshon, 2006; Newsome et al., 1989; Shadlen and Newsome, 2001).

MI fulfills two main roles in current study: A) MI is used to identify the most informative voxels, which we posit are also the most homogenous and hence representative of underlying neural populations. B) MI provides a means of evaluating how cognitive manipulations, such as spatial deployments of attention, can systematically influence the amount of information encoded about behaviorally relevant stimuli. Since we use MI for the dual purpose of selecting highly representative voxels as well as for observing changes in the information content of these voxels with attention, we were careful to avoid non-independence errors. This is especially important in the present context because MI is positively correlated with the multiplicative scaling parameter

of a VTF. To avoid issues of circularity, we used independent selection and evaluation phases to first select high MI voxels and then to evaluate the influence of attention on their response profiles. The MI score of a voxel was calculated using data from all scans save one (selection phase) and the VTFs were computed using data from the remaining scan (evaluation phase), and this process was iterated through all unique combinations of holding-one-scan-out. Similarly, for the analysis of the change in MI with attention (Fig. 9), we rank-ordered voxels based on MI derived from unattended stimuli only, and then computed the relative increase in MI due to attention (a further split-half analysis was done as well, see Supplemental Fig. 13). The use of a cross-validation procedure ensures that task-independent machine noise was not responsible for the observed attentional effects. Given these measures and the fact that machine noise cannot systematically and significantly vary between attended and unattended perceptual states, we conclude that the data presented here reflect attentional modulations of BOLD signals and are not due to any selection bias or non-independence errors in our methodology.

The observation of multiplicative scaling in the most informative voxels is consistent with previous single-unit studies in primates (McAdams and Maunsell, 1999). On the other hand, the extremely large additive scaling factors we report (see Fig.6) are not typically observed at the single-unit level, where such effects are generally modest. However, our data and simulations suggest a resolution to this apparent discrepancy. Consider a hypothetical voxel that contains a heterogeneous distribution of orientation-selective neurons. Such a voxel will have poor selectivity and should produce a relatively flat tuning function. Here, a purely multiplicative scaling at the single unit level would manifest as a feature non-specific (additive) increase in BOLD activation. For example, attending to 45° should primarily boost the firing rate of neurons tuned to 45°, which will produce heightened activation in the voxel. However, because the voxel contains an equal number of neurons that prefer every other orientation, attending to a different orientation (e.g. 90°, 135°, etc.) should produce an equivalent increase in the activation level.

Thus, when measured at the coarse spatial resolution afforded by fMRI, multiplicative scaling in a VTF will likely be underestimated and, more importantly perhaps, translated into a response that appears additive. The magnitude of this 'smearing' should increase as the distribution of neural tuning preferences within a voxel becomes more uniform (see simulation results, Supplemental Fig. 2). Note that we are not disputing the existence of additive scaling per se, but simply suggesting that estimates based on fMRI might be generally skewed in favor of additive over multiplicative scaling. In contrast, using an information-theoretic methodology allowed us to distill a subset of voxels that likely captured changes in a more homogeneous distribution of orientation-selective neurons, thereby providing a better window into the nature of attentional modulations in neuronal populations. Accordingly, the more informative voxels showed clear signs of multiplicative scaling, in line with most single-unit studies. Therefore, we stress that choosing the most representative voxels is key to improving the precision of inferences based on fMRI data, and MI provides such a criterion for voxel selection.

The role of multiplicative scaling in increasing the precision of population codes is relatively straightforward: it increases the mean dynamic range of responses to the feature set, thereby increasing the information content of neural representation (Butts and Goldman, 2006; Doya et al., 2007). However, the role of additive scaling is not immediately apparent. From an information-theoretic perspective, a uniform translation of a tuning function (without any other accompanying changes in signal characteristics) should not affect the coding precision of that tuning function (Cover and Thomas, 1991). Neuronal noise in visual cortex is typically thought to be near-Poisson in nature (Shadlen and Newsome, 1994; Averbeck et al., 2006). This implies that if there was a feature non-specific increase in signal (additive scaling), noise would not increase proportionately ($\Delta\text{Noise} \propto \Delta\sqrt{\text{Signal}}$), leading to an increase in SNR of the individual responses (Mitchell et al., 2007). However, there is a disconnect when analyzing SNR of single responses

and the MI of population response profiles. MI of a tuning curve, be it neural or voxel-based, not only depends on the noise characteristics of single responses (noise entropy) but also on the mean dynamic range of responses (total entropy) (See Methods). Thus a pure additive scaling that increases the SNR of single-unit responses would ultimately increase the noise entropy without increasing the mean dynamic range (yielding no net improvement in information content). On the other hand, multiplicative scaling of tuning curves would simultaneously increase the mean dynamic range of responses as well as increase the noise in responses, the benefit of the former outweighing the detriment of the latter.

When comparing figures 8 and 9, the observed increase in MI with attention within high MI voxels seems almost equally well-correlated with additive scaling as it is with multiplicative scaling. However, it is unlikely that additive scaling caused the observed attention-related increase in MI on theoretical grounds (see above), and because there was no attention-related increase in the information content of the lowest MI voxels, even though these same voxels showed significant additive scaling (compare Figs. 8 and 9 as well as supplemental figs. 12 and 13). Thus, additive scaling alone is not sufficient to induce a relative increase in MI with attention. Instead, the ratio of attended-to-unattended MI does not go above 1 until multiplicative gain becomes evident in the higher-MI voxels (towards the left-hand side of Figure 9). Therefore, the observed increase in MI is more likely related to multiplicative scaling as opposed to additive scaling, and our data do not appear to shed light on the functional role of additive scaling in influencing the information content of population codes per se.

MI possesses the property that it is directly proportional to the amplitude, or the multiplicative scaling parameter, of the VTFs (or any Gaussian tuning function). However, even though we observed a multiplicative scaling that accompanied an increase in MI with attention, we cannot conclude that the increase in MI was entirely due to the multiplicative scaling of the

tuning function. First, the noise entropy component $H(B|\Theta)$ of MI can also independently influence the overall value of MI, so it is possible that part of the increase in MI is related to an attention-related reduction in trial-by-trial noise within single units, as well as an attention-related decorrelation across neurons on a within trial basis (Mitchell et al., 2007; Mitchell et al., 2009; Cohen and Maunsell, 2009). Since we cannot make precise measurements of neural noise using our methodology, we cannot separate the proportion of MI attributable to an increase in total entropy with multiplicative scaling and the proportion attributable to changes in noise entropy. Second, Connor et al. (1996, 1997) has shown that spatial attention can shift the spatial receptive field of neurons towards a relevant stimulus, effectively increasing the number of responsive cells (Womelsdorf et al., 2006, 2008; Hamker et al., 2008). Given certain conditions – such as when opposite tuned neurons are correlated – an increase in the number of responsive neurons can increase the quality of encoding (Abbott & Dayan, 1999; Shamir & Sompolinsky, 2006). Thus, although multiplicative gain is one factor that should increase MI, noise decorrelation and shifting of spatial receptive fields may play an important role as well.

A source of continuing controversy regards the nature of attentional modulation of contrast response functions (CRFs, e.g. Reynolds and Heeger, 2009). Based on our results, our assertion that spatial attention induces both multiplicative scaling and additive scaling seems to be in direct conflict with previous studies that found only additive scaling of the CRF (Buracas and Boynton, 2007). One might also interpret multiplicative scaling in VTFs to be evidence that spatial attention induces response gain or activity gain on the CRF (Williford and Maunsell, 2006). However, the present experiment was conducted at a single contrast level, and even though we found a significant multiplicative scaling, we cannot determine if the *magnitude* of this multiplicative scaling remains constant across all contrast levels. If it remains constant or increases, then a response gain of the CRF would be expected; otherwise, either an additive gain,

activity gain or a contrast gain of the CRF might be induced, depending on the exact nature of the relationship between multiplicative scaling and contrast (and other factors as well, such as the relationship between stimulus size and the size of neural receptive fields; see Reynolds and Heeger, 2009). Thus, our results do not directly address this issue and further experiments will be required to specify the exact relationship between attentional gain at the level of VTFs and attentional modulations in CRFs.

General conclusions: Here, we provide evidence that spatial attention improves the information content of population responses by inducing multiplicative scaling of population response profiles, which was hitherto un-documented in human visual cortex. This multiplicative scaling should theoretically improve the encoding precision of population codes that represent the features of a relevant stimulus, and our data provide empirical support for this role of spatial attention in perception.

Chapter 2, in full, is a reprint of the material as it appears in "Spatial attention improves the quality of population codes in human visual cortex", Saproo S., Serences J., Journal of Neurophysiology, 2010 Aug;104(2):885-95. The dissertation author was the primary investigator and author of this paper. I thank John Serences for his contribution. I thank Ymkje Anna De Vries, Geoff Boynton, Craig McKenzie and Ed Awh for helpful discussions towards this material.

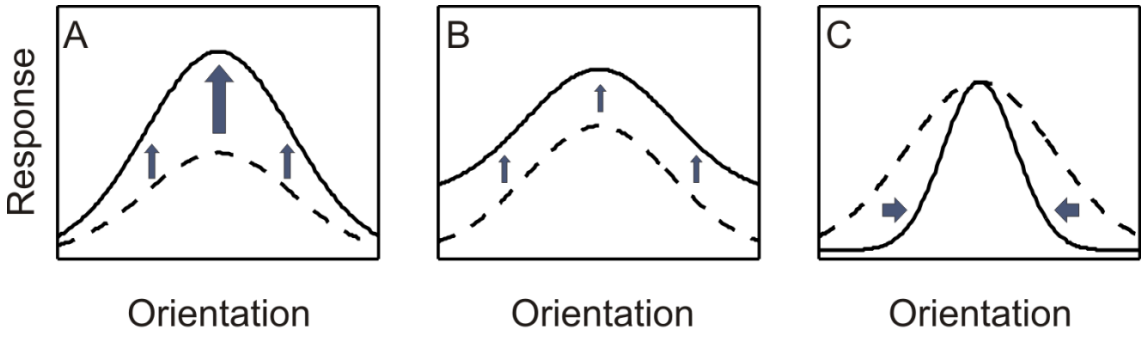


Figure 2.1. Schematic depiction of the possible types of attention-induced scaling in the neuronal tuning function and consequently in population response profiles; (A) Multiplicative scaling, or feature-dependent increase in response amplitude where the increase depends on the proximity of the attended stimulus feature to the preferred feature of the neuron, (B) Additive scaling, or feature non-specific increase in response amplitudes, and (C) Bandwidth scaling, or change in standard deviation of response profile.

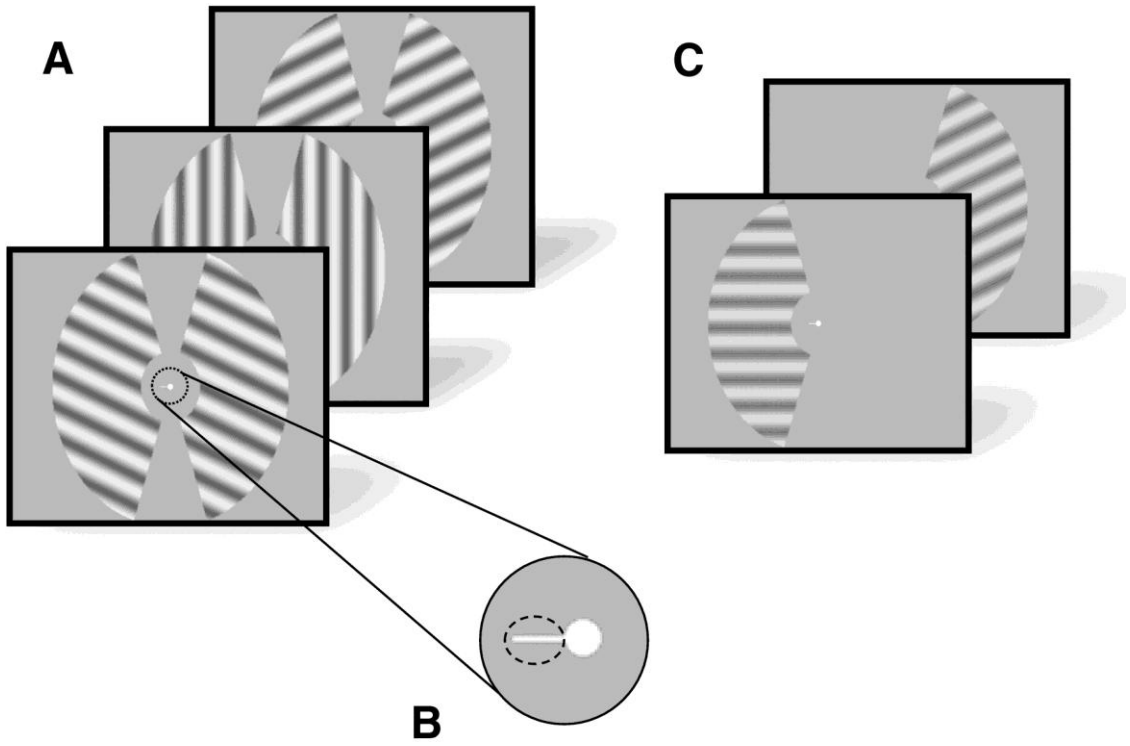


Figure 2.2. (A) Schematic of the task performed by subjects during fMRI scanning. Subjects attended to a flickering (2Hz) sinusoidal grating that was rendered in one of eight possible orientations [0°, 22.5°, 45°,…, 157.5°]. Subjects were required to continuously fixate on the spot at the center of the display, and to attend to either the left or the right stimulus based on a small central cue (figure panel B). Subjects pressed a button whenever they detected a slight dimming of the attended stimulus and ignored an equally probable dimming of the unattended stimulus. Target contrast decrements were titrated to maintain detection accuracy at approximately 75-80%. Stimuli in both hemi-fields were rendered in the same orientation to negate the influence of global feature-based attentional modulations (see Methods). (C) Independent localizer scans were used to identify the most spatially selective voxels in V1-hV4. The stimulus was similar to that used in the attention task (above) except that it was visible in only one hemi-field, to which the subject had to direct his/her attention.

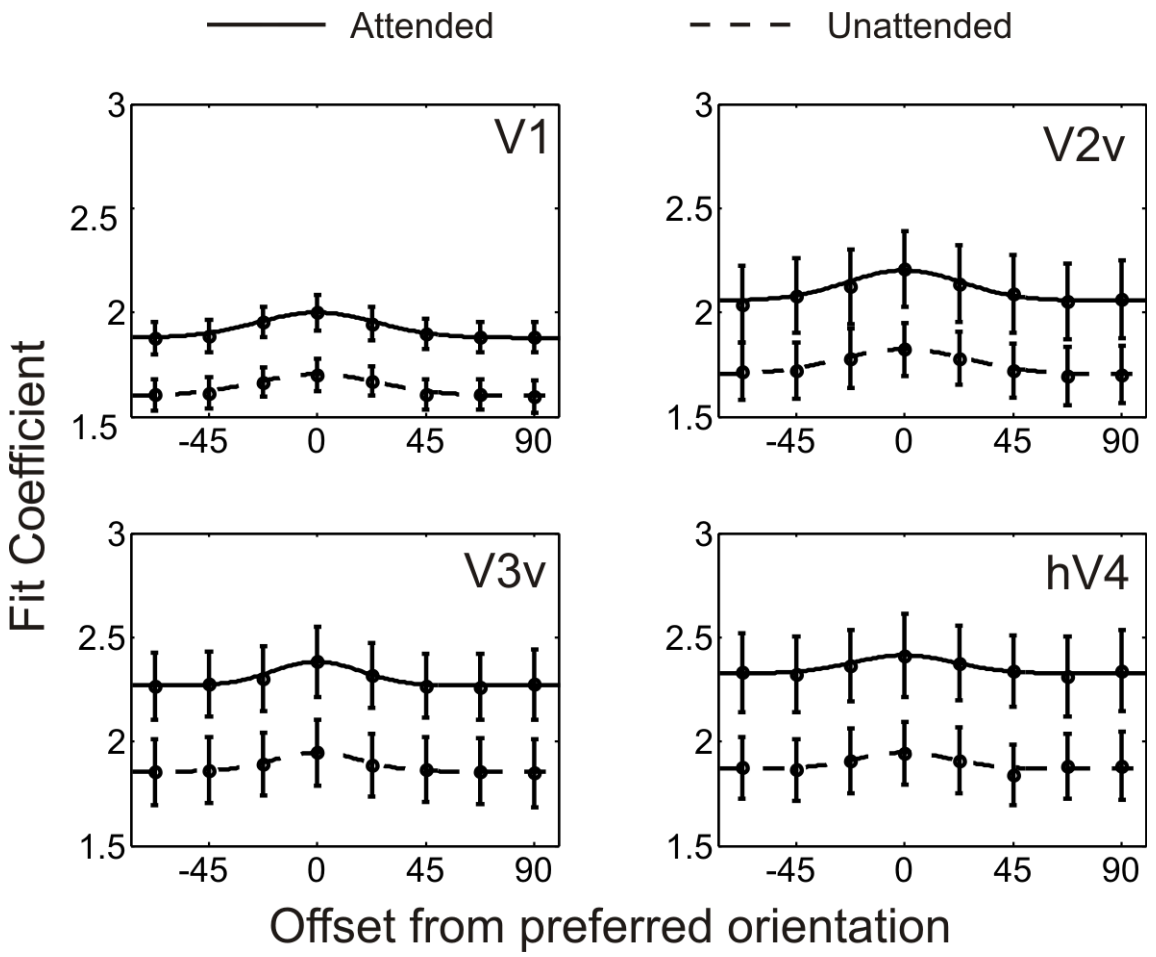


Figure 2.3. Voxel tuning functions (VTFs) with attention (solid curve) and without attention (dashed curve) based on responses in V1, V2v, V3v and hV4. These mean tuning functions were produced by centering all VTFs (for a visual area) at their preferred orientation and then averaging across subjects (lines represent best fitting circular Gaussian, see Methods). Error bars reflect ±1 S.E.M. across subjects.

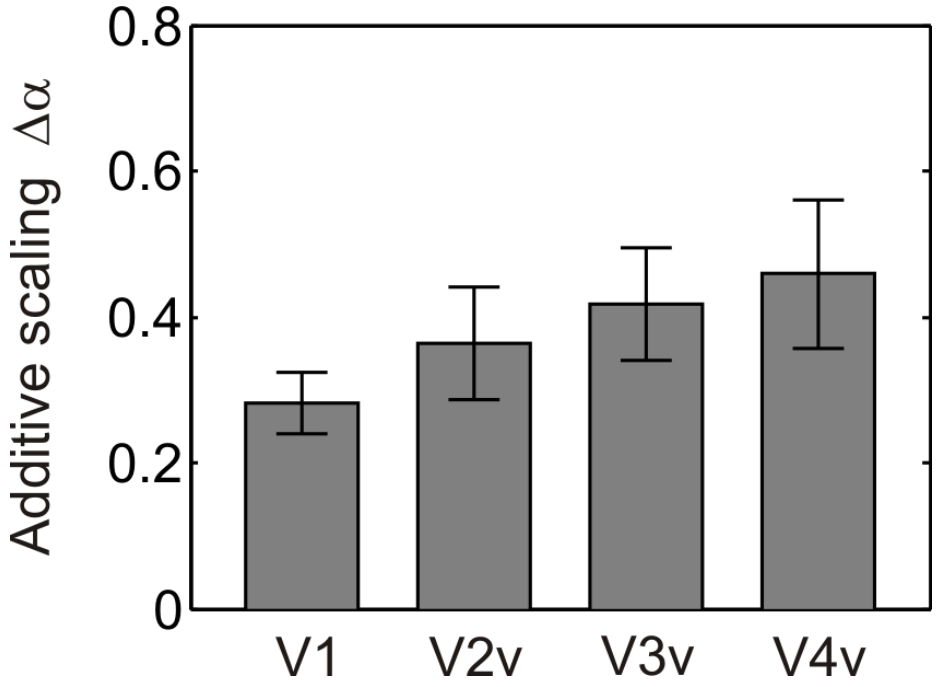


Figure 2.4. Additive scaling parameter of best fitting Gaussian function in areas (V1-hV4). Error bars reflect ±1 S.E.M. across subjects.

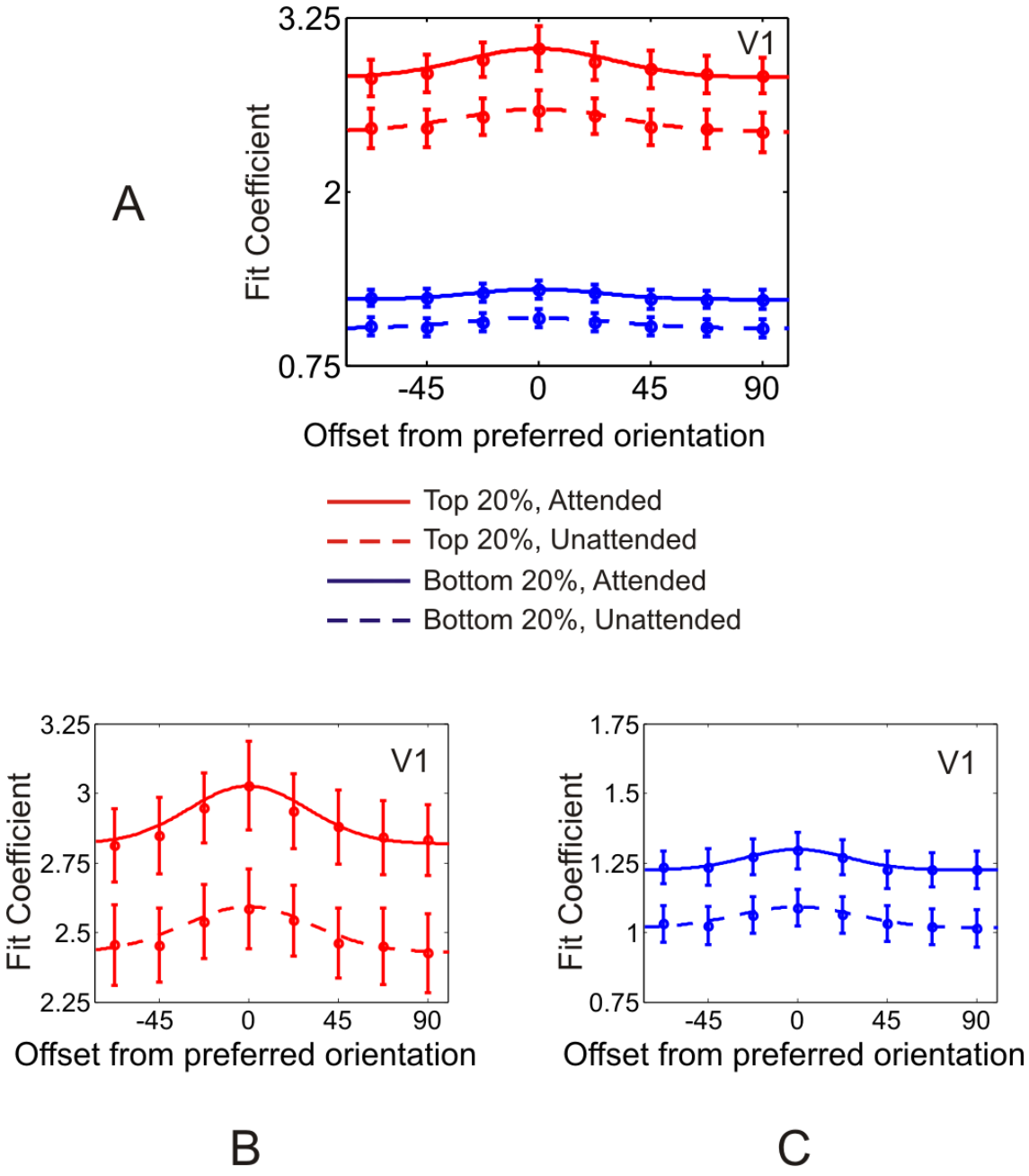


Figure 2.5. (**A**) Tuning functions for top 25% (red) and bottom 25% (blue) of voxels in V1 ranked by MI score, with attention (solid curves) and without attention (dotted curves). Tuning functions of high MI voxels (red lines, see also panel **B**) show significant multiplicative and additive scaling, whereas low MI voxels show primarily additive scaling (blue lines, see also panel **C** and Figures 6 and 7). Y-axis refers to the magnitude of the fit coefficients (beta weights) estimated using the GLM (see Methods), and note that panels **B** and **C** have been derived from panel A to better highlight the difference in shape between the VTFs from high MI and low MI voxels (with y-axis of both panels B and C covering equal range to permit a direct comparison). Error bars reflect ±1 S.E.M. across subjects.

Figure 2.6. Additive scaling with attention for each MI quartile in V1, V2v, V3v, and hV4. Left panels show additive scaling parameter derived separately from attended data and unattended data. Right panels show the net change (attended *minus* unattended) in the parameter with attention. Top row shows data averaged across all visual areas, remaining rows show data from each visual area. Error bars reflect ±1 S.E.M. across subjects.

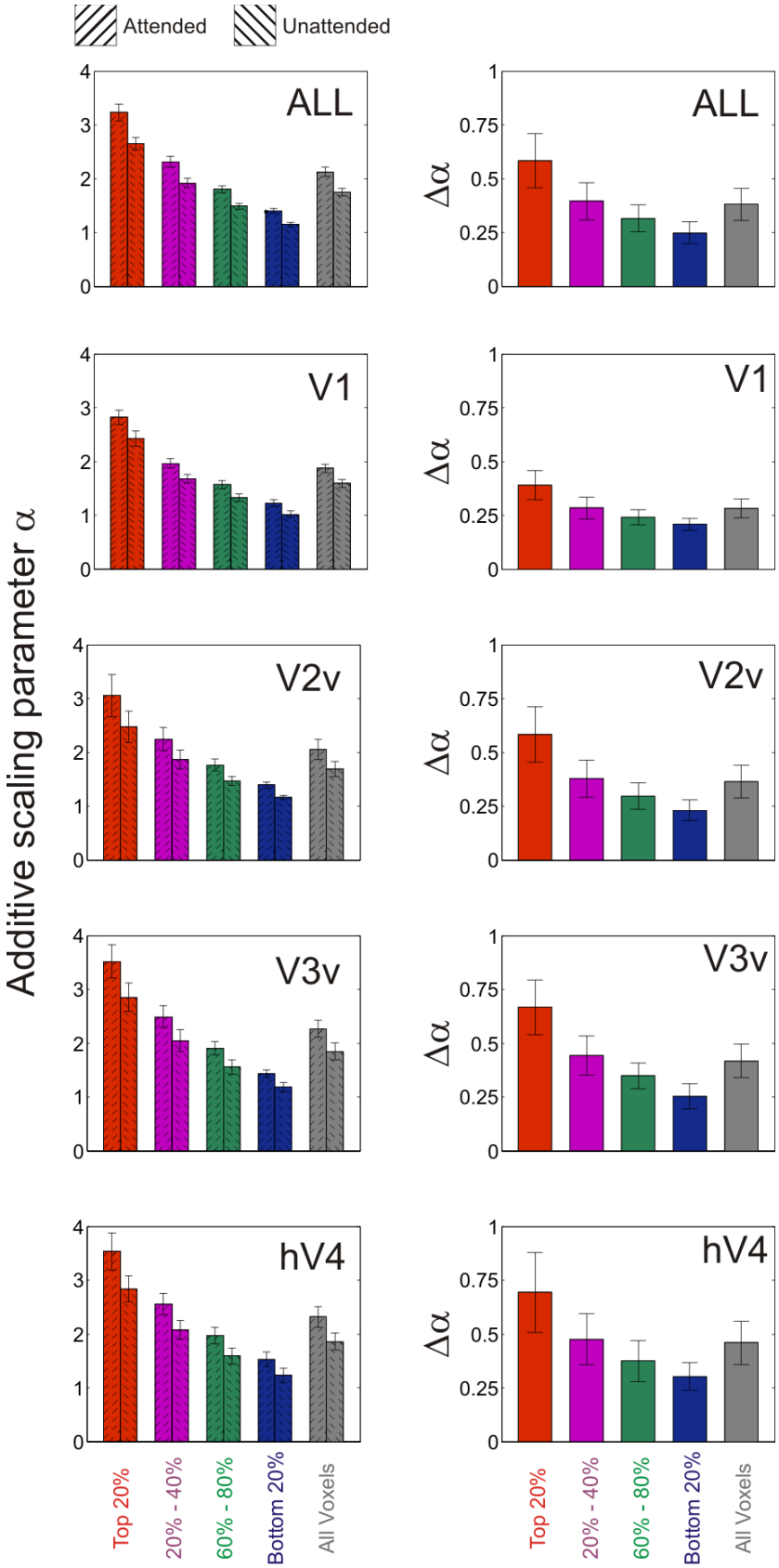

Attended
Unattended
ALL
V1
V2v
V3v
hV4
Additive scaling parameter α
Δα
Top 20%
20% - 40%
60% - 80%
Bottom 20%
All Voxels

Figure 2.7. Multiplicative scaling with attention for each MI quartile in V1, V2v, V3v, and hV4. Left panels show multiplicative scaling parameter derived separately from attended data and unattended data. Right panels show the net change (attended *minus* unattended) in the parameter with attention. Top row shows data averaged across all visual areas, remaining rows show data from each visual area. Error bars reflect ±1 S.E.M. across subjects.

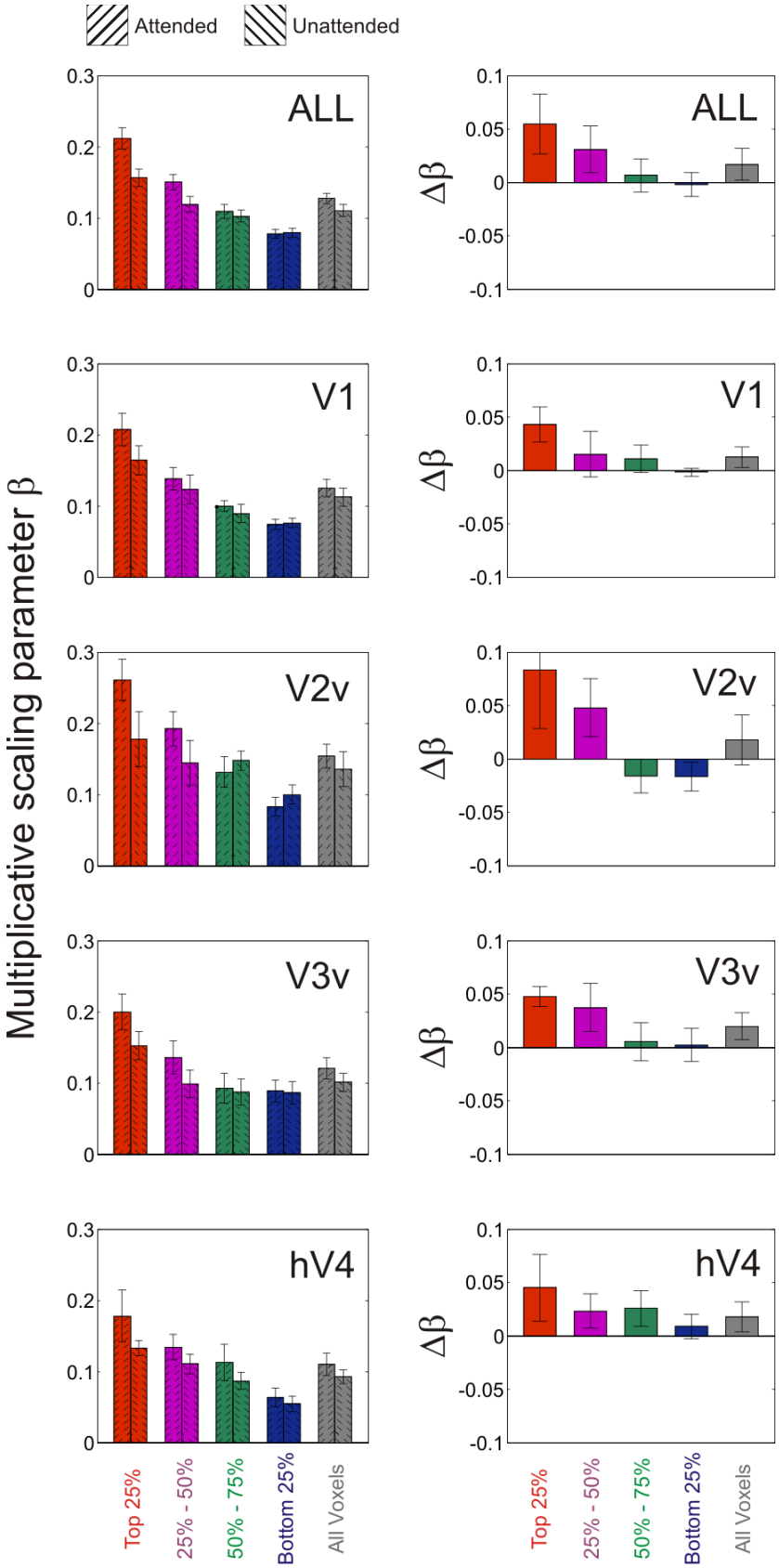
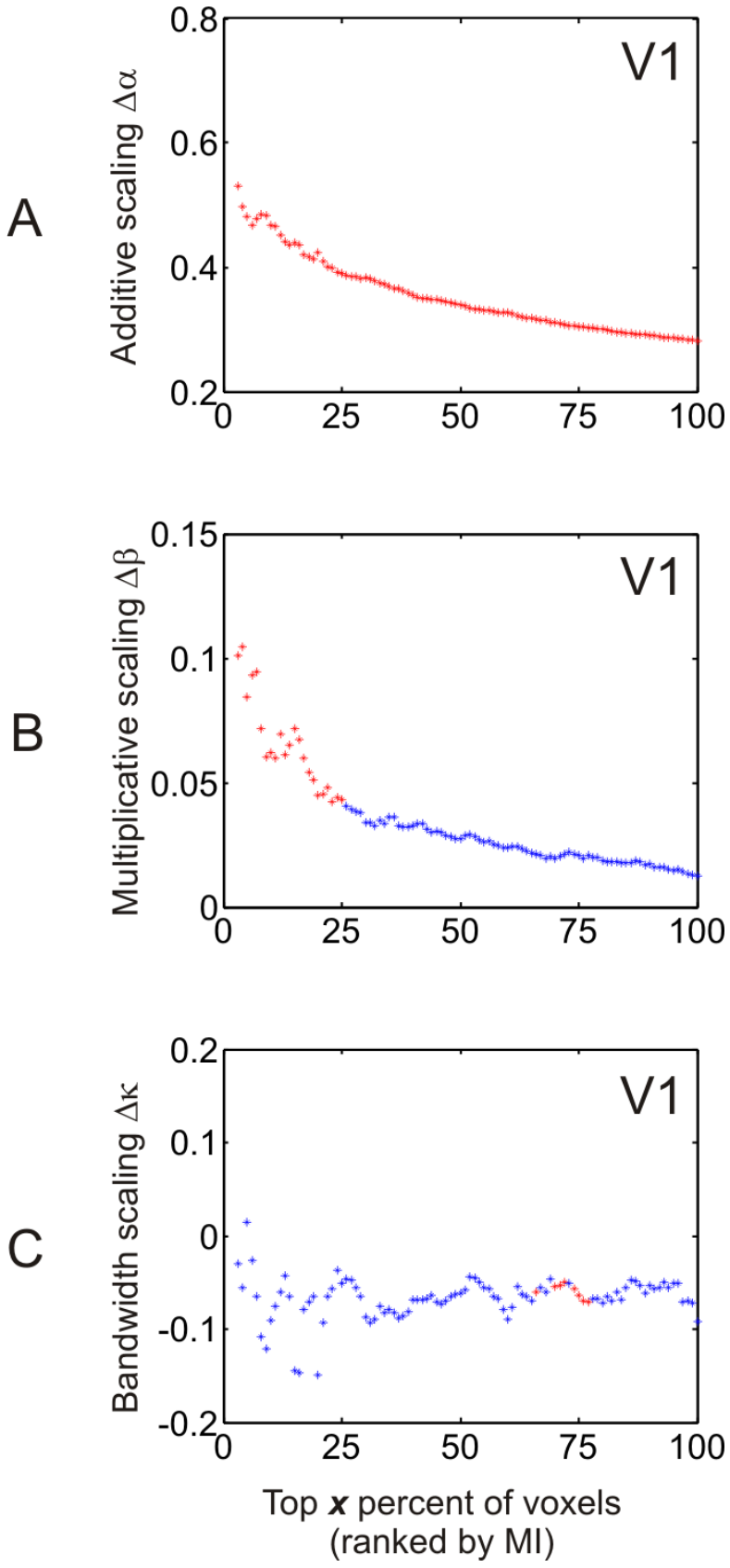

Attended
Unattended
Multiplicative scaling parameter β
ALL
V1
V2v
V3v
hV4
Δβ
0.3
0.2
0.1
0
0.1
0.05
0
-0.05
-0.1
Top 25%
25% - 50%
50% - 75%
Bottom 25%
All Voxels

Figure 2.8. (A) Additive scaling, (B) Multiplicative scaling, and (C) Bandwidth scaling (attended *minus* unattended) for voxels in V1 ranked by their MI score. Each point on x-axis depicts an aggregation of top x% of voxels ranked by their MI score. Corresponding y-axis depicts the mean additive scaling, multiplicative scaling and bandwidth scaling with attention for mean voxel tuning function of that group. Red color indicates data points that reached significance by repeated measures t-test ($p<0.05$). Data points for which mean RMSE of fit (across subjects) deviated more than 2 standard deviations from overall RMSE mean (across all subjects and datapoints), were excluded; this was only an issue for the smallest aggregation of high Mutual Information voxels (< top 3%).


A
B
C
V1
V1
V1
Additive scaling Δα
Multiplicative scaling Δβ
Bandwidth scaling Δκ
0.8
0.6
0.4
0.2
0.15
0.1
0.05
0
-0.1
-0.2
0
25
50
75
100
Top x percent of voxels
(ranked by MI)

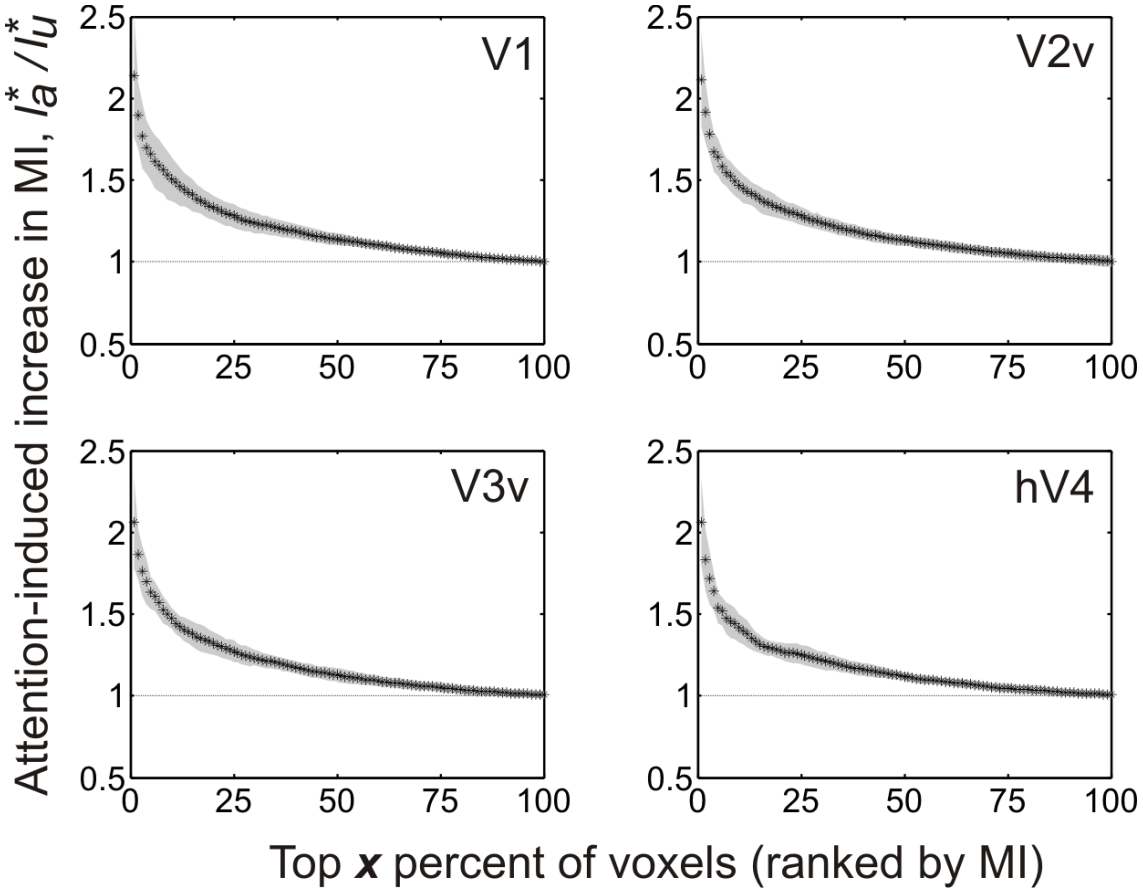
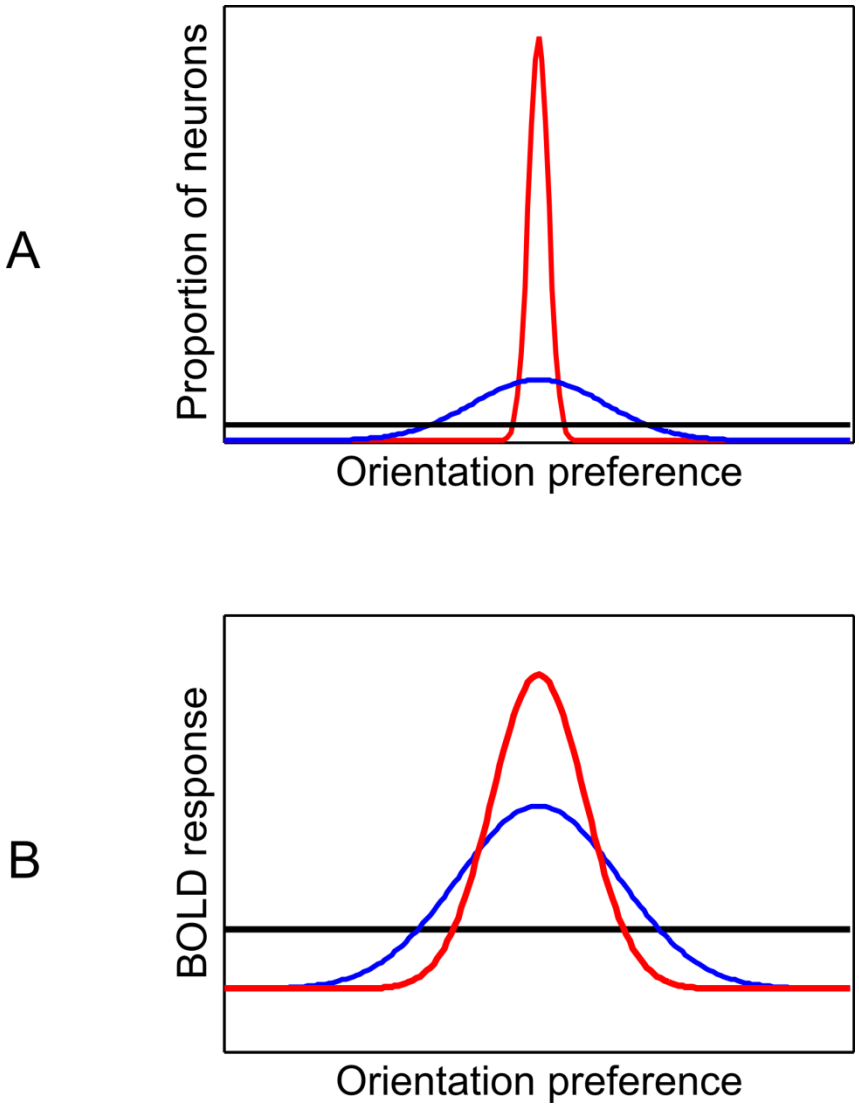


Figure 2.9. Ratio of normalized MI (see Methods) for voxels in V1, V2v, V3v and hV4 for attended and unattended stimuli. Voxels were sorted by normalized MI based on unattended responses only. The abscissa marks different groups of top voxels (by percentage of the total voxels ranked by their MI score). /* refers to normalized MI (see Methods) derived from unattended responses and attended responses, respectively. Ordinate refers to their ratio. Shaded gray patch around data points highlights the standard deviation of values between subjects. High MI voxels showed higher increase in overall information content with attention.

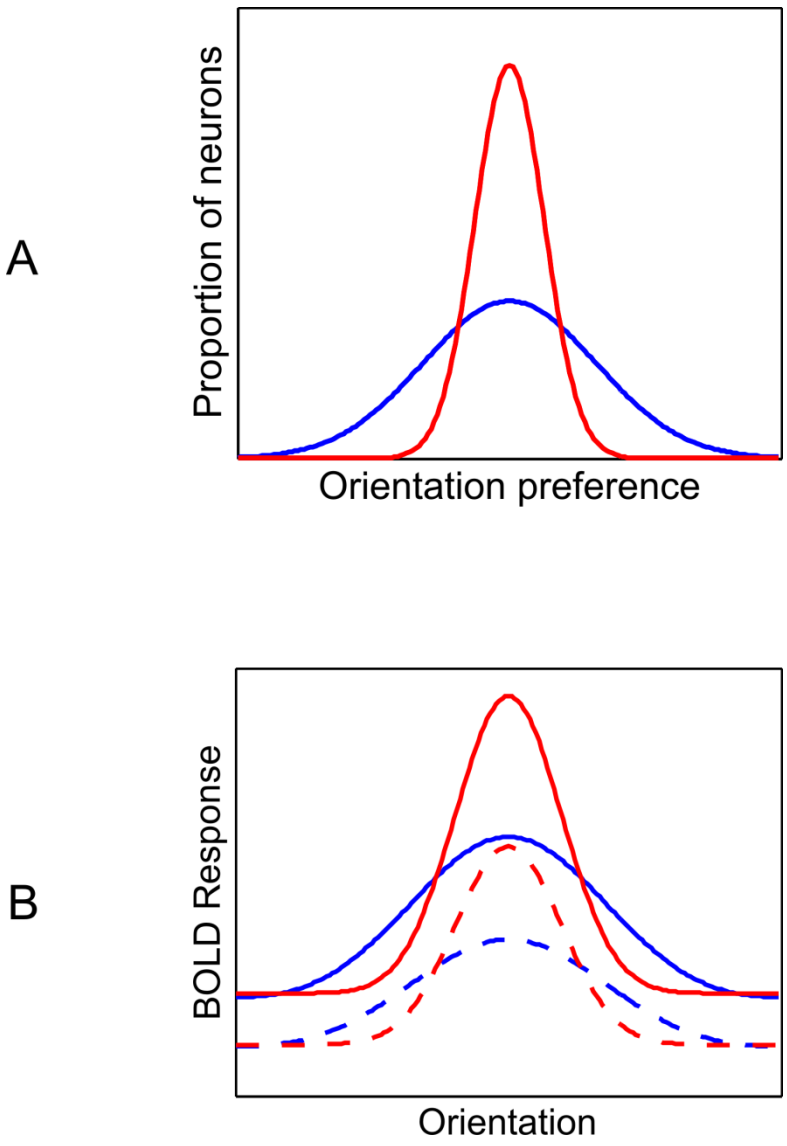
Table 2.1 Average size across subjects for each visual area

Visual area	Mean Voxels	Std Deviation
V1	551.43	163.68
V2v	251.79	105.12
V3v	233.29	81.93
hV4	237.57	113.25

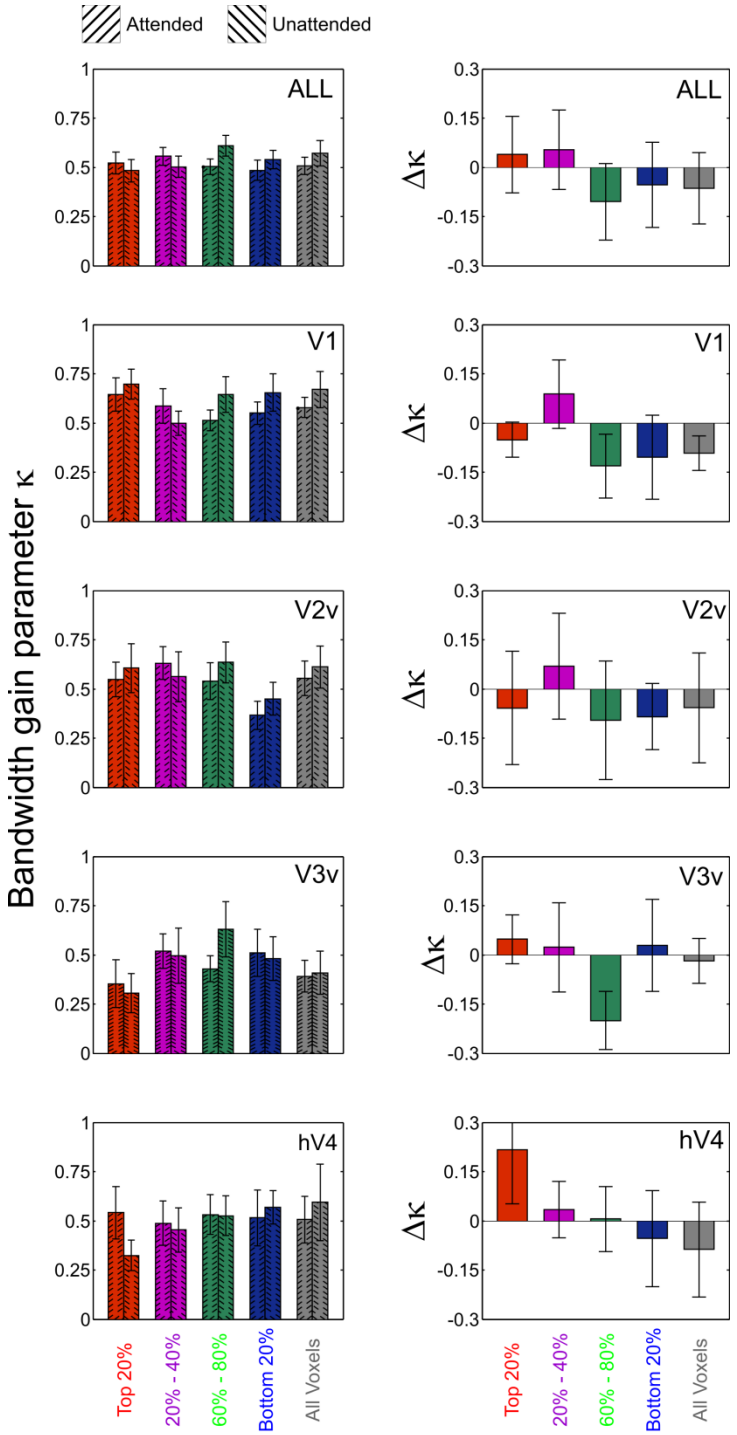
2.5 SUPPLEMENTAL MATERIAL



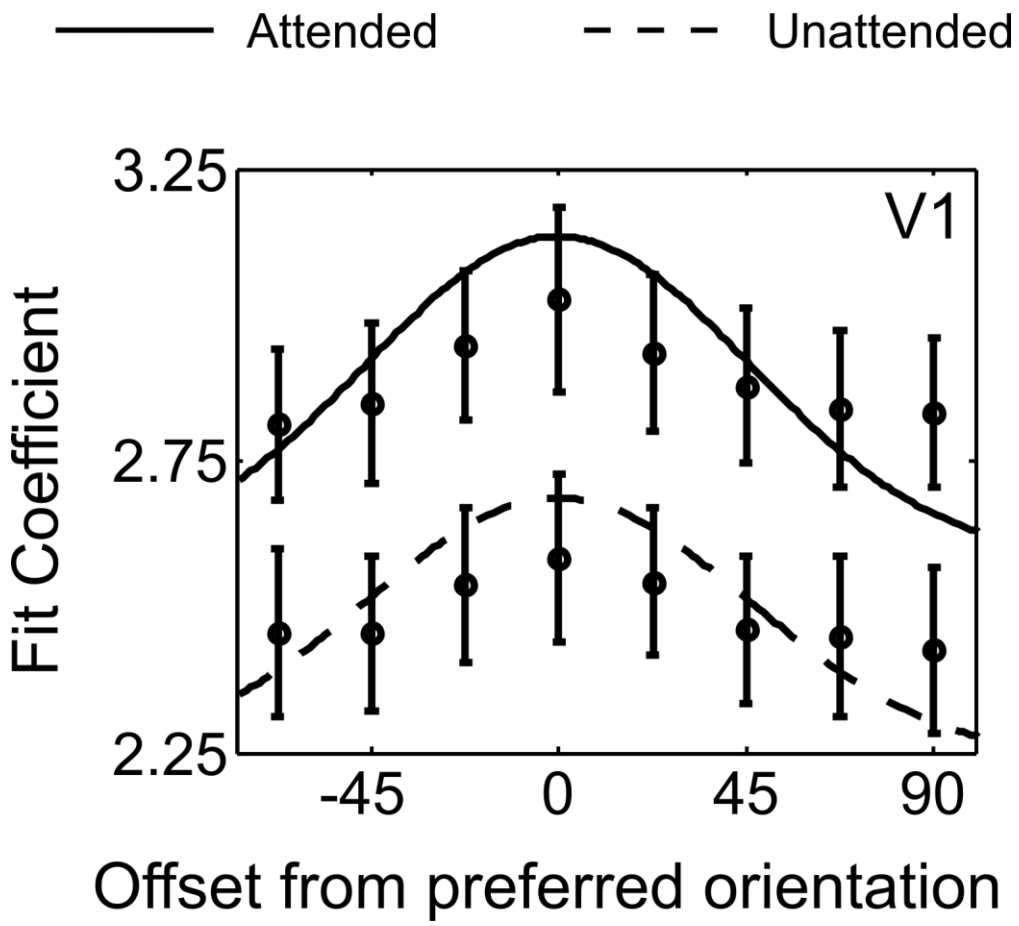
Supplemental Figure 2.1 Results of a simulation showing how the distribution of feature preferences in the neural population comprising a voxel affects the expected shape of the corresponding voxel tuning function (VTF). (A) The distribution of orientation preferences in 3 synthetic voxels (black = uniform distribution, blue = slightly biased distribution, and red = highly biased distribution). (B) Assuming for the sake of simplicity that neural activity is linearly related to the BOLD response, then the black, blue, and red lines depict the corresponding VTFs predicted for each distribution shown in panel A. While the precise shape of the VTFs predicted for each voxel will depend on the exact function that maps neural activity and the BOLD response, the qualitative relationship between the lines should hold even under nonlinear (but monotonic) mappings.



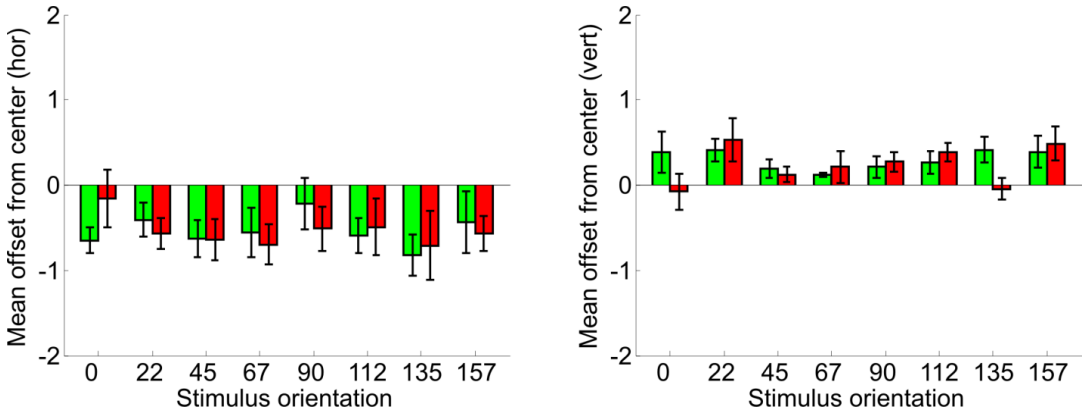
Supplemental Figure 2.2 Results of a simulation showing how the distribution of feature preferences in the neural population comprising a voxel affects the expression of attentional scaling in the corresponding VTF. (A) The distribution of orientation preferences in 2 synthetic voxels (blue = slightly biased distribution, and red = highly biased distribution). (B) Dashed blue and red lines depict VTFs for the two synthetic voxels, in the absence of attention. Solid blue and red lines depict VTFs for the same voxels under attention. Model neurons comprising the two synthetic voxels shared the same tuning characteristics when simulating the condition of inattention. Subsequently, the neurons comprising the two synthetic voxels received the same amount of additive scaling (20:1) and multiplicative scaling (1.5:1), simulating the hypothetical effects of attention. The resulting multiplicative scaling in the red VTF was larger than that in the blue VTF. However, additive scaling was marginally higher in the blue VTF because the relatively uniform distribution of neural tuning preferences within the synthetic voxel caused some of the multiplicative scaling to be translated into additive scaling (confirmed by fitting the VTF model described in Methods to the simulated data points).



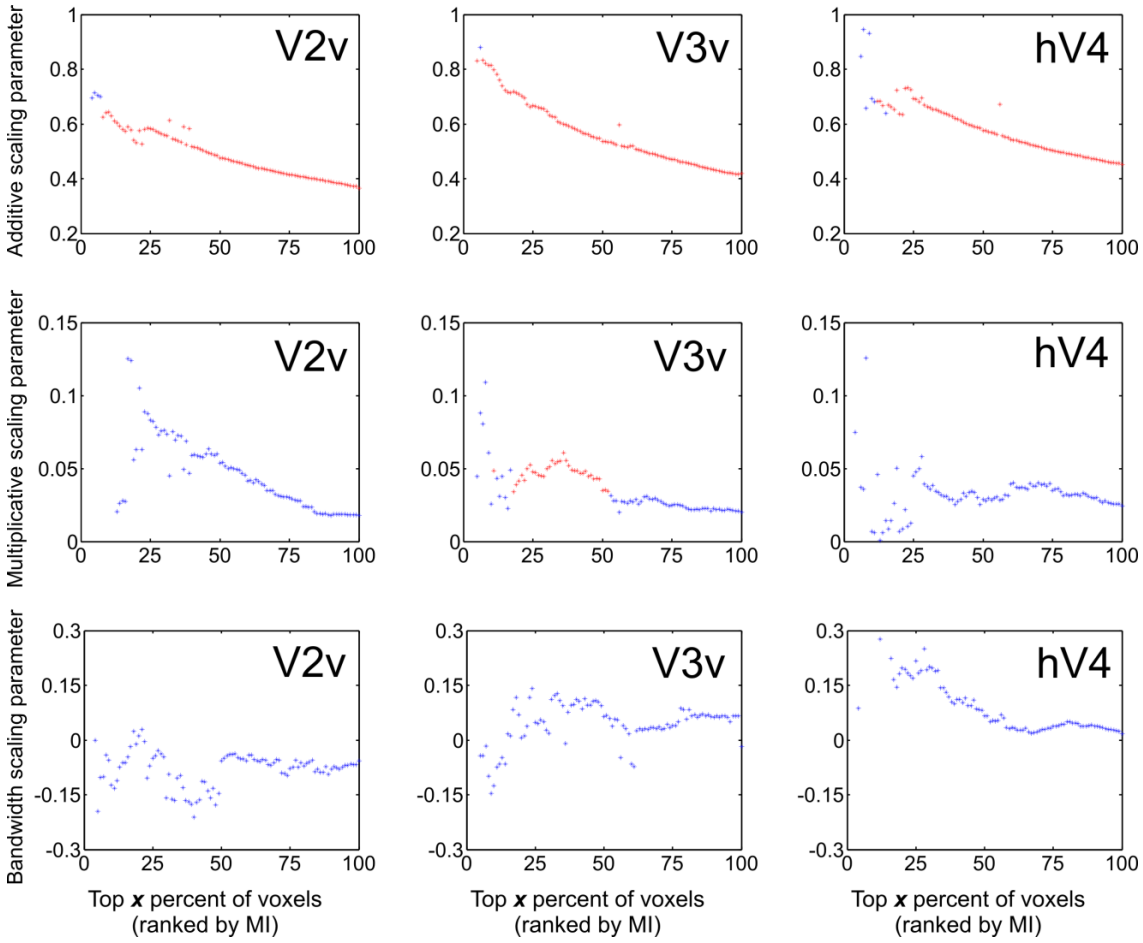
Supplemental Figure 2.3 Effects of attention on bandwidth scaling (or concentration parameter, κ) of best fitting circular Gaussians (attended *minus* unattended) for each MI quartile in V1, V2v, V3v and hV4. No significant changes were observed. Top row shows data averaged across all visual areas, remaining rows show data from each visual area. Error bars reflect ±1 S.E.M. across subjects.



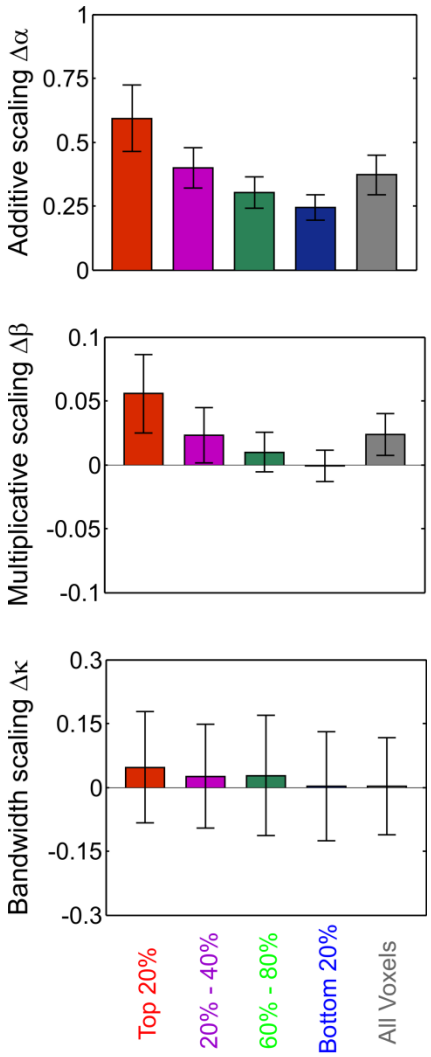
Supplemental Figure 2.4 Voxel tuning function with attention (solid curve) and without attention (dashed curve) based on responses in V1 when fit using activity gain model (see Methods). These mean tuning functions were produced by centering all VTFs (for a visual area) at their preferred orientation and then averaging across subjects (lines represent best fit of activity gain model). As can be seen, the model fit the data poorly when fit parameter values were constrained to be positive (i.e. non-negative additive scaling parameter values were disallowed because negative activation levels are not plausible in the present context). Because of the relatively poor fit, the model was dropped from further consideration. Error bars reflect ±1 S.E.M. across subjects.



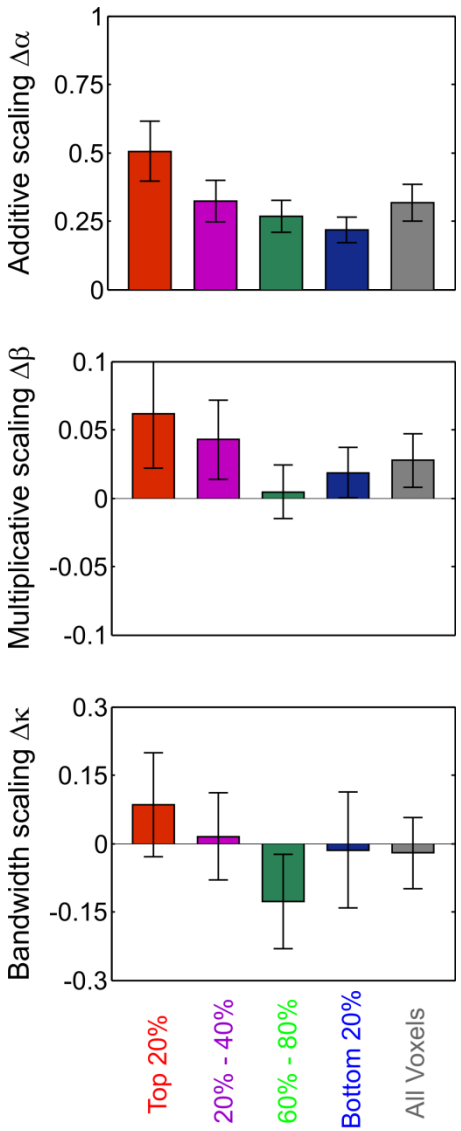
Supplemental Figure 2.5 Mean X and Y position of saccade endpoints (in degrees of visual angle from the center fixation spot, see Fig. 2) binned based on the orientation of the stimuli: green bars indicate mean eye position on trials where the left stimulus was attended, and red bars indicate mean eye position when the right stimulus was attended. No systematic differences in eye position were observed as a function of the attended location (left vs. right).



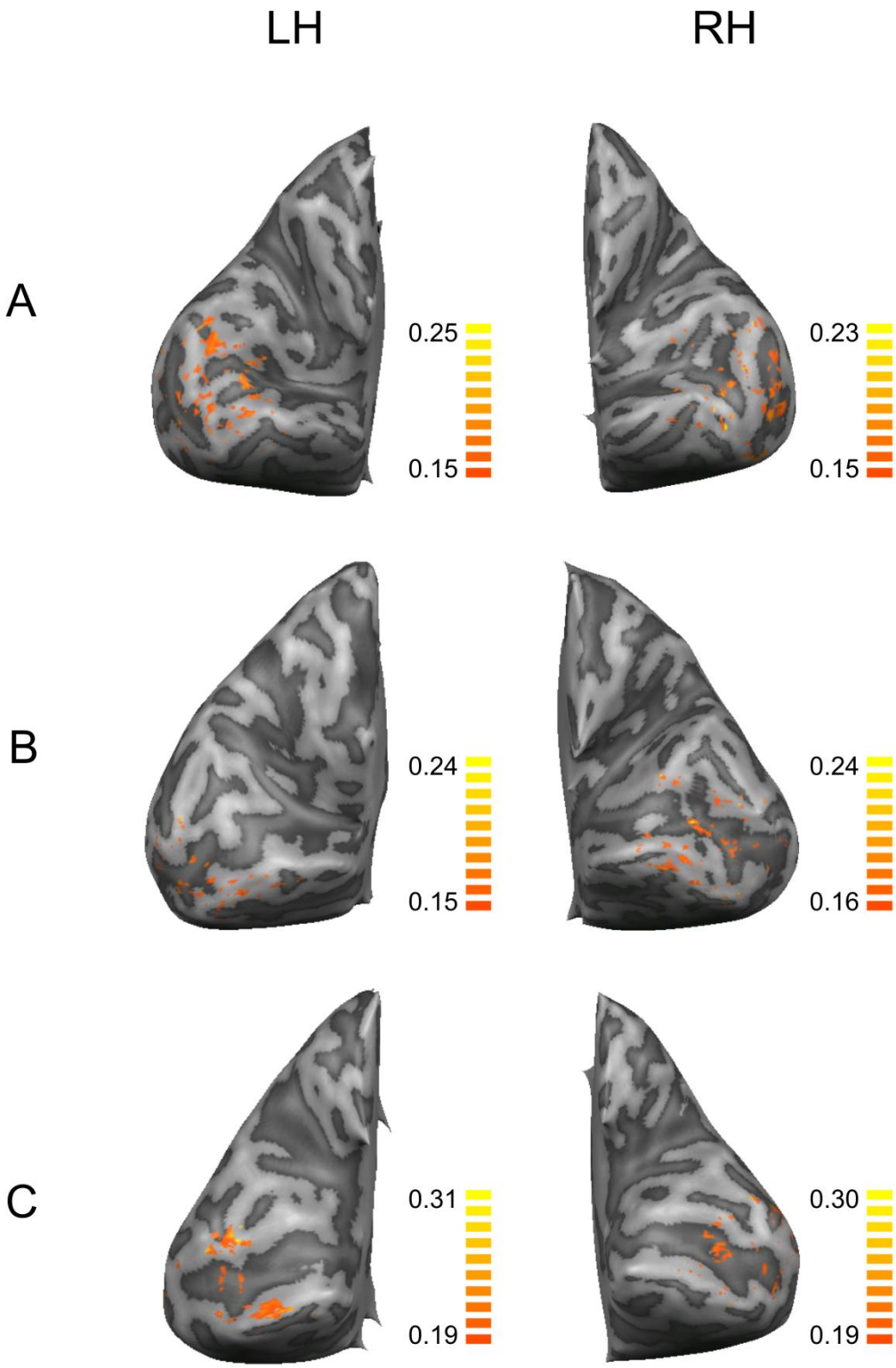
Supplemental Figure 2.6 Additive scaling, Multiplicative scaling, and Bandwidth scaling (attended *minus* unattended) for voxels in V2v, V3v and hV4 ranked by their MI score. Red datapoints indicate significance by repeated measures t-test ($p<0.05$). Multiplicative scaling sometimes could not be plotted for groups of high MI voxels because aggregation across a small number of voxels made the parameter estimates for the associated voxel tuning functions very unreliable, causing the model to fail to converge or to have an unusually high RMSE (see Materials and Methods). Also, there were approximately ½ as many voxels in areas V2v, V3v, and hV4 as in area V1, so the scatterplots shown here are inherently less stable than those based on data from V1 (Figure 8).



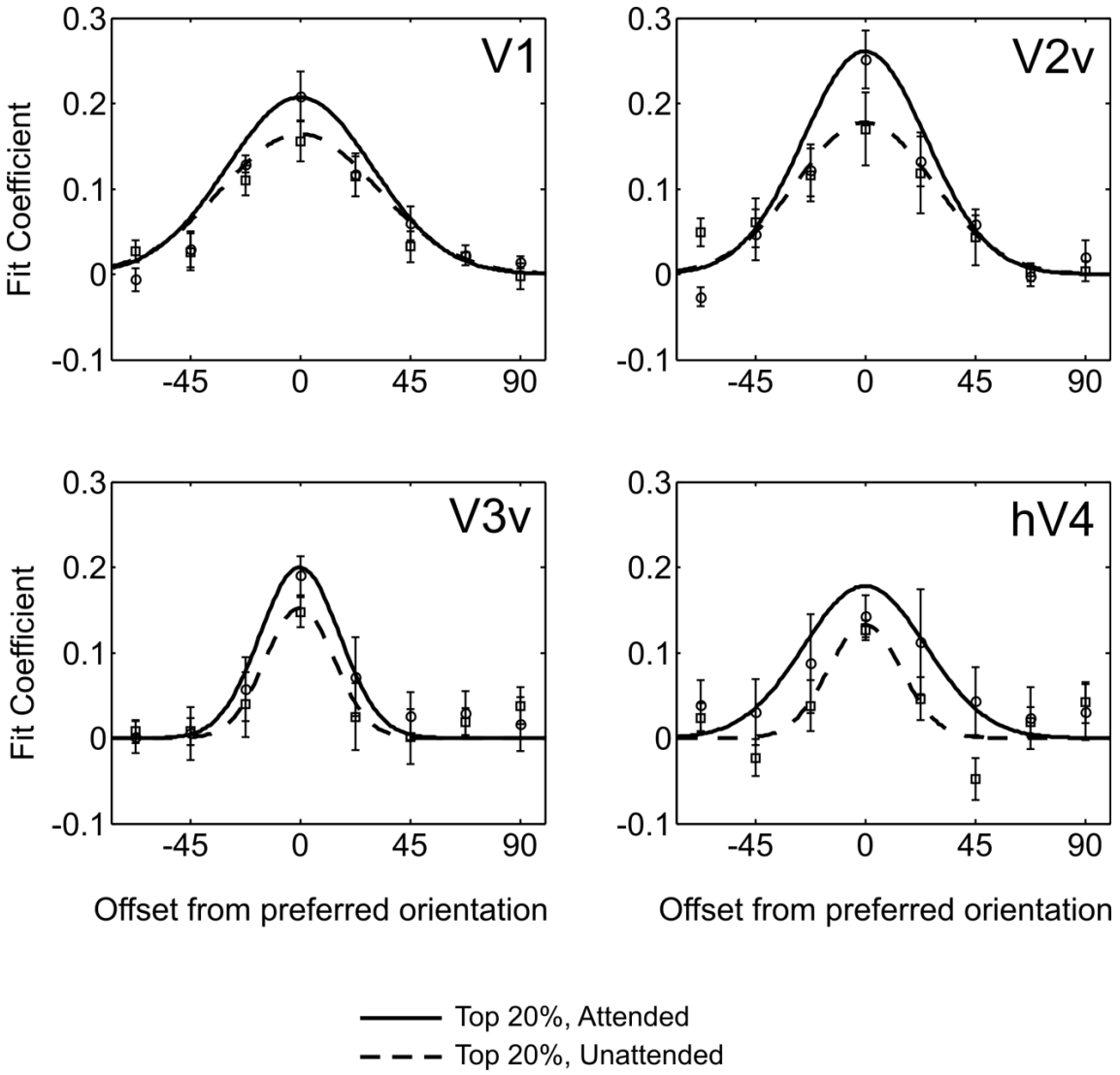
Supplemental Figure 2.7 In the main analysis, the feature preference of each voxel was assigned based on the orientation preference that produced the maximum response in that voxel. However, we also adopted another approach in which we used a GLM to fit a series of circular Gaussians with different preferred orientations and bandwidths (using a set of Gaussian reference functions that peaked at each orientation and a set of bandwidths that ranged from.1: .1: 1, for a total of 8 preferred orientations * 10 possible bandwidths = 80 different GLMs). Then, we selected the scaled Gaussian from this set of 80 that accounted for the most variance in the response of the voxel, and used the parameters of that Gaussian to estimate orientation preference (specifically, the central tendency of this best fitting Gaussian). While this approach is slower than using a simple MAX rule, it is still computationally expedient and it provides a more principled way of assigning orientation preference. However, this new method did not yield a qualitatively different pattern of data compared to the max rule. Shown here are additive scaling, multiplicative and bandwidth scaling with attention, collapsed across all visual areas; note the similarity between these patterns and those shown in Figures 6, 7, and Supplemental Figure 3. As in the main data set, additive scaling increased with MI ($F(3,18)=16.27$, $p<0.001$), as did multiplicative scaling ($F(3,18)=5.59$, $p<0.007$). No significant changes in bandwidth were observed with attention. Error bars reflect ±1 S.E.M. across subjects.



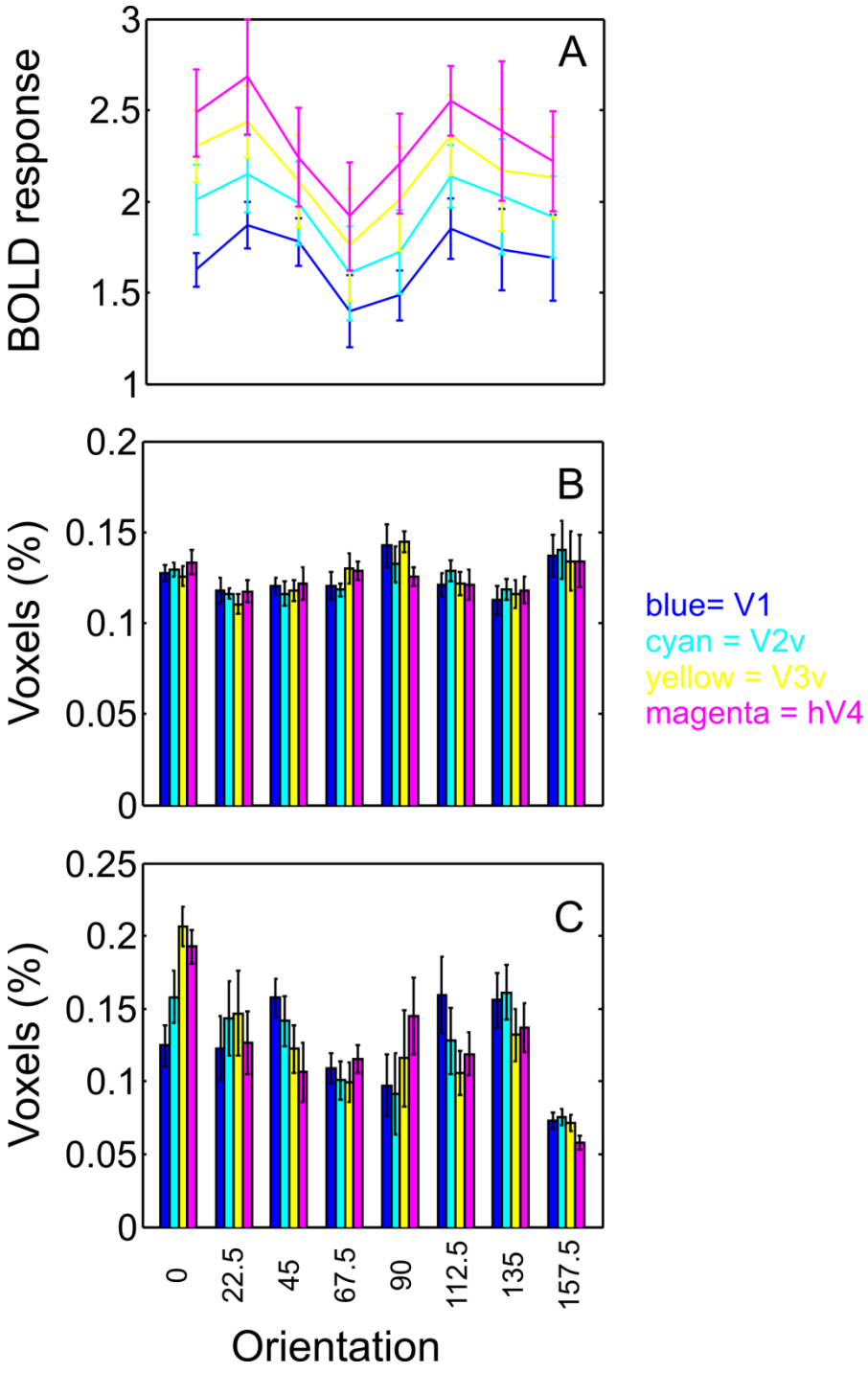
Supplemental Figure 2.8 Additive, Multiplicative and Bandwidth scaling – collapsed across all visual areas – when assigning orientation preference without correcting for main effects of stimulus orientation that had a common influence on the response of every voxel. As in the main data set, additive scaling increased with MI ($F(3,18)=16.1$, $p<0.001$), as did multiplicative scaling ($F(3,18)=4.9$, $p<0.02$). No significant changes in bandwidth were observed with attention. Error bars reflect ±1 S.E.M. across subjects.



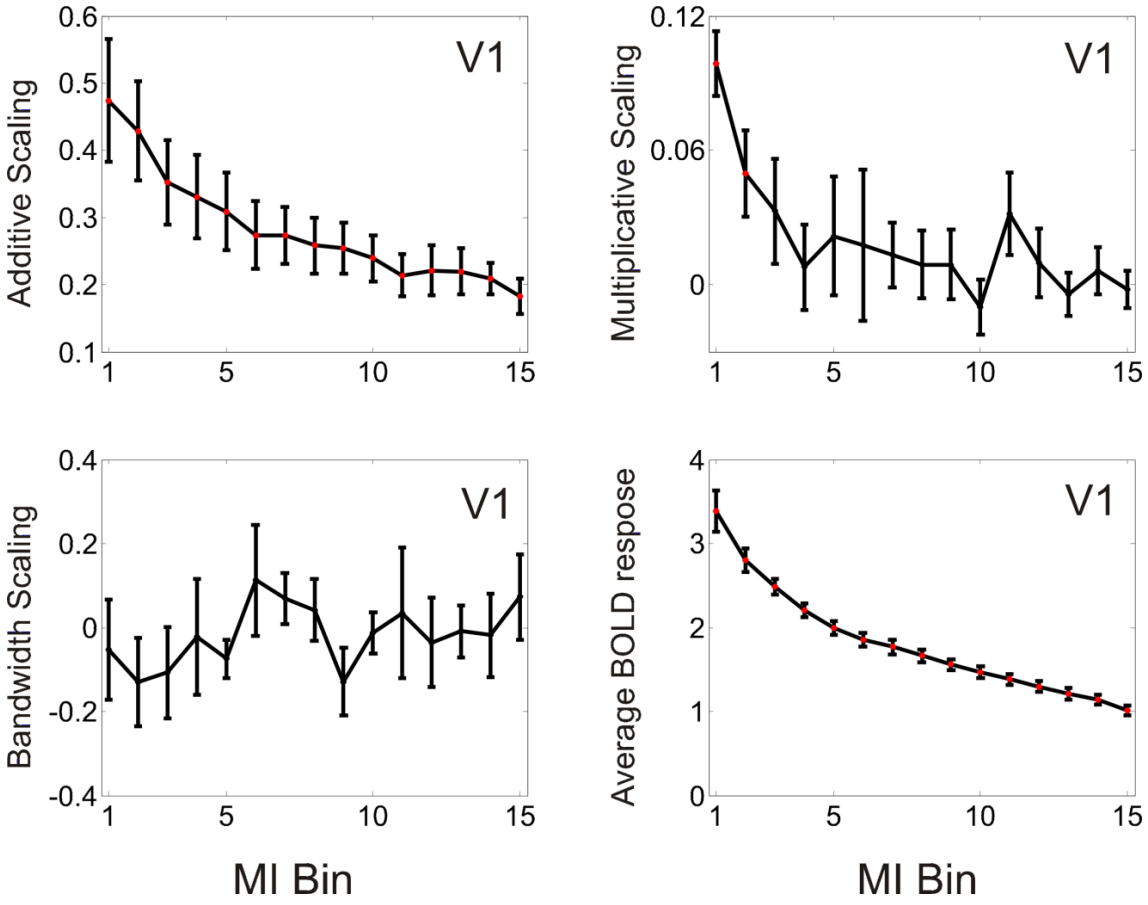
Supplemental Figure 2.9 Cortical location of high MI voxels (top 25%) in V1 for 3 subjects, projected onto computationally inflated representations of each subjects cortical surface. Color legend denotes normalized MI for voxels (which can range between 0 and 1).



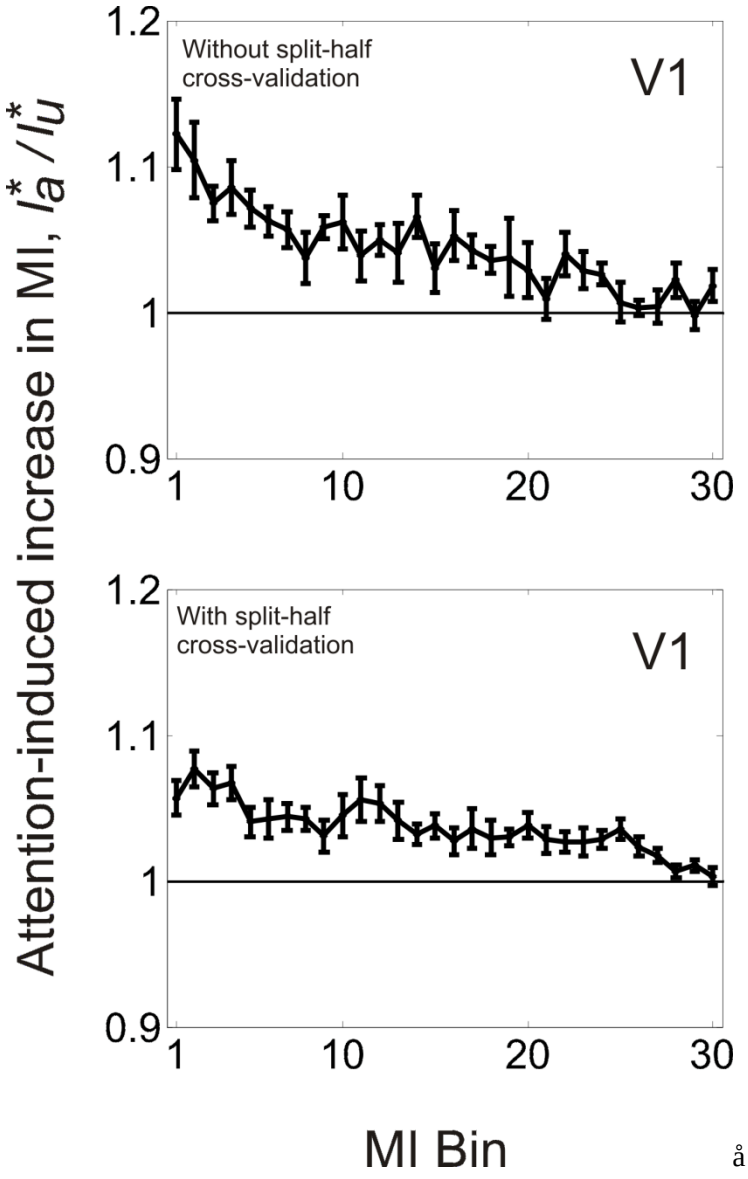
Supplemental Figure 2.10 Voxel Tuning functions in V1, V2v, V3v and hV4 for top 25% of voxels (ranked by overall MI) for attended stimuli (solid) and unattended stimuli (dashed) after removing additive scaling ($\Delta\alpha$) from both raw data and fitted curves. Error bars reflect ±1 S.E.M. across subjects.



Supplemental Figure 2.11(A) Average BOLD response across all orientations. B) Distribution of orientation selectivity of voxels after correcting for main effects of stimulus orientation that had a common influence on the response of every voxel. C) Distribution of orientation selectivity without correction for main effects. Error bars reflect ±1 S.E.M. across subjects.



Supplemental Figure 2.12 (A) Additive scaling, (B) Multiplicative scaling, (C) Bandwidth scaling (attended *minus* unattended), and (D) overall BOLD response magnitude for voxels in V1 ranked by their MI score. Each point on x-axis depicts an independent bin of voxels within top x% of voxels ranked by their MI score. Corresponding y-axis depicts the mean additive scaling, multiplicative scaling and bandwidth scaling with attention for mean voxel tuning function of that group. Red color indicates data points that reached significance by repeated measures t-test ($p<0.05$). Error bars reflect ±1 S.E.M. across subjects.



Supplemental Figure 2.13 Ratio of normalized MI (see Methods) for voxels in V1, V2v, V3v and hV4 for attended and unattended stimuli. (A) Voxels were sorted into independent bins based on their normalized MI score computed only from unattended responses. (B) Same as A, but voxels were sorted into independent bins based on the unattended data using only ½ of the data, and the ratio between attended and unattended MI was then evaluated in the other half of the data (and then this procedure was repeated using the other half of the data set to sort voxels into MI bins and to evaluate the effects of attention). The abscissa marks different groups of voxels based on their MI bin. refers to normalized MI (see Methods) derived from unattended responses and attended responses, respectively. Ordinate refers to the ratio . High MI voxels showed higher increase in overall information content with attention. Error bars reflect ±1 S.E.M. across subjects.

2.6 APPENDIX

Analysis relating voxel tuning function and neural tuning function.

The sole intention of the following theorems is to establish a link between the neural tuning function of a *hypothetical* homogenous voxel composed of neurons selective for the same feature and that voxel's tuning function, or VTF. Given this link, the rest of the assumptions underlying our simulations (Supplemental Figures 1 and 2), i.e. that a heterogenous voxel with Gaussian orientation selectivity will still have a Gaussian VTF, follow (and these assumptions also follow from the work of Nevado et al., 2004). By no means are we suggesting that any real voxel that we empirically measured using fMRI was comprised of neurons that all prefer the same orientation. However, previous work by Nevado et al. (2004) assumed a linear relationship between neural responses and BOLD responses. Next, they derived the quantitative relationship between neural MI and voxel level MI. Here, we wanted to relax the assumption of linearity and to show that if we only assume a monotonic relationship between neural responses and BOLD responses (consistent with most available data: e.g. Logothetis et al., 2001; Logothetis, 2003), that a VTF with a Gaussian profile is still possible from a homogenous voxel comprised of underlying neural populations with Gaussian tuning function. Once this relationship has been established, we can also infer that voxels with heterogeneous composition would have a Gaussian VTF as long as the distribution of orientation selectivity remains unimodal (see next paragraph for more on this point). This general model forms the backbone of simulations done to produce Supplemental Fig.1 and Supplemental Fig. 2, where VTFs of voxels with different neuronal compositions were produced by linear addition of homogenous neural tuning functions from each comprising neural population accompanied by divisive normalization (similar to Nevado et al., 2004).

While the assumption of unimodality in the neural response profile underlying a VTF is certainly idealized and strictly unrealistic, there is a large body of neurophysiological literature that suggests that feature selective columns in visual cortex are arrayed in a roughly continuous manner across the cortical sheet (e.g., Albright, 1984; Britten, 1998; Mountcastle, 1997; Hubel and Wiesel, 1969). Therefore, since adjacent neural clusters do not differ substantially in their orientation preference, a situation where a voxel contains 2 or more populations of neurons that differ substantially in their orientation preference, without any neuronal populations with intermediate selectivity, should only rarely arise. Voxels should thus generally have a near Gaussian distribution of orientation selectivity, with the spread of selectivity governed by the size of the voxel and heterogeneity of the comprising neurons (see also Nevado et al., 2004). Again, though it is possible that neural population response profiles with two or more distinct peaks may sometimes arise within a single voxel, but we posit that this is an exception rather than a rule. Moreover, such voxels would tend to be excluded from our analyses since they would not have sharp VTFs and consequently also have lower MI. In either case, the proofs below represent a first step towards a more formal characterization between the activity of feature-selective neurons and their projection onto the relatively coarse spatial scale of BOLD fMRI.

Theorems

Let $N(\theta_0,\theta)$ represent the average neural tuning function of a population of analogously tuned neurons, all having θ_0 as the preferred orientation. Let $B(\theta_0,\theta)$ represent the corresponding voxel tuning function derived by sampling the neural population represented by $N(\theta_0,\theta)$.

We propose transformation function f_T: $N \rightarrow B$ that expresses the relationship between neural activity and BOLD activation, such that $B(\theta_0,\theta) = f_T(N(\theta_0,\theta))$.

From previous reports of the monotonic relationship between BOLD response and Neural response (Logothetis et al., 2001; Logothetis, 2003), we assume that f_T bears the property of monotonicity i.e. $f_T(\alpha) > f_T(\beta)$ iff $\alpha > \beta$.

$\therefore B(\theta_0,\theta_a) > B(\theta_0,\theta_b) \Leftrightarrow N(\theta_0,\theta_a) > N(\theta_0,\theta_b)$, $\forall$a,b s.t. $\{\theta_a, \theta_b\} \in$ set of all possible orientations, Θ **I.**

Note: In the following proofs, notation for functions *B* and *N* will omit preferred orientation θ_0 for simplicity.

Theorem 1. Voxel tuning function *B* will have maxima at the same angle as neural tuning function *N*.

Proof:

Let *N* have a maxima at $\theta = \theta_m$.

$\therefore N(\theta_{m-1}) < N(\theta_m) > N(\theta_{m+1})$, where $\theta_{m-1} < \theta_m < \theta_{m+1}$ **II.**

From **I**, $N(\theta_m) > N(\theta_{m+1}) \Rightarrow B(\theta_m) > B(\theta_{m+1})$, and **IIIa.**

$N(\theta_m) > N(\theta_{m-1}) \Rightarrow B(\theta_m) > B(\theta_{m-1})$ **IIIb.**

From **II**, **IIIa** and **IIIb**, $B(\theta_{m-1}) < B(\theta_m) > B(\theta_{m+1})$

$\therefore \theta = \theta_m$ is the maxima of *B*

Theorem 2. If neural tuning function *N* is unimodal then voxel tuning function *B* will also be unimodal.

Proof by refutation:

Let *N* be unimodal and have maxima at θ_m.

$\therefore N(\theta_{m-i}) < N(\theta_m) > N(\theta_{m+i})$ $\forall$i s.t. $\theta_{m\pm i} \in \Theta$ **IVa.**

And, $N(\theta_{\varphi-1}) < N(\theta_\varphi) < N(\theta_{\varphi+1})$ $\forall\varphi$ s.t. $\theta_\varphi < \theta_m$ and $\theta_\varphi \in \Theta$ **IVb.**

And, $N(\theta_{\varphi-1}) > N(\theta_\varphi) > N(\theta_{\varphi+1})$ $\forall\varphi$ s.t. $\theta_\varphi > \theta_m$ and $\theta_\varphi \in \Theta$ **IVc.**

Let *B* be bimodal and have maxima at θ_m and θ_p, with $\theta_m \neq \theta_p$. **Va.**

$\therefore B(\theta_{m-1}) < B(\theta_m) > B(\theta_{m+1})$

And, $B(\theta_{p-1}) < B(\theta_p) > B(\theta_{p+1})$

$\therefore N(\theta_{p-1}) < N(\theta_p) > N(\theta_{p+1})$ **Vb.**

However, if $\theta_p < \theta_m$ then **IVb** $\Rightarrow N(\theta_p) < N(\theta_{p+1})$, which contradicts **Vb**

And, if $\theta_p > \theta_m$ then **IVc** $\Rightarrow N(\theta_{p-1}) > N(\theta_p)$, which contradicts **Vb**

$\therefore \theta_p = \theta_m$ and this contradicts **Va**

Hence, the assumption of *B* being bimodal is invalid. This can also be extended by analogy (or induction) to the assumption of *B* being multimodal.

Therefore, if *N* is unimodal then *B* has to be unimodal as well.

Theorem 3. If neural tuning function *N* is Gaussian having preferred orientation θ_0, then voxel tuning function *B* is also going have preferred orientation θ_0 .

Proof:

If *N* is a Gaussian then *N* is going to be unimodal. Therefore from theorem 2, since *N* is unimodal *B* is also going to be unimodal. **VIa.**

Since *N* has θ_0 as its preferred orientation then *N* has maxima at θ_0. Therefore from theorem 1, *B* is also going to have maxima at θ_0. **VIb.**

$\therefore$ *B* is going to have θ_0 as its preferred orientation {From **VIa** and **VIb** }

Note: We did not infer analytically that *B* will also be an exact Gaussian if *N* has a Gaussian distribution. However, it can be inferred that *B* will have a Gaussian-like monotonic fall-off on either side if its mode { From **I, IVa, IVb** and **IVc** }.

CHAPTER 3

ATTENTION IMPROVES INFORMATION TRANSMISSION BETWEEN V1 AND MT

The efficiency of visual processing depends on precise neural communication between cortical areas. Using fMRI and an encoding model that estimates responses in populations of feature-selective, we investigated the impact of selective attention on communication between two areas – V1 and MT – that synergistically process motion information. Consistent with previous single-unit recording studies, we observed that attention modulates direction-selective populations in both V1 and MT in a similar manner; there are enhanced responses in neural populations tuned to the attended direction, and an attenuation of responses in neural populations tuned away from the attended direction. Furthermore, the modulation in cross-area correlation between similarly-tuned populations of V1 and MT neurons is inversely related to the modulation in their mean response. A computational model suggests that this pattern of response modulation and cross-area correlation is an outcome of motion computation in MT and attention-related modulations of local neuronal correlations within V1. Most importantly, these attentional modulations have a constructive functional role; they lead to an increase in the efficacy of neural communication – as measured by mutual information – between direction-selective populations of V1 and MT neurons.

3.1 INTRODUCTION

Sensory systems have a limited information-processing capacity. To mitigate this limitation, a subset of incoming sensory information is selectively processed based on relative behavioral relevance. This mechanism, referred to as selective attention, modulates the response characteristics of neurons encoding relevant stimuli. For instance, attending to a particular spatial location or visual feature modulates multiple aspects of neural firing: gain is increased (Martinez-

Trujillo and Treue 2004; McAdams and Maunsell 1999), the overall variability of neural responses decreases (Bressler and Silver 2010; Mitchell et al. 2007), and the structure of noise correlations changes (Cohen and Maunsell 2009; 2011; Mitchell et al. 2009). Collectively, these modulations should improve the signal-to-noise ratio of neural representations for attended stimuli in early visual areas such as V1 or MT (Desimone and Duncan 1995; Reynolds and Heeger 2009). More recently, attention has also been shown to increase the information contained in population responses in V1, thereby implying an improvement in the encoding capacity of population codes (Saproo and Serences 2010).

Although informative, this framework is based on attention research that has focused predominately on neural activity within a single visual area such as V1, V4, or MT. However, isolated activity in a single visual area is not sufficient to support perception as distinct areas communicate and co-ordinate during information processing. For instance, neurons in MT are thought to pool input from neurons in V1 to gain their complex direction-selective receptive fields (Beck and Neumann 2011; Nishimoto and Gallant 2011; Rust et al. 2006; Simoncelli and Heeger 1998). Therefore, an accurate formation of neural representation for stimulus direction in MT is facilitated by the quality of the neural representation in V1 as well as the accurate transfer of this information to MT. Motion perception is also critically facilitated by feedback information from MT to V1 (Pascual-Leone and Walsh 2001; Silvanto et al. 2005).

Importantly, the quality of information transfer between distinct neural populations, such as between neural populations in V1 and MT, is impacted by the stochastic nature of neural communication due to unreliable synaptic signaling and other biological processes that introduce channel noise (Conti and Wanke 1975; Manwani and Koch 2001; Schneidman et al. 1998; Vinje and Gallant 2002; White et al. 2000). According to Shannon's channel coding theorem for communication networks (Cover and Thomas 1991; Latham and Roudi 2009; MacKay 2004), an

improvement in encoding capacity of codes in upstream areas – such as found in V1 (Saproo and Serences 2010) – should improve the decoding of information when received by a downstream area – such as MT – after being transmitted over a hypothetical noisy channel. Therefore, given the ubiquity of such inter-cortical interactions and given their importance in perception (Felleman and Van Essen 1991), it is important to understand the impact of selective attention on the dynamics of information transmission between sensory areas.

Using BOLD fMRI and a forward encoding model (Brouwer and Heeger 2011; 2009; Nishimoto and Gallant 2011; Serences and Saproo 2011), we investigated the impact of selective attention on: 1) the mean response of direction-selective populations within V1 and MT, 2) noise correlations between similarly tuned populations in V1 and MT, and ultimately, and 3) overall mutual information between V1 and MT. The pattern of modulation of mean response was found to be similar to previous single-unit studies (Martinez-Trujillo and Treue 2004; Treue and Maunsell 1999), but the pattern of cross-area correlations was found to be inversely related to mean response. This result can be explained using a linear- nonlinear cascade model of motion computations (Rust et al., 2006) that is constrained by recent empirical observations about the influence of attention on pair-wise noise correlations (Cohen and Maunsell 2011). Finally, the data suggest that these attention-related modulations in response amplitude and cross-area correlation lead to an overall increase in bidirectional information coupling between V1 and MT, suggesting a role of attention in promoting efficient communication between anatomically segregated populations of sensory neurons that synergistically encode basic sensory features.

3.2 METHODS

Stimulus and Task: The stimulus consisted of an annulus of white dots moving on a gray background (Figure 1), also known as a random dot kinetogram. In each 5 second trial, the dots moved in one of 9 possible direction (0°, 40°, 80°, …, 320°) with 100% coherence and constant

speed (6 pixels per second). The contrast of each dot was sampled from a normal distribution with a mean and variance of 0.41 and 0.1 respectively. There were 2 types of trials: one where subjects were instructed to respond to a transient change in the mean contrast (0.05-0.2, mean~0.12) of moving dots (*attend contrast*), and the other where subjects had to respond to a transient change (5-20°, mean~15°) in the direction of moving dots (*attend motion*). The timing of the target onset was randomized for each trial. Note that each trial contained both targets (brief change in direction as well as brief change in contrast), however subjects were to only respond to one and ignore the other based on the instructions given at the start of the scan (a scan was a block of 36 trials). There were 4 scans for each trial type, with 4 trials per scan for each direction of motion. Task difficulty for two trial types was matched for each subject (average accuracy ~65%).

Data acquisition and post-processing: BOLD fMRI data was collected from 12 neurologically healthy subjects having age between 18 and 35 years and who were affiliated with University of California, San Diego. Each subject trained on both task types for at least an hour prior to the actual fMRI scanning session.

MRI scanning was carried out on a GE MR750 3-Tesla scanner equipped with an 8-channel SENSE head coil at the Center for Functional Magnetic Resonance Imaging (CFMRI), University of California, San Diego. Anatomical images were acquired using a MPRAGE T1-weighted sequence, which yielded images with a $1mm^3$ resolution (TR/TE = 11/3.3ms, TI = 1100ms, 150 slices, flip angle=18^0, with no SENSE acceleration). Functional images were acquired using a 2D gradient echo planar imaging (EPI) pulse sequence, which covered the whole brain with 33 oblique transverse slices. Slices were acquired in sequential order with 3mm thickness and 1mm gap to avoid slice crosstalk; thus a 50mm thick slab was acquired (TR/TE =

1000/30ms, flip angle = 90°, image matrix = 64 (AP) × 64 (RL), with FOV = 240mm (AP) × 240mm (RL), SENSE factor = 2, voxel size = 3mm × 3mm × 3mm, IF =128Hz).

Data analysis was performed using BrainVoyager QX (v 1.86; Brain Innovation, Maastricht, The Netherlands) and custom time-series analysis routines written in Matlab (version 10.1; The Math Works, Natick, Massachusetts). All EPI images were slice-time corrected, motion-corrected (both within and between scans) and high-pass filtered (3 cycles/run) to remove low frequency temporal components from the time-series.

Independent Region of Interest (ROI) localizer scans: Two independent ROI localizer scans were run for each subject to identify voxels in V1 and MT that responded to the spatial position occupied by the stimuli. The stimuli used in localizer scans were similar to the stimuli used in main experiment scans, however with following key differences: A) Rather than having 1 annulus of moving dots, there were two annuli (outer and inner) that taken together covered the same spatial region as the annulus used in main experiment scans. However, on any trial, either the inner annulus or the outer annulus was displayed. The dots in the outer annulus were slightly bigger in size compared to the inner annulus. B) In a given trial, the dots either moved coherently in one direction (motion trials) or they randomly flickered at their static positions (snow trials). C) Subjects task was to respond to a change in speed of moving dots or to a change in contrast of flickering dots, as the case may be. D) There were multiple targets (change in speed, or change in contrast) within each trial and subjects were asked to respond to each.

To identify voxels that responded to the retinotopic position of the stimulus aperture, data from the functional localizer scans were analyzed using a general linear model (GLM) that contained two regressors; one marking motion trials and other marking snow trials. Voxels that responded more to motion compared to snow within each visual area were retained for further

analysis if they passed a statistical threshold of $p<.01$, corrected for multiple comparisons using the False Discovery Rate (FDR) algorithm implemented in BrainVoyager QX.

Retinotopic mapping procedures: Retinotopic mapping was obtained in one to two scans per subject, using a checkerboard stimulus and standard presentation parameters (stimulus flickering at 8 Hz and subtending 60° of polar angle; Engel et al., 1994; Sereno et al., 1995). This procedure was used to identify ventral visual areas V1, V2v, and V3v. To aid in the visualization of early visual cortical areas, we projected the retinotopic mapping data onto a computationally inflated representation of each subject's gray/ white matter boundary. Visual areas were then defined manually according to visual markers of hemi-field and quarter-field reversals (Wandell et al. 2007).

Data Analysis:. Hemodynamic response function was estimated using a finite impulse response (FIR) model based de-convolution of BOLD time series across all voxels in an ROI (Kay et al. 2008). The magnitude of the event-related BOLD response in each voxel of ROI was estimated using a GLM and a design matrix that modeled the response in each trial as the scalar multiplier (beta weight) of a canonical double-gamma function.

Encoding model: BOLD responses have been shown to be indirectly but linearly related to neural activity (Heeger et al. 2000; Kahn et al. 2011). However, BOLD response from a voxel at current resolutions would be a linear mixture of responses from many sub-populations of neural with different feature-selectivity and response profiles (Boynton 2005; Saproo and Serences 2010). In order to get the most accurate response estimate for direction-selective neural populations in V1 and MT, we decomposed voxel responses into direction-selective *channel* responses using a forward encoding model (Brouwer and Heeger, 2009, 2011; Serences and Saproo, 2011). Each voxel response was modeled as a linear sum of weighted responses of different direction-selective channels. A set of 9 channels was used, with the direction selectivity

of channels linearly spaced between 0 and 360 degrees (0°, 40°, 80°, …, 320°), and each channel's tuning function was modeled either using a $\sin^4$ function or a Kronecker delta function (we present results from both analysis, which are qualitatively similar). The $\sin^4$ is a steerable filter while a delta function forms an orthonormal basis set; the latter being relevant when computing cross-area correlations. A general linear model was set up: B = (W)C, where B is a matrix of voxel responses for all trials, C is a matrix of channel responses for all trials and W is a matrix that provides channel weights for each voxel. BOLD data was divided into two parts; training set B_{train} and testing set B_{test}. Training set was used to compute ordinary least-squares estimate of W from the model equation (1), assuming unit channel responses C_u, using equation (2) Afterwards, testing set B_{test} was used to compute channel responses C_{test} using previously computed weights W, using equation (3). Data was partitioned into training and testing set iteratively and using hold-2 scans-out principle (one attend motion scan and one attend contrast scan on each iteration).

$$B_{train} = WC_u \quad (1)$$

$$W = B_{train}C_u^T(C_uC_u^T)^{-1} \quad (2)$$

$$C_{test} = (W^TW)^{-1}W^TB_{test} \quad (3)$$

Multivariate Mutual Information: Shannon's channel coding theorem provides a method of assessing the efficacy of different codes in a noisy communication system (such as in Figure 6A). The theorem states that for any coding scheme that is used to transfer information over a noisy channel, the probability of correct decoding by the receiver depends on the Mutual Information (MI) between sent codewords and received codewords (Cover and Thomas 1991; Latham and Roudi 2009). Therefore, an increase in mutual information across a channel would imply an increase in the efficacy of the coding scheme used for communication over that channel.

. Given a robust model of motion computation that depends on communication between V1 and MT (Rust et al. 2006) and the fact that the majority of projections into MT are directly from V1 (Born and Bradley 2005), we modeled V1 to MT communication systems as a noisy channel with unknown but fixed noise characteristics. Given, this assumption, the efficacy of neural communication about the attended feature between V1 and MT can be measured by mutual information between the responses of direction-selective neural populations in the two areas. We used a multi-variate measure of mutual information (Chai et al. 2009) that computed total mutual information between all direction-selective channel responses in V1 and those in MT. A multivariate measure has an advantage over univariate measure that it would account for any synergistic encoding of information by channels (Schneidman et al. 2003), and provide a more accurate picture of cross-area information coupling. The equation for multi-variate Mutual Information between two *d*-dimensional response vectors *X* and *Y* with *n* observations is

$$I(X,Y) = -\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{p_k(x_i)p_k(y_i)}{p_k(x_i,y_i)}\right) \quad (4)$$

$$p_k(x_i) = \frac{k}{n-1}\left(\frac{\Gamma\left(d/2+1\right)}{\pi^{d/2}}\right)\left(\frac{1}{r_k(x_i)^d}\right) \quad (5)$$

Here, Γ is the gamma function, x_i is the ith observation in *X*, and $r_k(x_i)$ is the distance of between x_i and its kth nearest neighbor (similarly for y_i). We set k equal to $\sqrt{n}$ in our analysis since k should scale with the size of data (Misra et al. 2010; Mnatsakanov et al. 2008). Dimensionality *d* was set equal to the number of channels i.e. 9, and *n* was set equal to the number of trial responses. Bias correction on mutual information estimates was done by randomly shuffling data labels a thousand times, computing mutual information each time, and finally averaging mutual information derived from each shuffled data to yield an estimate of bias.

Model simulation of cross-area correlation for MT motion computation: We set up 60 neural pools consisting of >100 neurons each (*N*). All neurons had a canonical circular Gaussian tuning function with bandwidth (full width half maximum) of 90°, however each pool of neurons had a unique direction preference chosen from the set of 60 possible directions (6°, 12°, 18°, …, 360°). The variance of neurons in each pool (σ_i^2) was set equal to their mean firing rate in response to a stimulus (Poisson distribution). The correlations among each pool of neurons was set to be inversely proportional to their mean firing rate (Cohen and Maunsell, 2011) (e.g. a firing rate of 40Hz yielded a correlation coefficient (c_i) of 0.15, 60Hz yielded c = 0.1375, and 20 Hz yielded c = 0. 1625). Thereafter, we set the trial-by-trial variance of the average of each pool $\sigma_i'^2$ to $c\sigma^2$ since $N>>1/c$ (Abbot and Dayan, 1999). The covariance matrix for the pool of neurons was then constructed according to the limited range correlation model (Snippe and Koenderinck, 1992; Abbot and Dayan, 1999); here covariance between ith and jth neuronal pool is given as $Q_{ij} = \sqrt{(\sigma_i'^2 * \sigma_j'^2)} * \rho^{|i-j|}$ where $\rho = \exp(-\Delta/L)$. Here, Δ is the distance between tuning preference of neural pools (6°), and L is the correlation length, which isset to 4 keeping $L<\Delta$ (Snippe and Koenderink 1992). Using the above expression of covariance matrix and a mean neural response based on Gaussian tuning function, the averaged response of each neural pool was computed for each stimulus presentation in the simulation using matlab function *mvnrnd*.

The response profile of direction-selective MT neurons was computed as follows: First, for each stimulus direction (6°, 12°, …, 360°), the averaged response of each V1 direction selective neural pool is computed. Thereafter, these pooled responses for each stimulus direction were passed through the Linear-Nonlinear MT motion computation (Rust et al., 2006), with

following model parameters; MT scaling non-linearity = 1, MT exponent nonlinearity = 2, V1 untuned normalization factor = 0.1, V1 self-tuned normalization factor = 0.1.

The effect of attention on V1 cells was simulated by changing the baseline (β) and amplitude (α) of their tuning function; $R(\theta) = \beta+\alpha*f(\theta;\kappa)$, where R is the response of a cell for a stimulus of direction θ. The bandwidth (κ) was kept unchanged. Parameters for unattended (baseline = 20Hz, amplitude = 40Hz) and attended condition (baseline = 10Hz, amplitude = 70Hz) were chosen to reflect the observations of suppression and enhancement at population level responses in our BOLD-based dataset. We computed V1 neural pool responses and MT responses over 100,000 trials and computed correlation between V1 responses and MT responses for neurons tuned to similar direction, as well as mean response profile for MT neurons. We find the results of the simulation to be consistent with empirical observations; we find an enhancement of mean response for MT neurons tuned to attended direction while a suppression in mean response for neurons tuned away from attended direction but more importantly we find that cross-area correlations between V1 and MT neurons with tuning similar to attended direction (low-offset tuning) had a decrease in cross-correlation whereas neurons with high-offset tuning had an increase in cross-correlation.

Analytical Examination: According to cascade model of MT motion computation (Rust et al. 2006; Simoncelli and Heeger 1998), the response of MT cell is given by $M_p = \Sigma_i w_i V_i$, where M_p is the response of MT cell tuned to direction p, V_i is the response of V1 cell tuned to direction i and w_i is the weighted scaling for V_i. Therefore, correlation between any two MT and V1 neurons tuned to the same direction p is given by

$$Corr(M_p, V_p) = \frac{Cov(M_p, V_p)}{\sigma^2_{M_p}\sigma^2_{V_p}} \qquad (6)$$

where, $Cov(M_p, V_p)$ is the covariance between the cells, $\sigma^2_{M_p}$ is the variance of MT cell, and $\sigma^2_{V_p}$ is the variance of the V1 cell. The terms can be expanded to

$$Corr(M_p, V_p) = \frac{Cov(\Sigma_i w_i V_i, V_p)}{\sigma^2_{\Sigma_i w_i V_i} \sigma^2_{V_p}} = \frac{\sum_i w_i Cov(V_i, V_p)}{\left(\sum_i w_i^2 \sigma^2_{V_i} + 2\sum_{i,j} w_i w_j Cov(V_i, V_j)\right)\sigma^2_{V_p}} \quad (7)$$

According to eq.6 above, a change in correlation can be affected by 3 variables; the weights w_i , the covariance matrix for V1 cells $Cov(V_i, V_j)$, or the variance of V1 cells $\sigma^2_{V_i}$. In the model simulation, we assumed that the linear weights w_i do not get affected by short-term changes in attentional state, since they might be a functional of synaptic strength and probably are affected by longer term cognitive factors such as perceptual learning. Since approximately an equal number of neurons undergo an increase in mean response as well as a decrease, the term $\sum_i w_i^2 \sigma^2_{V_i}$ might remain largely unchanged, assuming Poisson firing statistics, but also since it is smaller in magnitude compared to other components. Cohen and Maunsell (2011) found that a change in correlation for neural populations was inversely related to a change in mean firing with attention. Given Poisson firing statistics, a changed in variance is also expected with a change in mean firing rate, therefore the relative magnitude that a change in covariance contributes to correlation is unclear. This observation is also applicable for cross-area correlations such in we found between V1 and MT. If we assume that the covariance matrix for V1 cells $Cov(V_x, V_y)$ does not get affected by the modulation of tuning functions, then only factor that affects cross-area correlation given eq.6 above is the variance of V1 cells. Specifically, it can be inferred that $Corr(M_p, V_p) \propto \frac{1}{\sigma^2_{V_p}}$ i.e. the change in correlation between similarly tuned V1 and MT cells is inversely proportional to change in the variance of the V1 cell. Since variance of neural response is largely a function of its mean firing rate, the implication of this analysis is that the change in

cross-area correlation we observe could simply be a consequence of the gain modulation of V1 cells. The observation that within area correlation modulation is a function of gain modulation might also be explained away as a consequence of a large change in variance with changing mean response levels, with marginal change in covariance between neurons. However, we cannot answer this broader question using our dataset.

3.3 RESULTS

We concurrently recorded BOLD activation levels from areas V1 and MT in 12 human subjects. In each 5 second trial, subjects viewed a circular annulus comprised of small dots that moved coherently at constant speed in one of 9 possible directions (see Figure 1). Subjects were asked to deploy attention to the entire annulus while maintaining fixation at the center of the annulus. Stimulus directions were linearly spaced between 0 and 360 degrees (0°, 40°, … , 320°), and there were an equal number of trials for each direction. At the start of a block of trials (which we call a 'scan'), the subjects were instructed to perform one of two distinct tasks in each trial; one task was to respond to small changes in the direction of motion of the dots (*attend-motion* condition), and the second task was to respond to small changes in mean contrast of the display (attend contrast or *ignore-motion* condition). After preprocessing fMRI data, we selected voxels within each ROI that showed significant direction-selectivity in independent localizer trials ($p<0.01$, corrected for multiple comparisons using False Discovery Rate algorithm). The selected voxels were used for further analysis, which involved comparison of BOLD activation between attend-motion trials and ignore-motion trials; since the displays were physically identical, any differences in the BOLD signal were attributed to feature-based attention.

First, we computed the average hemodynamic response function (HRF) for all selected voxels (see Methods section entitled 'Data Analysis' for details). Comparing HRFs between attend-motion and ignore-motion trials, we find a relative increase in hemodynamic response

after stimulus presentation, in both V1 and MT, when directing attention to motion direction (Fig. 2). The modulation was more pronounced in MT as compared to V1, consistent with a previous fMRI study (O'Craven et al. 1997). Furthemore, attentional modulation in V1 precedes similar modulation in MT by 2-3s; whereas V1's HRF had an increase during 1-2s post-stimulus 'initial dip' ($p<0.05$ at 2s), MT's HRF had an increase during 5-10s post-stimulus 'late rise' ($p<0.05$ at all time-points). This precedence effect suggests a casual nature of attention-related modulations in our data, consistent with the notion that MT cells pool inputs from V1 to derive their motion selectivity (Movshon and Newsome 1996; Rust et al. 2006; Simoncelli and Heeger 1998).

We next assessed attention-induced modulation in different direction-selective populations of areas V1 and MT. To do so, we decomposed voxel responses into direction-selective channel responses using a forward encoding model (Brouwer and Heeger 2011; 2009; Serences and Saproo 2011), which involved modeling the response of each voxel as a linear combination of weighted responses from nine direction-selective channels, each having a canonical tuning function with a different direction preference (0, 40, 80, …, 320°). We experimented with two different tuning functions for the basis set; Sine function (Figure 3)(Freeman et al. 1991) as well as Kronecker delta function (Figure 4A), see method section entitled 'Encoding model' for details. Using this forward encoding model, we converted the BOLD activation levels across the set of approximately 200-400 voxels that were measured on each trial to a set of 9 channel responses that represent the relative activation level in each hypothetical direction-selective channel. Attending to the direction of motion led to increase in the mean response for channels tuned to the stimulus direction (0, ±40° offset) while those channels that were tuned away from the stimulus direction (±120, 160° offset) were attenuated (Fig. 3 and Fig. 4A). This observation is in agreement with the previous reports of feature-based attentional modulations of firing rates of single direction-selective neurons in MT (Martinez-

Trujillo and Treue 2004). However, we found the magnitude of suppression to be more pronounced in V1 as compared to MT. This difference in the observed suppression between V1 and MT may further explain the rather small modulation of V1 HRF compared to MT HRF (in Figure 2), since approximately similar amounts of enhancement and suppression in distinct neural populations in V1 would not change an average measure of BOLD activity such as HRF.

We next investigated the impact of attention on cross-area correlations; we computed correlation between response vectors of like-tuned channels in V1 and MT, e.g. between the channel tuned to 0° in V1 and the channel tuned to 0° in MT. Figure 4B shows these cross-area correlations as a function of the distance between the direction preference of each channel and the direction of the stimulus that evoked the response on a given trial. We collapsed across negative and positive offset to improve power because the mean response profile (in Fig. 4A) was symmetrical around attended direction (i.e. around offset 0°). The results suggest that when motion-direction is attended, cross-area correlations decrease for channels tuned to the direction of the stimulus while cross-area correlations increase for channels tuned away from the direction of the stimulus. A comparison of the data in figures 4A and 4B reveals that attention modulation of cross-area correlation is inversely proportional to the modulation of mean response for the corresponding channels.

To understand this pattern of cross-area correlations, we simulated a well-known model of V1-to-MT information transfer during motion processing, and constrained the model based on recent empirical observations about attention modulation of within-area correlations. We simulated MT motion computation using the linear-nonlinear cascade model (Rust et al. 2006; Simoncelli and Heeger 1998); here, the response of a MT direction-selective cell is modeled as a probabilistic non-linear function of the sum of weighted responses from a population of V1 direction-selective cells (see Fig 5A and methods). However, the model was constrained in 3

ways: First, we assumed limited-range correlations between direction-selective cells in V1, i.e. the strength of correlations between any two direction selective cells was a function of the difference in their direction selectivity (see Fig 5B for an example and see Methods for a mathematical description)(Abbott and Dayan 1999; Jazayeri and Movshon 2006; Snippe and Koenderink 1992). Second, we made variance in trial-by-trial responses of V1 cells scale proportionally to their mean response according to the Poisson noise model (White et al. 2000). Third, noise correlations between similarly tuned V1 cells were scaled negatively with the mean response of those cells. This last constraint was kept in accordance with the recent observation by Cohen and Maunsell (2011) who showed that attention causes neural populations tuned to the attended feature to become less correlated while populations tuned away from the attended feature become more correlated. Moreover, they show that the magnitude of mean response of a like-tuned neural population is inversely correlated to the magnitude of pair-wise correlations within that population, both with and without spatial or feature attention.

To simulate effects of attention on cross-area correlation between V1 and MT, we modulated the mean responses of V1 direction-selective cells according to the modulation pattern observed empirically with attention (Figure 5C, contrast with figure 3) and derived the response of MT direction selective cells from the model. We simulated a large number (~10000) of trails with and without attention to direction of motion. The simulation produced a pattern of attention-induced modulation of MT's mean response profile that is similar to the pattern observed empirically (Figure 5D). Furthermore, the pattern of cross-area correlations output by the simulation (Figure 5E) is also qualitatively similar to the empirically-observed pattern. Attention produced a decrease in the cross-area correlation between direction-selective populations tuned to the attended feature in V1 and MT, while cross-area correlations increased for populations tuned away from the attended feature.

Finally, we investigated whether the observed attention-induced modulations of amplitude and cross-area correlation increased the overall efficiency of information transfer between V1 and MT. According to Shannon's channel coding theorem (Cover and Thomas 1991; Latham and Roudi 2009; MacKay 2004), if we assume that the conduit of information between V1 and MT forms a noisy channel of unspecified noise characteristics, then the efficacy of neural communication would be proportional to mutual information between direction-selective channel responses in V1 and MT (Fig. 6A,B). Accordingly, we measured mutual information between all direction-selective channels of V1 and MT for attend-motion and ignore-motion conditions. Furthermore, it has been suggested that neural populations employing population codes can increase efficiency by encoding information synergistically (Brenner et al. 2000). Therefore, we measured mutual information using a multivariate method of computation (Chai et al. 2009) since it would account for any latent synergy between channels within the same area, thereby providing a more accurate picture of the overall coupling between V1 and MT. We observe that when attention is directed towards the direction of motion, mutual information between V1 direction-selective channels and MT direction-selective channels undergoes a significant increase ($p<0.01$, Figure 6C). This result suggests a significant role of attentional modulations towards promoting efficient communication of task-relevant information between distinct cortical regions.

3.4 DISCUSSION

Survival critically depends on the ability to effectively react to a dynamic environment. This requires gathering information about the environment continuously via various sensory organs, and then combining this multi-modal information to produce an internal perceptual model of the environment. In this cortical hierarchy, usually the neural activity in one visual area depends upon input from another area, such as motion computation in MT being facilitated by information received from V1. Here, we provide data suggesting that attention not only

modulates the quality of neural representations within individual visual areas but also facilitates accurate communication of information between distant neural populations. Improving communication efficacy between such cortical nodes should improve the accuracy of internal perceptual model and therefore the speed and accuracy of perceptual decisions.

The disparity between the magnitude of hemodynamic response function (HRF) modulation in MT and V1 might have several plausible explanations. This disparity might be a consequence of summing small attention modulations in V1 to produce a large effect in MT since each MT direction-selective neuron is thought to pool the activity of a large group of V1 neurons (Rust et al. 2006; Simoncelli and Heeger 1998). On the other hand, it could simply be an attribute of MT's robust direction-selectivity as compared to V1; since BOLD response is proportional to the pooled activity of many neurons in each voxel; BOLD response in V1 – with its more heterogeneous population of feature-selective neurons as compared to MT – might under-represent the modulations due to direction-selective neurons. Alternatively, since HRF is a measure of average response from all neural subpopulations in an ROI, the observed disparity of HRF modulation between V1 and MT could be a consequence different pattern of modulations in neural subpopulations in the two areas. For instance, there is almost an equal number of neural populations that undergo enhancement and suppression in V1 (Figure 3), which may lead to little effect on HRF from V1. However, the magnitude of suppression in MT is almost negligible as compared to the magnitude of enhancement, leading to the large modulation of HRF we observe there.

Shannon's channel coding theorem states that for any given noisy channel, there is an upper bound on the achievable efficacy of communication using any coding scheme; this bound is called *channel capacity*. For a given noisy communication system, an optimal code should achieve a communication rate that equals channel capacity. We modeled V1 and MT as an

isolated communication system, with the channel between them having unknown but fixed noise properties. In this system, the measured pattern of direction-specific channel responses in V1 and MT form the sent and received messages respectively. Therefore, a measurement of mutual information between them should provide a measure of the efficacy of the communication model (Fig 6). The assumed communication model between V1 and MT is grounded in having both a well established motion-computation model that relies on transfer of information between V1 and MT (Rust et al. 2006; Simoncelli and Heeger 1998) as well as the fact that most direct anatomical projections into MT come from V1 (Born and Bradley 2005). Furthermore, conscious perception of motion has also been found to be critically facilitated by feedback information from MT to V1 (Pascual-Leone and Walsh 2001; Silvanto et al. 2005). In spite of evidence of predominantly direct connection between V1 and MT, even if we assume that some of the motion-relevant information is communicated indirectly from V1 to MT, our abstract communication model would subsume that special case, and our conclusions would still hold. The case where our conclusions would lose some of their relevance would be a scenario wherein a third anatomical structure is causing most of the correlated direction-specific activity in both V1 and MT. We view such a possibility unlikely because although there are direct thalamic connections to both V1 and MT, they are dwarfed by the numerosity of direct connections between V1 and MT. Moreover, the agreement of our model simulation – which does not assume any thalamic interference – with empirical observations suggests that MT cells get most of their functional selectivity by pooling information directly from V1.

Although, attention modulation of mean response profile (or tuning functions) in V1 and MT as well as the modulation of cross-area correlations might be equally contributing factors towards the increase in mutual information, we believe – based on our simulations – that the increase in cross-area mutual information might be largely driven by the modulation of mean response in V1. The simulation shows that the nature of MT computation – as modeled by

cascade linear-non-linear pooling of V1 responses – can give rise to both the observed attention modulation of MT mean response as well as the modulation of cross-area correlations. Therefore, it is likely that MT cells largely inherit attention modulations from V1, which is in agreement with the conclusions of another study that found that MT cells inherited modulations in V1 cells due to sensory adaptation (Kohn and Movshon 2003). A quantitative analysis of the cascade model (see Methods) also points towards similar conclusion, but under 2 assumptions; that the amount of suppression and enhancement across different V1 populations is roughly equal, and that covariance matrix does not get impacted by attention. The first assumption (equal amount of suppression and enhancement) is supported by our data (Fig 3,4), however the second assumption is more nuanced. Under a Poisson noise model, an increase or decrease in mean response should lead to an increase or decrease in the variance of a sensory neuron respectively. Therefore, an attention-induced modulation of mean response would have a direct impact on the observed attention-induced modulations of local correlations – if there is no change in covariance between cells – in a manner similar to that observed by Cohen and Maunsell (2011). If local correlations are indeed byproduct of gain modulation, then our analysis suggest that the modulation of cross-area correlation would also be a direct outcome of gain modulation in V1.

It has been suggested that neurons can communicate or exchange information by synchronizing their native oscillations with each other, typically in gamma band (Engel et al. 2001; Fries 2005). There is emerging consensus that attention enhances synchrony between distinct neural populations in different cortical regions (Womelsdorf and Fries 2007), such as between FEF and V4 as found recently (Gregoriou et al. 2012). If attention led to an increase in synchrony between direction-selective neurons in V1 and MT in our experimental task, it would also lead to similar empirical observations as we found. For instance, due to increased synchrony, distinct V1 and MT direction-selective neurons might functionally behave as a single pool of neurons. This would not only cause an increase in mutual information between V1 and MT, but

would also produce the same pattern of cross-area noise correlations as within-area noise correlations. However, since BOLD signal is in an indirect measure of neural activity at best, it is not clear whether a direct comparison between our results and those derived from single-unit recordings is useful or even appropriate. Nevertheless, given the slower dynamics of BOLD signal, any synchrony underlying our results would probably be due to synchronization in slower bands such as alpha (8-14Hz) and beta (14-30Hz), which are also known to be modulated by attention (Courtemanche et al. 2003; Gross et al. 2004; Palva et al. 2005; Palva and Palva 2007; Sauseng et al. 2005; Suffczynski et al. 2001; Tallon-Baudry 2004; Thut et al. 2006).

Data processing inequality (Cover and Thomas 1991) states that any information processing should mostly reduce overall information contained in a signal (or data). In the context of neural systems, this inequality suggests that successive stages of sensory filtering should diminish the overall information in sensory signals. For instance, although V1 contains both orientation information and motion information, MT mostly contains motion information. However, this *filtering* is necessary for extracting relevant information required for decision-making. Ultimately, the aim of all successive stages of filtering of sensory information is the generation of a behavioral output. Therefore, accurate behavior would require preserving more information about behaviorally relevant stimuli at each data processing stage along visual hierarchy. Our results suggest that attention achieves this in part by changing neural representations for certain features and objects such that they might become more robust to channel noise during inter-cortical communication, and lead to more accurate downstream representations. This is a new direction towards explaining the functional role of attention-induced gain modulation, which so far has been primarily considered to enhance the signal to noise ratio of neural representation of attended features, while diminishing the same for unattended features. In the larger perspective, the recent impetus towards understanding consciousness as a result of simultaneous broad-based dialogue between multiple brain areas

(Boly et al. 2011; van Gaal and Lamme 2011) underscores the importance of future investigations into how within-area modulations due to cognitive factor such as attention and adaptation impact cross-area communication throughout visual hierarchy towards the ultimate aim of uncovering the nature of perception.

Chapter 3, in part, is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material, while Serences J. was co-author.

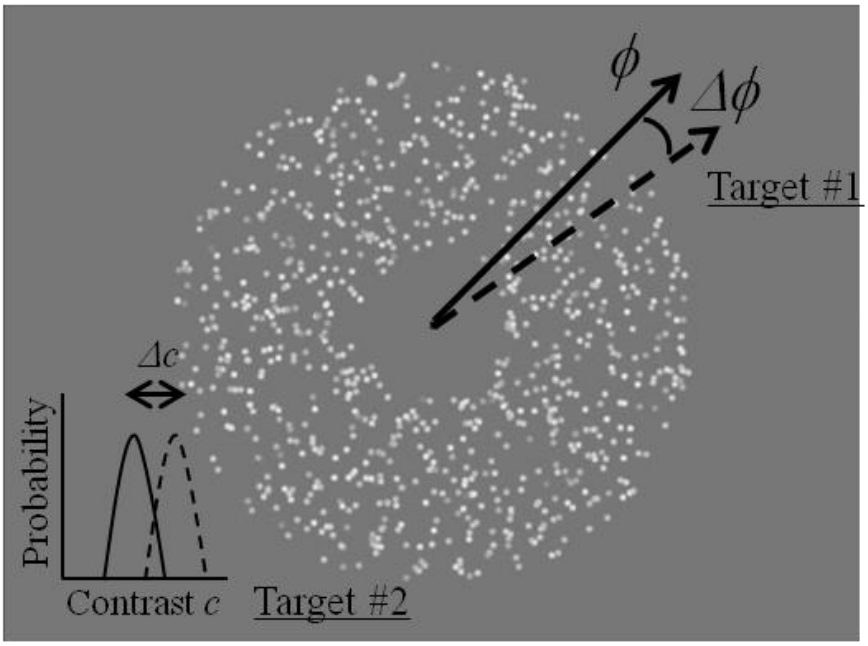


Figure 3.1. Shows a static snapshot of stimulus during a trial. On each trial, all white dots (with mean contrast c but a small variance in individual contrast) moved coherently at a constant speed in direction $\phi \in (0°, 40°, 80°, \ldots, 320°)$. On 'attend motion' trials, subjects had to detect a brief and transient change in direction, given by $\Delta\phi$. $\Delta\phi$ was titrated to yield an average task accuracy of 75%. On 'attend contrast' trials, subjects had to detect a brief and transient change in mean contrast (Δc) of the display. Δc was chosen to yield an average accuracy of 75%. Note that in each trial, both targets were present i.e. the display changed direction as well as changed mean contrast at random and mutually independent times. However, subjects were instructed to respond to one target but ignore the other depending on the instructions provided at the beginning of a block of trials.

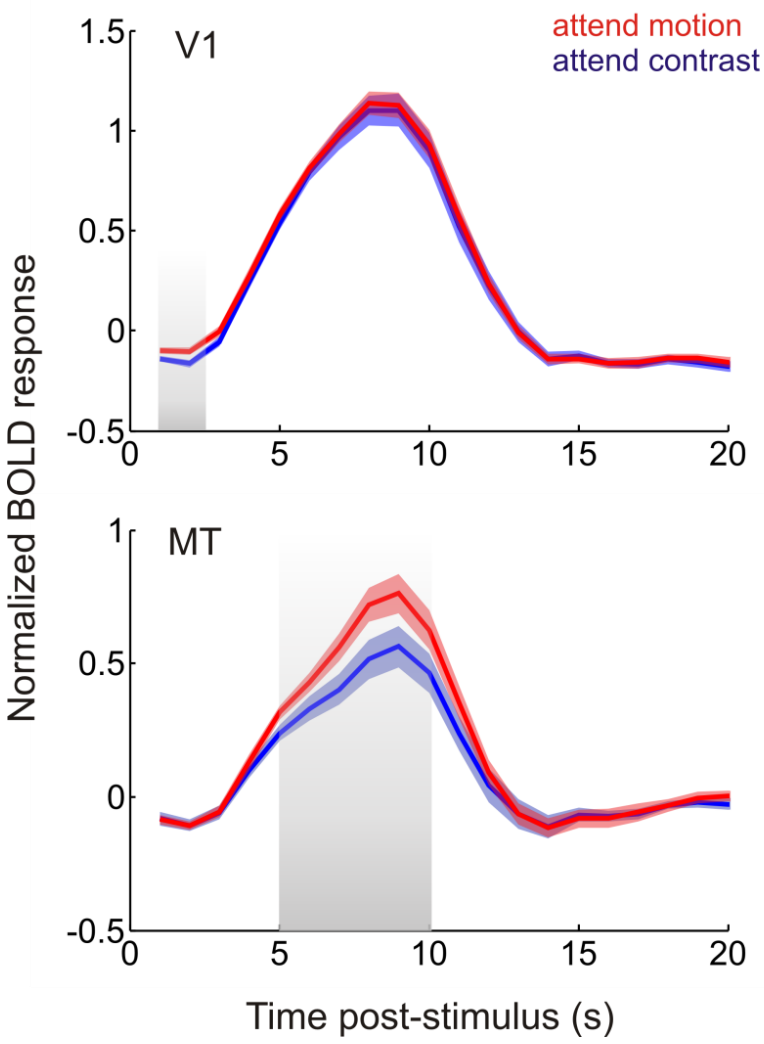


Figure 3.2 Shows hemodynamic response function (HRF) computed from 'attend motion' (red curve) and 'attend contrast' (blue curve) trials in V1 and MT. Gray portion in each panel shows time-points post stimulus that had a significant increase with attention; 2s for V1 and 6-10s for MT ($p<0.05$). The attention modulation of HRF in MT is preceded by similar modulation in V1, suggesting a largely sequential flow of information from V1 to MT. Shaded region around the curves depicts SEM.

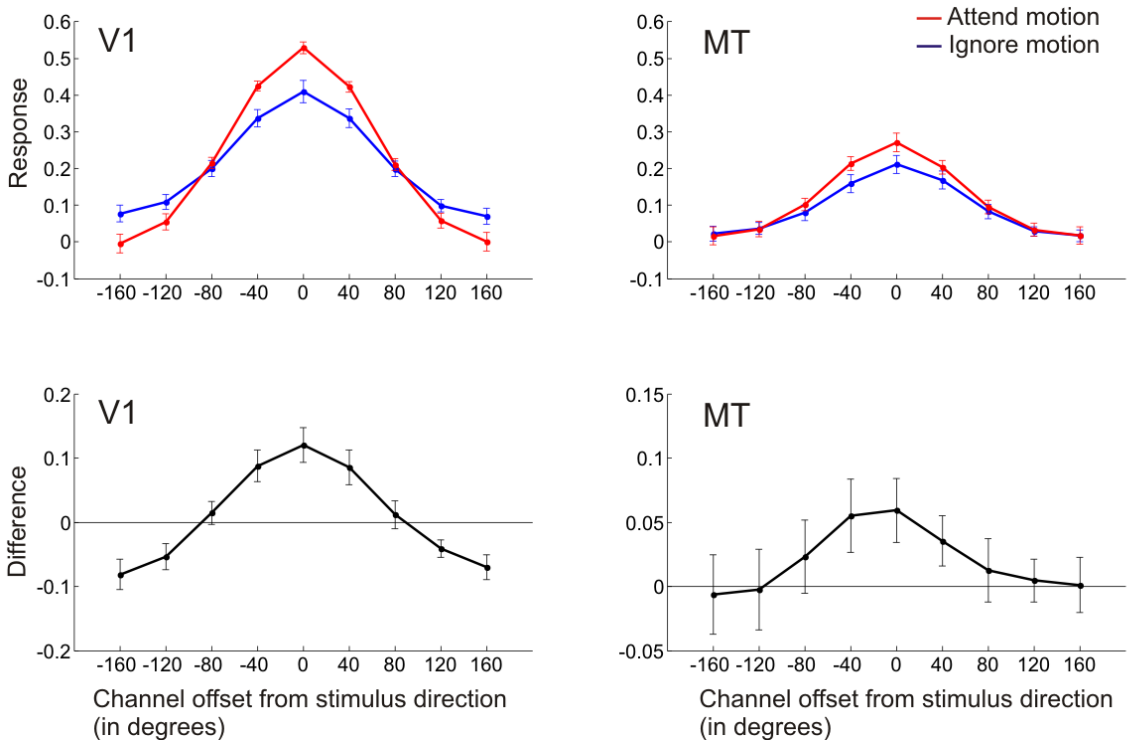


Figure 3.3 Normalized BOLD response for direction-selective neural populations in V1 and MT tuned to a direction offset (indexed by x-axis) with respect to stimulus direction. Blue curve depicts response profile when attention was directed away from stimulus direction (to contrast), and red curve depicts response profile when attention was directed towards stimulus direction. Response profiles here were generated using forward model (see Methods) with Sine^4 basis functions. Bottom Panels depict the difference between 'attend motion' and 'ignore motion' condition, highlighting the effect of attention in neural response profiles.

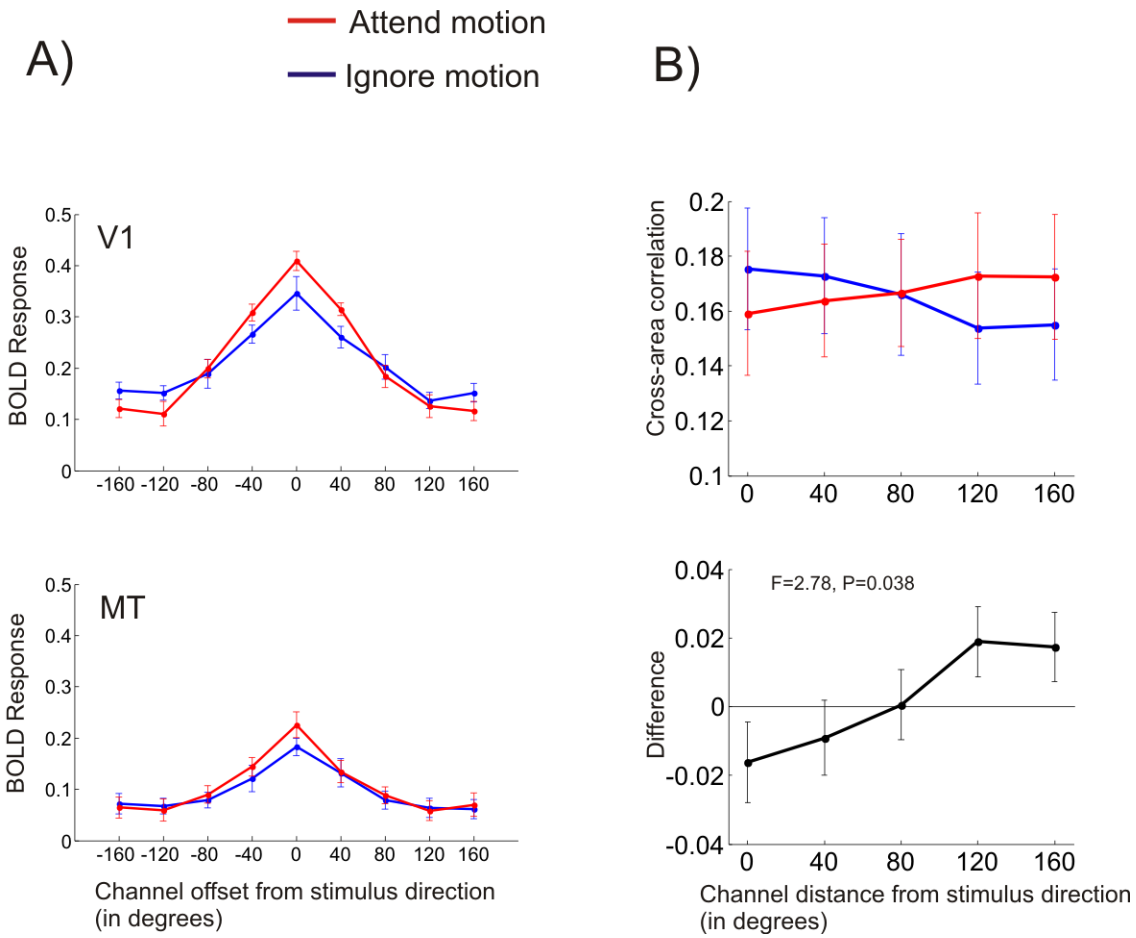
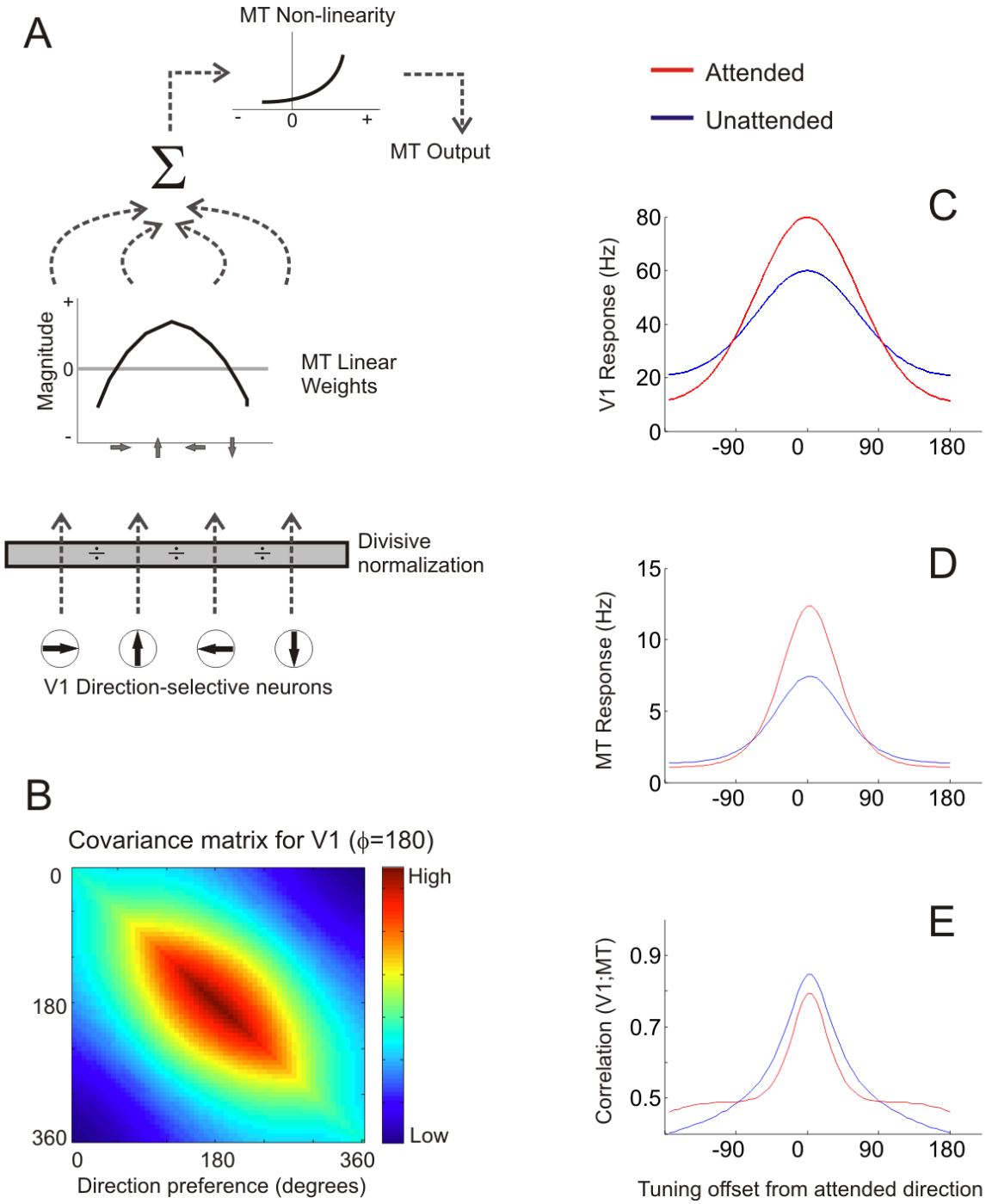


Figure 3.4 A) Similar to figure 3 but with Kroneckar delta basis function used in forward model. Normalized BOLD response for direction-selective neural populations in V1 and MT tuned to a direction offset (indexed by x-axis) with respect to stimulus direction. B) Panel shows correlation between similarly tuned direction-selective populations (channels) from V1 and MT, as shown in panel A. X-axis indexes the distance of the channel from stimulus direction; channels with same negative and positive offset were averaged to increase statistical power.

Figure 3.5 A) Cascade model of MT motion computation (Rust et al. 2006; Simoncelli and Heeger 1998): Direction-selective cells undergo divisive normalization consisting of tuned and untuned components. Afterwards, the response of each cell is weighted with a linear weighting scheme and thereafter all weighted responses are summed. Finally, the response of MT direction-selective cell is determined by passing the summed responses through a non-linear function. B) The example of covariance matrix of V1 cells in the simulations. The covariance depended on two factors; first, the variance of cells was proportional to their mean response, and second the correlation between cells was a function of the difference between their tuning preference (i.e. limited range correlation). Panel shows covariance when attended stimulus is moving in 180 degrees, and covariance matrix would depend on stimulus direction. C-E) The results of simulation run on cascade model of MT motion computation to ascertain the effect of attention on cross-area correlations. C) Mean population response profile for V1 direction selective cells, both with and without attention, manually set by the experimenter according to observed pattern of modulation in BOLD data. D) Mean population response profile for V1 direction selective cells, both with and without attention, as output by simulation of cascade model. This is similar to the pattern observed in the experimental data. E) Shows cross-area correlation profile between V1 and MT direction selective cells as output by the simulation; cross-area correlation decreases for populations that have an increase in mean response with attention, whereas the opposite effect is observed for populations that have a decrease in mean response. This is qualitatively similar to the pattern observed in our experimental BOLD data.


A
MT Non-linearity
-
0
+
MT Output
Σ
Magnitude
+
0
-
MT Linear
Weights
Divisive
normalization
V1 Direction-selective neurons
B
Covariance matrix for V1 (ϕ=180)
0
180
360
0
180
360
High
Low
Direction preference (degrees)
Attended
Unattended
C
V1 Response (Hz)
80
60
40
20
0
-90
0
90
180
D
MT Response (Hz)
15
10
5
0
-90
0
90
180
E
Correlation (V1;MT)
0.9
0.7
0.5
-90
0
90
180
Tuning offset from attended direction

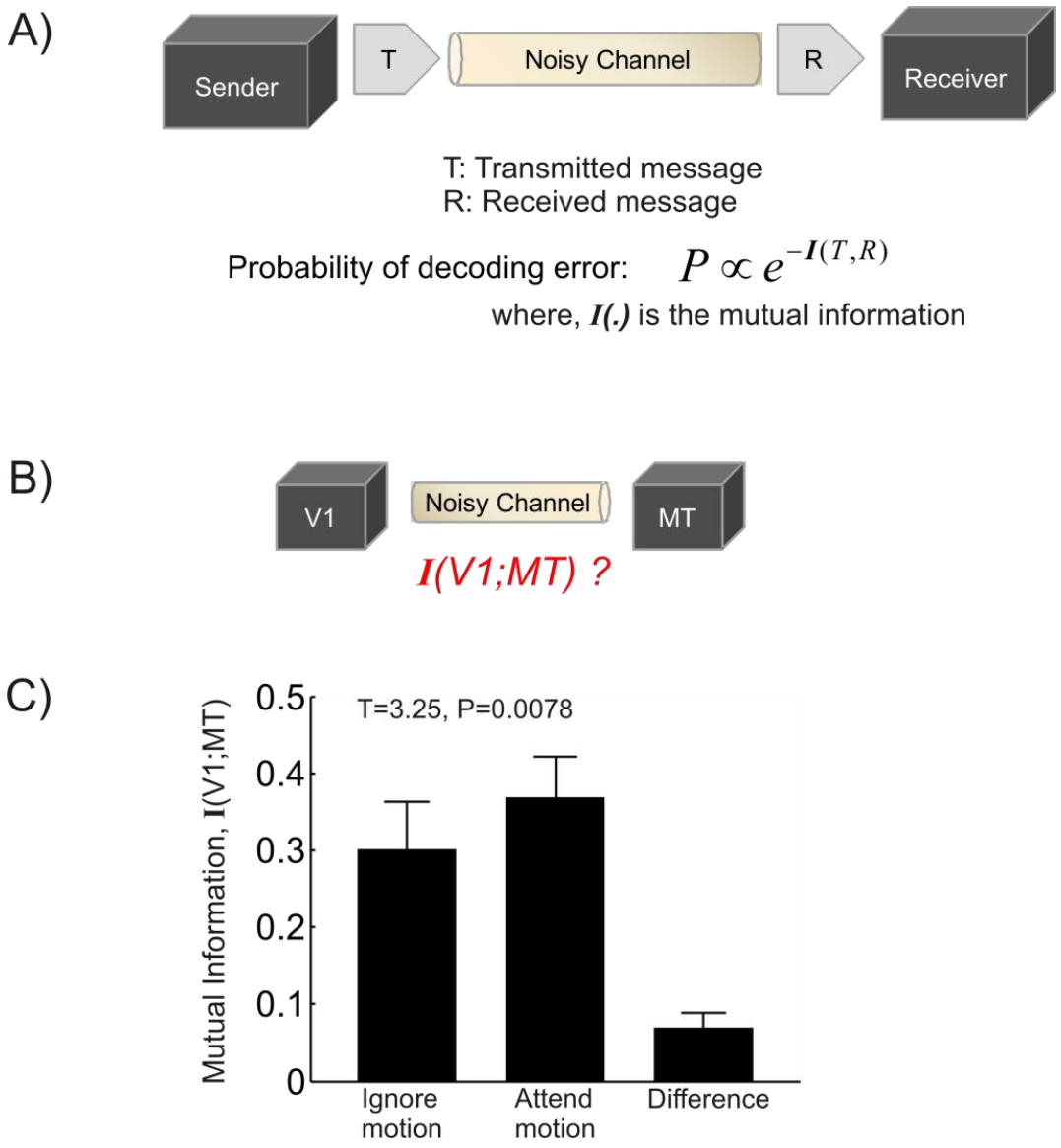


Figure 3.6 A) Shows a generic communication model, and the probability of error based on Shannon's channel coding theorem. B) Shows our communication model for V1 and MT. The efficacy of communication here would be estimated by measuring mutual information between direction-selective channels between the two areas when responding to the same stimulus. C) Shows the mutual information between direction-selective channels of V1 and MT for 'Ignore Motion' trials, 'Attend motion' trials, and the difference between the two that estimates the net increase in communication efficacy as a result of feature attention. There is a significance increase in communication efficacy with attention to direction of motion ($p<0.01$).

FUTURE SCOPE

The theoretical framework, empirical data and model simulations have thrown up a host of interesting questions that require future theoretical and empirical investigations to answer. Firstly, trying to fully uncover the nature and mechanism underlying inter-neuronal correlations and understanding the effect of attention and adaptation on those correlations is currently an open challenge. The preliminary work with MT motion computation model in Chapter 3 hints at the possibility that attention modulation of correlations might simply be a byproduct of gain modulation and not an independent functional mechanism in itself. However, more biophysically accurate simulations and/or novel experiments at single-unit level would be needed to confirm or deny this hypothesis.

Secondly, the neural modulations for attention and adaptation have been researched extensively but in isolation. Understanding how attention and adaptation can jointly mediate sensory information processing is an avenue of investigation that is both important and interesting. Although, I attempted to reconcile the two cognitive phenomena within the framework of dynamic neural coding, a more thorough and rigorous analysis is required, perhaps supplemented with experimental investigations. For instance, although attention and adaptation modulate neural gain in diametrically opposite ways, the dissociation can be simplistically understood as a push-pull mechanism for maintaining a metabolically efficient neural code. However, the modulation of covariance with attention and adaptation is nuanced since the relationship of gain modulation with modulation of covariance is not consistent between attention and adaptation. Understanding this complex interaction might prove to be important in future.

Thirdly, although we found an increase in information coupling between V1 and MT, given the indirect nature of BOLD signal it is not clear how neural populations get synchronized across distal regions of cortical networks. It has been suggested that neurons communicate through

synchronization of their oscillatory activity. Furthermore, recent empirical evidence has suggested that attention may lead to enhanced synchrony in multiple frequency bands. Although, it is likely that an increase in oscillatory synchronization led to the observed increase in coupling of BOLD signals between V1 and MT, however, we cannot ascertain whether the increase in synchrony occurred in higher frequency bands (such as gamma frequency) or lower frequency bands (such as alpha or beta) or both. To answer these questions might require similar experiments but using EEG or MEG as measurement modality.

Finally, aforementioned investigations have shown exciting possibilities for using more computationally sophisticated methods to analyze BOLD data and thereby facilitate investigations that previously lay outside the ambit of fMRI. One promising direction is to exploit the inherent multiplicity of neural composition across voxels to better assess the modulations at neural level by different cognitive factor such as attention and adaptation. In this regard, the two successful methods that were used to extract more accurate estimates of neural modulations than available in raw voxels – forward model and mutual information – could be combined in novel ways to create even more powerful investigatory tools for fMRI analyses.

REFERENCES

Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. Neural Comput 11:91-101.

Albrecht DG, Farrar SB, Hamilton DB (1984) Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. J Physiol 347:713-739.

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183-193.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358-366.

Averbeck BB, Lee D (2006) Effects of noise correlations on information encoding and decoding. J Neurophysiol 95:3633-3644.

Balasubramanian V, Kimber D, Berry MJ, 2nd (2001) Metabolically efficient information processing. Neural Comput 13:799-815.

Barlow H (2001) Redundancy reduction revisited. Network 12:241-253.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Sensory Communication (Rosenbluth WA, ed), pp 217-234. Cambridge, MA: MIT Press.

Beck C, Neumann H (2011) Combining feature selection and integration--a neural model for MT motion selectivity. PLoS One 6:e21254.

Bialek W, Rieke F (1992) Reliability and information transmission in spiking neurons. Trends in neurosciences 15.

Boly M, Garrido MI, Gosseries O, Bruno MA, Boveroux P, Schnakers C, Massimini M, Litvak V, Laureys S, Friston K (2011) Preserved feedforward but impaired top-down processes in the vegetative state. Science 332:858-862.

Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157-189.

Borst A, Theunissen FE (1999) Information theory and neural coding. Nat Neurosci 2:947-957.

Boynton GM (2005) Imaging orientation selectivity: decoding conscious perception in V1. Nat Neurosci 8:541-542.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207-4221.

Boynton, GM (2005) Attention and visual perception. Curr Opin Neurobiol. 15:465-9.

Brenner N, Bialek W, de Ruyter van Steveninck R (2000a) Adaptive rescaling maximizes information transmission. Neuron 26:695-702.

Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000b) Synergy in a neural code. Neural Comput 12:1531-1552.

Bressler DW, Silver MA (2010) Spatial attention improves reliability of fMRI retinotopic mapping signals in occipital and parietal cortex. NeuroImage 53:526-533.

Britten KH (1998) Clustering of response selectivity in the medial superior temporal area of extrastriate cortex in the macaque monkey. Vis Neurosci 15:553-558.

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

Brouwer GJ, and Heeger DJ (2011) Cross-orientation suppression in human visual cortex. J Neurophysiol 106(5):2108-19.

Buracas GT, Boynton GM (2007) The effect of spatial attention on contrast response functions in human visual cortex. J Neurosci 27:93-97.

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

Carandini M, and Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci 13: 51-62

Chai B, Walther DB, Beck DM, Fei-Fei L (2009) Exploring Functional Connectivity of the Human Brain using Multivariate Information Analysis. In: Neural Information Processing Systems.

Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12:1594-1600.

Cohen MR, Maunsell JH (2011a) Using neuronal populations to study the mechanisms underlying spatial and feature attention. Neuron 70:1192-1204.

Cohen MR, Maunsell JH (2011b) When attention wanders: how uncontrolled fluctuations in attention affect performance. J Neurosci 31:15802-15806.

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

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

Conti F, Wanke E (1975) Channel noise in nerve membranes and lipid bilayers. Q Rev Biophys 8:451-506.

Cortes JM, Marinazzo D, Series P, Oram MW, Sejnowski TJ, van Rossum MC (2011) The effect of neural adaptation on population coding accuracy. J Comput Neurosci.

Courtemanche R, Fujii N, Graybiel AM (2003) Synchronous, focally modulated beta-band oscillations characterize local field potential activity in the striatum of awake behaving monkeys. J Neurosci 23:11741-11752.

Cover TM, Thomas JA (1991) Elements of Information Theory. New York: Wiley.

Dayan P, Abbott LF (2001) Theoretical Neuroscience. Cambridge: MIT Press.

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

Ditterich J, Mazurek ME, Shadlen MN (2003) Microstimulation of visual cortex affects the speed of perceptual decisions. Nat Neurosci 6:891-898.

Doya K, Ishii S, Pouget A, Rao RPN (2007) Bayesian brain : probabilistic approaches to neural coding. Cambridge, Mass.: MIT Press.

Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. J Neurosci 31:14272-14283.

Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci 2:704-716.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. Nature 369:525.

Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. Nature 412:787-792.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

Freeman WT, Adelson EH (1991) The design and use of steerable filters. Cambridge, Mass.: Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology.

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends in cognitive sciences 9:474-480.

Fuhrmann Alpert G, Sun FT, Handwerker D, D'Esposito M, Knight RT (2007) Spatio-temporal information analysis of event-related BOLD responses. NeuroImage 34:1545-1561.

Galambos R, Davis H (1943) The response of single auditory-nerve fibers to acoustic stimulation. Journal of Neurophysiology 6:39-57.

Gardner JL, Sun P, Waggoner RA, Ueno K, Tanaka K, Cheng K (2005) Contrast adaptation and representation in human early visual cortex. Neuron 47:607-620.

Geisler WS, Diehl RL (2002) Bayesian natural selection and the evolution of perceptual systems. Philos Trans R Soc Lond B Biol Sci 357:419-448.

Genovese CR, Lazar NA, Nichols TE (2002) Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. NeuroImage **15**:870-878.

Gitelman DR (2002) ILAB: a program for postexperimental eye movement analysis. Behav Res Methods Instrum Comput 34:605-612.

Gregoriou GG, Gotts SJ, Desimone R (2012) Cell-type-specific synchronization of neural activity in FEF with V4 during attention. Neuron 73:581-594.

Gross J, Schmitz F, Schnitzler I, Kessler K, Shapiro K, Hommel B, Schnitzler A (2004) Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. Proc Natl Acad Sci U S A 101:13050-13055.

Gutnisky DA, Dragoi V (2008) Adaptive coding of visual information in neural populations. Nature 452:220-224.

Hamker FH, Zirnsak M, Calow D, Lappe M (2008) The peri-saccadic perception of objects and space. PLoS Comput Biol 4:e31.

Hamming RW (1950) Error detecting and error correcting codes. Bell System Technical Journal 29.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.

Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686-691.

Heeger DJ, Huk AC, Geisler WS, Albrecht DG (2000) Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? Nat Neurosci 3:631-633.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106-154.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215-243.

Hubel DH, Wiesel TN (1969) Anatomical demonstration of columns in the monkey striate cortex. Nature 221:747–750.

Huffman DA (1952) A Method for the Construction of Minimum-Redundancy Codes. Proceedings of the IRE 40:1098-1101.

Ivey C, Apkarian AV, Chialvo DR (1998) Noise-induced tuning curve changes in mechanoreceptors. J Neurophysiol 79:1879-1890.

Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. Nat Neurosci 9:690-696.

Kahn I, Desai M, Knoblich U, Bernstein J, Henninger M, Graybiel AM, Boyden ES, Buckner RL, Moore CI (2011) Characterization of the Functional MRI Response Temporal Linearity via Optical Control of Neocortical Pyramidal Neurons. J Neurosci 31:15086-15091.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679-685.

Kang K, Shapley RM, Sompolinsky H (2004) Information tuning of populations of neurons in primary visual cortex. J Neurosci 24:3726-3735.

Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282:108-111.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751-761.

Kastner S, Ungerleider LG (2001) The neural basis of biased competition in human visual cortex. Neuropsychologia 39:1263-1276.

Kay KN, David SV, Prenger RJ, Hansen KA, Gallant JL (2008) Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. Hum Brain Mapp 29:142-156.

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355.

Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4:219-227.

Kohn A, Movshon JA (2003) Neuronal adaptation to visual motion in area MT of the macaque. Neuron 39:681-691.

Kojadinovic I (2005) On the use of mutual information in data analysis : an overview. In: Applied Stochastic Models and Data Analysis (ASMDA 2005). Brest, France.

Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM Journal of Optimization Vol. 9 Number 1:112-147.

Latham PE, Roudi Y (2009) Mutual Information. Scholarpedia 4.

Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturforsch C 36:910-912.

Laughlin SB, Sejnowski TJ (2003) Communication in neuronal networks. Science 301:1870-1874.

Levy WB, Baxter RA (1996) Energy efficient neural codes. Neural Comput 8:531-543.

Liu T, Mance I (2011) Constant spread of feature-based attention across the visual field. Vision Res 51:26-33.

Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci 23:3963-3971.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.

Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol 77:24-42.

MacKay DJC (2004) Information theory, inference, and learning algorithms, Reprinted with corrections. Edition. Cambridge, UK ; New York: Cambridge University Press.

Manwani A, Koch C (1999) Detecting and estimating signals in noisy cable structure, I: neuronal noise sources. Neural Comput 11:1797-1829.

Manwani A, Koch C (2001) Detecting and estimating signals over noisy and unreliable synapses: information-theoretic analysis. Neural Comput 13:1-33.

Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME (2007) Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. PLoS Biol 5:e19.

Marcelja S (1980) Mathematical description of the responses of simple cortical cells. J Opt Soc Am 70:1297-1300.

Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr Biol 14:744-751.

Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 49:1127-1147.

McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431-441.

Misra N, Singh H, Hnizdo V (2010) Nearest Neighbor Estimates of Entropy for Multivariate Circular Distributions. Entropy 12:1125-1144.

Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron 55:131-141.

Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron 63:879-888.

Mnatsakanov RM, Misra N, Li S, Harner EJ (2008) kn - Nearest Neighbor Estimators of Entropy. Mathematical Methods of Statistics 17:261-277.

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

Mountcastle VB (1997) The columnar organization of the neocortex. Brain 120 (Pt 4):701-722.

Movshon JA, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J Neurosci 16:7733-7741.

Muller JR, Metha AB, Krauskopf J, Lennie P (1999) Rapid adaptation in visual cortex to the structure of images. Science 285:1405-1408.

Nemenman I, Bialek W, de Ruyter van Steveninck R (2004) Entropy and information in neural spike trains: progress on the sampling problem. Phys Rev E Stat Nonlin Soft Matter Phys 69:056111.

Nevado A, Young MP, Panzeri S (2004) Functional imaging and neural information coding. NeuroImage 21:1083-1095.

Newsome WT, Britten KH, Movshon JA (1989) Neuronal correlates of a perceptual decision. Nature 341:52-54.

Nienborg H, Cumming B (2010) Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? Curr Opin Neurobiol.

Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. J Neurosci 31:14551-14564.

Niven JE, Laughlin SB (2008) Energy limitation as a selective pressure on the evolution of sensory systems. J Exp Biol 211:1792-1804.

O'Connor DH, Fukui MM, Pinsk MA, Kastner S (2002) Attention modulates responses in the human lateral geniculate nucleus. Nat Neurosci 5:1203-1209.

O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. Nature 401:584-587.

O'Craven KM, Rosen BR, Kwong KK, Treisman A, Savoy RL (1997) Voluntary attention modulates fMRI activity in human MT-MST. Neuron 18:591-598.

Olshausen BA, Field DJ (1996a) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607-609.

Olshausen BA, Field DJ (1996b) Natural image statistics and efficient coding. Network 7:333-339.

Palva JM, Palva S, Kaila K (2005) Phase synchrony among neuronal oscillations in the human cortex. J Neurosci 25:3962-3972.

Palva S, Palva JM (2007) New vistas for alpha-frequency band oscillations. Trends in neurosciences 30:150-158.

Panzeri S, Magri C, Logothetis NK (2008) On the use of information theory for the analysis of the relationship between neural and imaging signals. Magn Reson Imaging 26:1015-1025.

Panzeri S, Schultz SR, Treves A, Rolls ET (1999) Correlations and the encoding of information in the nervous system. Proc Biol Sci 266:1001-1012.

Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for the sampling bias problem in spike train information measures. J Neurophysiol 98:1064-1072.

Pascual-Leone A, Walsh V (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science 292:510-512.

Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27:1226-1238.

Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1985) Visual Cells in the Temporal Cortex Sensitive to Face View and Gaze Direction. Proceedings of the Royal Society of London Series B, Biological Sciences 223:293-317

Pestilli F, Viera G, Carrasco M (2007) How do attention and adaptation affect contrast sensitivity? J Vis 7:9 1-12.

Pinker S, Bloom P (1990) Natural language and natural selection. Behavioral and Brain Sciences 13:707-784.

Posner MI, Snyder CR, Davidson BJ (1980) Attention and the detection of signals. J Exp Psychol 109:160-174.

Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nat Rev Neurosci 1:125-132.

Pouget A, Dayan P, Zemel RS (2003) Inference and computation with population codes. Annu Rev Neurosci 26:381-410.

Pouget A, Deneve S, and Latham PE (2001) The relevance of Fisher Information for theories of cortical computation and attention. In: Visual attention and cortical circuits, edited by Braun J, Koch C, and Davis JL. Cambridge: MIT Press.

Quian Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10:173-185.

Recanzone GH, Guard DC, Phan ML (2000) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol 83:2315-2331.

Reich DS, Mechler F, Victor JD (2001) Independent and redundant information in nearby cortical neurons. Science 294:2566-2568.

Ress D, Backus BT, Heeger DJ (2000) Activity in primary visual cortex predicts performance in a visual detection task. Nat Neurosci 3:940-945.

Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168-185.

Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. Neuron 26:703-714.

Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1999) Spikes : exploring the neural code. Cambridge, Mass. ; London: MIT.

Rust NC, Mante V, Simoncelli EP, and Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nat Neurosci 9: 1421-1431.

Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. Nat Neurosci 5:631-632.

Saproo S, Serences JT (2010) Spatial attention improves the quality of population codes in human visual cortex. J Neurophysiol 104:885-895.

Sasaki Y, Rajimehr R, Kim BW, Ekstrom LB, Vanduffel W, Tootell RB (2006) The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. Neuron 51:661-670.

Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, Gruber WR, Birbaumer N (2005) A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur J Neurosci 22:2917-2926.

Schneidman E, Bialek W, Berry MJ, 2nd (2003) Synergy, redundancy, and independence in population codes. J Neurosci 23:11539-11553.

Schneidman E, Freedman B, Segev I (1998) Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. Neural Comput 10:1679-1703.

Sclar G, Lennie P, DePriest DD (1989) Contrast adaptation in striate cortex of macaque. Vision Res 29:747-755.

Serences JT, Boynton GM (2007) Feature-based attentional modulations in the absence of direct visual stimulation. Neuron 55:301-312.

Serences JT, Saproo S (2011) Computational advances towards linking BOLD and behavior. Neuropsychologia.

Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. NeuroImage 44:223-231.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889-893.

Series P, Latham PE, Pouget A (2004) Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. Nat Neurosci 7:1129-1135.

Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. Proc Natl Acad Sci U S A 90:10749-10753.

Shadlen MN, Newsome WT (1994) Noise, neural codes and cortical organization. Curr Opin Neurobiol 4:569-579.

Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18:3870-3896.

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916-1936.

Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. Neural Comput 18:1951-1986.

Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD (2006) Adaptive filtering enhances information transmission in visual cortex. Nature 439:936-942.

Silvanto J, Cowey A, Lavie N, Walsh V (2005) Striate cortex (V1) activity gates awareness of motion. Nat Neurosci 8:143-144.

Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vision Research 38:743-761.

Snippe HP, Koenderink JJ (1992) Information in channel-coded systems: correlated receivers. Biol Cybern 67:183-190.

Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. Phys Rev E Stat Nonlin Soft Matter Phys 64:051904.

Strong SP, Koberle R, De Ruyter van Steveninck R, Bialek W (1998) Entropy and information in neural spike trains. Physical Review Letters 80:197-200.

Suffczynski P, Kalitzin S, Pfurtscheller G, Lopes da Silva FH (2001) Computational model of thalamo-cortical networks: dynamical control of alpha rhythms in relation to focal attention. Int J Psychophysiol 43:25-40.

Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon CH, Kim SG, Tong F (2010) Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. J Neurosci 30:325-330.

Tallon-Baudry C (2004) Attention and awareness in synchrony. Trends in cognitive sciences 8:523-525.

Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006) Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J Neurosci 26:9494-9502.

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

Treue S, Maunsell JH (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. J Neurosci 19:7591-7602.

Tsotsos JK, Culhane SM, Wai WYK, Lai Y, Davis N, Nuflo F (1995) Modeling visual attention via selective tuning. Artificial Intelligence 78:507-545.

van Gaal S, Lamme VA (2011) Unconscious High-Level Information Processing: Implication for Neurobiological Theories of Consciousness. Neuroscientist.

van Hateren JH, van der Schaaf A (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond B 265:359-366.

Vinje WE, Gallant JL (2002) Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. J Neurosci 22:2904-2915.

Wainwright MJ (1999) Visual adaptation as optimal information transmission. Vision Res 39:3960-3974.

Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. Neuron 56:366-383.

Wark B, Lundstrom BN, Fairhall A (2007) Sensory adaptation. Curr Opin Neurobiol 17:423-429.

White AL, Carrasco M (2011) Feature-based attention involuntarily and simultaneously improves visual performance across locations. J Vis 11.

White JA, Rubinstein JT, Kay AR (2000) Channel noise in neurons. Trends in neurosciences 23:131-137.

Williford T, Maunsell JH (2006) Effects of spatial attention on contrast response functions in macaque area V4. J Neurophysiol 96:40-54.

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

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

Womelsdorf T, Fries P (2007) The role of neuronal synchronization in selective attention. Curr Opin Neurobiol 17:154-160.